Московский физико-технический институт
(государственный университет)
Лаборатория структурных методов анализа данных
в предсказательном моделировании (ПреМоЛаб)

На правах рукописи

УДК 519.8

Нестеров Юрий Евгеньевич

Алгоритмическая выпуклая оптимизация

Диссертация

на соискание ученой степени

доктора физико-математических наук

по специальности 01.01.07 - вычислительная математика

МОСКВА – 2013

Оглавление

Введение

Актуальность темы и степень ее разработанности

Настоящая диссертация посвящена разработке новых методов решения нелинейных выпуклых задач оптимизации. Теория методов оптимизации является одной из самых востребованных областей численного анализа. Наиболее продвинутая ее часть посвящена решению выпуклых задач. Эти постановки базируются на солидном математическом фундаменте, разработанном в основном в первой половине 20го столетия математиками Г.Минковским, К.Каратеодори, Э.Хелли, В.Фенхелем, А.Александровым и другими (см., например, [74, 56, 6, 1]). Поначалу выпуклый анализ развивался независимо от теории экстремальных задач. Однако после основополагающей монографии Р.Т.Рокафеллара [25] центр развития этой науки окончательно сместился в сторону теории оптимизации [13]. В настоящее время существует большое количество прекрасных книг, как по выпуклому анализу, так и по его оптимизационным приложениям (см., например, [2, 3, 4, 5, 8, 10, 13]). Очень важно, что выпуклые задачи представляют собой практически единственный класс оптимизационных задач, допускающих построение методов с глобальными скоростными характеристиками, приемлемыми для большинства практических приложений. Это обстоятельство привело к появлению большого количества интересных методов и подходов. Основные приоритетные результаты в области выпуклой оптимизации принадлежат отечественным ученым А.Антипину, Ф.П.Васильеву, Е.Г.Гольштейну, В.Ф.Демьянову, Ю.Г.Евтушенко, Ю.М. Ермольеву, А.Д.Иоффе, В.Г.Карманову, Б.С.Мордуховичу, А.С.Немировскому, Б.Т.Поляку, Р.А.Поляку, Б.Н.Пшеничному, А.М.Рубинову, В.М.Тихомирову, Л.Г.Хачияну, Н.З.Шору, Д.Б.Юдину, и другим (см., например, [7, 8, 9, 11, 12, 50, 14, 23, 24, 117]). Результаты по теории сложности экстремальных задач могут рассматриваться как естественное развитие традиций, заложенных еще Л.В.Канторовичем [61].

Однако начиная с выдающейся работы Н.Кармаркара, опубликованной в 1985г., и примерно до 2000г. развитие теории и методов оптимизации было в основном связано с прогрессом в теории полиномиальных методов внутренней точки. Были получены новые и очень эффективные методы решения задач линейного программирования, которые существенно подняли планку соревнования с традиционным симплекс-методом. Была разработана общая теория самосогласованных функций [95], которая позволяла строить полиномиальные методы внутренней точки для всех выпуклых задач с явной структурой.

Появилась теория конических нелинейных задач, в рамках которой удалось построить длинношаговые прямо-двойственные методы внутренней точки, способные ускоряться и стартовать из недопустимых точек. Новые эффективные методы появились в квадратичном и полуопределенном программировании. Очень важно, что прогресс в этом направлении сопровождался и инициировался детальным теоретическим анализом эффективности предлагаемых алгоритмов. Оценки скорости сходимости стали важным аргументом в пользу новых методов. Может быть даже более важным, чем результаты предварительных численных экспериментов.

К концу этого периода выяснилось, что все это время запросы пользователей росли быстрее, чем возможности новых методов. Громадный прогресс в скорости компьютеров, оперативной памяти и возможностях вспомогательных программ, небывалая доступность различной информации через Интернет, дешевизна и доступность вычислительных средств - все это вместе на порядки упростило процесс создания математических моделей и, как неизбежное следствие, существенно увеличило их размеры. Худшей ситуации для методов внутренней точки придумать было нельзя, так как трудоемкость этих методов растет (по крайней мере) пропорционально кубу размерности.

Цели и задачи

В связи с этим в начале 2000х годов встал вопрос о частичной реабилитации почти забытых методов градиентного типа. Однако полностью вернуться на позиции середины 80х годов оказалось невозможным. Дело в том, что к 1985г. теория методов выпуклой оптимизации представляла собой цельный монолит [24], завершенный в основном усилиями А.С.Немировского и Д.Б.Юдина. Разработанная ими теория сложности [79], дающая верхние оценки на возможную эффективность методов минимизации для различных классов задач, была подкреплена оптимальными методами, которые эти оценки как раз и реализовывали. Предположения их вычислительной модели (оракул типа черный ящик) полностью соответствовали существовавшему тогда стилю написания оптимизационных пакетов программ, где пользователю предлагалось написать подпрограмму вычисления значения функции и градиента, закрытую для самого пакета [12]. Таким образом, в то время казалось, что все важные вопросы в этой области были уже заданы и (почти) все отвечены.

Однако к 2000му году этот монолит дал трещину. Дело в том, что область реальных приложений для полиномиальных методов внутренней точки практически совпадает с областью применения черноящичных методов. И там где новые методы были применимы (т.е. были в состоянии сделать хотя бы одну итерацию) они как правило оказывались и в теории и на практике лучше неулучшаемых оптимальных методов. Кажущееся противоречие легко объяснялось тем, что методы внутренней точки для построения хорошего (самосогласованного) барьера требовали доступа к внутренней структуре оракула, т.е. они не были черноящичными.

Таким образом, к началу 2000х годов стало ясно, что некоторые из особо пессимистических границ теории сложности можно скорее всего отодвинуть - для этого был по крайней мере один пример (самосогласованные функции). И было понятно как - надо использовать структуру задачи. Этими соображениями и объясняется основное направление исследований автора, приведшее к появлению излагаемых ниже результатов.

Основной целью настоящей диссертации является разработка более мощных и быстрых методов оптимизации, которые значительно расширяют возможности, обрисованные в стандартной теории сложности.

Новизна, методология и методы исследования

Все приводимые ниже результаты являлись на момент их публикации новыми. Большинство из них было получено в 2002-2007 годах и опубликовано ближе к концу десятилетия в ведущих оптимизационных журналах. Все они связаны с разработкой новых методов оптимизации для различных классов выпуклых задач.

Новые методы обладают улучшенными глобальными скоростными характеристиками и/или возможностями аппроксимации двойственных решений. Их разработка стала возможной благодаря уточнению понятия модели целевой функции. Полезно различать три типа моделей.

1. Геометрические модели. В них обычно обосновывается факт включения некоторого простого выпуклого множества, содержащего текущую точку в более сложный объект (допустимое множество, надграфик целевой функции). В результате появляется возможность сдвинуться в нужном направлении, сохраняя при этом допустимость решения. В качестве примеров можно привести следующие ситуации.

- *Дикинский эллипсоид.* Этот эллипсоид, определяемый гессианом самосогласованной функции лежит в допустимой области.

- *Аппроксимация 1го порядка.* Для функции $f(x)$, чей градиент удовлетворяет условию Липшица с константой L, можно написать следующую аппроксимацию:

  $$f(y) \quad \leq \quad f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2} L \|y - x\|^2.$$

  Минимизируя эту аппроксимацию на допустимом множестве можно улучшить значение целевой функции.

- *Аппроксимация 2го порядка.* Для функции $f(x)$, у которой гессиан является липшицевым с константой , верна верхняя оценка

  $$f(y) \quad \leq \quad f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle + \tfrac{1}{6} M \|y - x\|^3.$$

  Минимизируя эту аппроксимацию получаем модифицированный вариант метода Ньютона, обладающий глобальными гарантиями эффективности.

- *Верхняя аппроксимация для системы уравнений.* Для уравнения $F(x) = 0$, где отображение $F : \mathbb{R}^n \to \mathbb{R}^m$ имеет липшицев якобиан ∇F с константой M, можно написать верхнюю оценку

  $$\|F(y)\| \quad \leq \quad \|F(x) + \nabla F(x)(y - x)\| + \tfrac{1}{2} M \|y - x\|^2.$$

  Минимизация этой оценки дает модифицированный метод Гаусса-Ньютона.

2. Алгебраические модели. В них описывается конкретный способ представления целевой функции. Для выпуклых функций естественной формой представления является максимум (дискретный или непрерывный) выпуклых функций. Например, может случиться что нам известно следующее представление:

$$f(x) \quad = \quad \max_{u \in Q_2} \{ \langle Ax, u \rangle - \phi(u) \}.$$

Эта модель оказывается полезной в прямо-двойственных субградиентных методах, где она позволяет строить приближенные двойственные решения. При благоприятных обстоятельствах, эта модель может использоваться в технике сглаживания, что на порядок увеличивает скорость сходимости соответствующих методов.

3. Структурные модели. В них описывается структура целевой функции. Обычно в ней выделяются два взаимодействующих объекта, один простой и один сложный. Простой объект может иметь плохие функциональные характеристики. Однако он должен допускать решение некоторых вспомогательных задач с его участием в явном виде. Примером может является составная целевая функция, которая формируется как сумма гладкой выпуклой функции и простой выпуклой функции общего вида. Простая функция может быть сколь угодно плохой, даже разрывной (например, индикаторной функцией выпуклого множества).

В этой работе мы покажем как и насколько увеличиваются возможности методов оптимизации, если они построены с учетом вышеупомянутых моделей целевой функции (или их комбинаций). Мы приведем новые алгоритмы следующих типов.

- Прямо-двойственные субградиентные методы для решения негладких выпуклых задач (параграфы 2.1, 2.2).

- Методы минимизации составных функций (параграф 2.3).

- Методы решения вариационных неравенств (Глава 3).

- Методы второго порядка (Глава 4).

- Алгоритмы техники сглаживания (Глава 5).

- Методы для нахождения решений выпуклых задач с относительной точностью (Глава 6).

Теоретическая и практическая значимость работы

В работе предлагаются новые методы для нескольких важных классов выпуклых задач. Все методы являются новыми. При этом они превосходят известные алгоритмы по крайней мере в одном из следующих аспектов:

- вычисление дополнительных характеристик решения (например, наряду с прямым решением, аппроксимируется также и решение двойственной задачи, или производится контроль точности решения);
- обладают гарантиями глобальной эффективности (например, новые методы второго порядка);
- сходятся быстрее, чем разрешает стандартная теория сложности (например, алгоритмы техники сглаживания).

Подобная эффективность многих из разработанных методов ранее считалась недостижимой. Поэтому можно ожидать существенное увеличение размеров решаемых задач, обладающих выпуклой структурой.

Положения выносимые на защиту

- Прямо-двойственные методы для минимизации негладких функций. Обладают неулучшаемой оценкой скорости сходимости. Позволяют формировать приближенное решение двойственной задачи и осуществлять контроль точности.
- Барьерный субградиентный метод. Позволяет минимизировать негладкие функции на множествах заданных самосогласованным барьером. Обладает оптимальной скоростью сходимости. Позволяет максимизировать положительные вогнутые функции с относительной точностью.
- Градиентные методы минимизации составных функций. Применимы к функциям, представимых в виде суммы гладкой функции, заданной черенοящичным оракулом и простой функцией общего вида. Позволяют сохранить скорость сходимости, характерную для гладкого случая.
- Методы решения вариационных неравенств с гладким оператором. Обладают неулучшаемой скоростью сходимости в классе черноящичных методов. Являются двойственным аналогом экстраградиентного метода.
- Методы решения квазивариационных неравенств. Существенно расширен класс неравенств, обладающих единственным решением и построен быстрый метод их решения.

- Кубическая регуляризация метода Ньютона. С помощью кубической верхней оценки целевой функции, построен новый метод второго порядка. Метод работает и на невыпуклых функциях. На выпуклых задачах он сходится со скоростью $O(1/k^2)$, где k - это счетчик итераций. Это первый метод второго порядка, обладающий такой скоростью сходимости на вырожденных выпуклых задачах.

- Ускоренный метод Ньютона. За счет использования небольшой памяти, кубический метод Ньютона удается ускорить. Теперь он сходится как $O(1/k^3)$.

- Модифицированный метод Гаусса-Ньютона. Предлагается новый метод для решения систем нелинейных уравнений. Он основан на использовании верхней квадратичной аппроксимации для нормы невязки. На невырожденных задачах допускает глобальную оценку эффективности.

- Сглаживание явной модели целевой функции. Для подходящей модели, эта техника позволяет решать негладкие задачи с трудоемкостью $O(1/\epsilon)$ итераций метода градиентного типа, где ϵ - это требуемая точность решения. Тем самым преодолевается неулучшаемая оценка $O(1/\epsilon^2)$, справедливая для черноящичных методов.

- Условие обратного зазора. Методы, основанные на этой интерпретации позволяют применять технику сглаживания к более широкому классу задач, в частности включающему сильно выпуклые функции. Это улучшает оценку трудоемкости до $O(1/\epsilon^{1/2})$.

- Техника сглаживания в полуопределенной оптимизации. Показывается, как применять технику сглаживания к симметричным функциям собственных значений симметрических матриц.

- Оптимизация с относительной точностью. Показывается как получать приближенные решения с относительной точностью для специальных классов однородных задач. Получаемые оценки трудоемкости практически не зависят от исходных данных.

- Эллипсоидальная аппроксимация выпуклых тел используется для нахождения правильной евклидовой нормы, позволяющей существенно ускорить скорость сходимости методов оптимизации.

Публикации и апробация результатов

Результаты, включенные в данную работу, опубликованы в 19 статьях в ведущих отечественных и международных журналах (см. [15] - [20], [84] - [94], [96], [97]). В работе также использовались фрагменты из монографий автора [21], [83] и русской версии [22]. В двух работах [96] и [97], написанных в соавторстве, соискателб принадлежат результаты, указанные в положениях, выносимых на защиту.

Результаты, включенные в диссертацию, неоднократно докладывались на научных семинарах в ведущих университетах России и зарубежом: семинар кафедры Оптимальное Управление, МехМат МГУ, руководитель - профессор В.М.Тихомиров (апрель 2012); семинар лаборатории Адаптивных и робастных систем им. Я.З.Цыпкина, ИПУ РАН, руководитель - профессор Б.Т.Поляк (март 2011); семинар лаборатории ПреМоЛаб, МФТИ, руководитель - профессор В.Г.Спокойный (декабрь 2011, апрель 2012); семинар по оптимизации, Технологический Университет Джорджии (США), руководитель - профессор А.С.Немировский (апрели 2008 - 2012гг.); семинар по исследованию операций института IFOR (ETHZ, Zurich), руководитель - профессор H.-J.Luethi (март 2008, сентябрь 2011).

В качестве пленарных и приглашенных докладов, результаты автора докладывались на многочисленных международных конференциях:

International Conference on Operations Research, Цюрих, 2011;

International conference “Foundations of Computational Mathematics”, Будапешт, 2011;

International Conference on Machine Leraning (NIPS), Ванкувер, 2010;

International Conference on Optimization and Applications (CIRM), Барселона, 2010;

International Workshop on Quadratic Optimization, Левен, 2010;

IPAM Workshop on Continuous Optimization, Лос Анжелес, 2010; .

International Congress of Mathematicians, Хидерабад, 2010;

International conference OPTIMA, Черногория, 2009;

20th International Symposium on Mathematical Programming, Чикаго, 2009;

7th EUROPT Workshop on Operations Research, Ремანген, 2009;

7th Brazilian Conference on Continuous Optimization, Кампинас, 2008;

International Conference on Optimization and Operations Research, Северобайкальск, 2008;

School "New Paradigms in Optimization Аскона, 2008;

International conference "High Performance Optimization Амстердам, 2008;

INFORMS meeting on Continuous Optimization, Атланта, 2008;

School on Continuous Optimization, Эйн Геди, 2007;

International Conference on Continuous and Discrete optimization, Ватерлоо, 2007;

International Conference on Nonlinear Optimization, Люмини, 2007;

School on Operations Research and Optimization, Зиналь, 2007;

International Workshop on Continuous Optimization VOCAL, Будапешт, 2006;

International Symposium on Mathematical Programming. Рио-де-Жанейро, 2006;

International Conference on Semidefinite Programming. Сингапур, 2006;

Annual INFORMS meeting. Сан-Франциско, 2005;

International conference on positive polynomials. Люмини, 2005;

International Workshop on Nonlinear Optimization, Обервольфах, 2005;

French-German-Spain conference on Operations Research. Авиньон, 2004;

International conference "High Performance Optimization Амстердам, 2004;

International workshop on semidefinite programming. Ватерлоо, 2004.

На основе приводимых результатов были подготовлены курсы лекций, которые были прочитаны в различных европейских университетах:

- Университет г.Льеж (Бельгия), апрель 2012.
- Независимый Университет, Москва, сентябрь 2012.
- Ecole nationale de la statistique et de l'administration économique (Paris Tech), Париж (Франция), ноябрь 2012.
- Университет г.Вены (Австрия), май 2013.
- Традиционная Математическая Школа по Управлению и Оптимизации, Сенеж, Моск.обл., июнь 2013.

Глава 1

Предварительные результаты

1.1 Классификация выпуклых функций

1.1.1 Нормы и производные

Всюду ниже мы работаем с конечномерным вещественным векторным пространством, обычно обозначаемым буквой E (возможно, с индексом). Это пространство снабжается нормой $\| \cdot \|_E$, точный смысл которой при необходимости будет уточнен. Пространство линейных функций на E (*двойственное пространство*) обозначается через E^*. Для $s \in E^*$ и $x \in E$ будем обозначать через $\langle s, x \rangle_E$ значение линейной функции s в точке x. Другими словами, это *скалярное произведение* s и x. Конечно же, наиболее важным случаем является $E = E^* = \mathbb{R}^n$. Тогда

$$\langle s, x \rangle \;=\; \sum\limits_{i=1}^{n} s^{(i)} x^{(i)}, \quad x, s \in \mathbb{R}^n.$$

Для пространства квадратных матриц $\mathcal{M}^{m,n}$ вводится *фробениусово* скалярное произведение:

$$\langle S, X \rangle \;=\; \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} S^{(i,j)} X^{(i,j)}, \quad X, S \in \mathcal{M}^{m,n}.$$

Пространство симметрических $n \times n$-матриц $\mathcal{S}^n$ интерпретируется как подпространство в $\mathcal{M}^{n,n}$ с соответствующим определением скалярного произведения.

Норма в двойственном пространстве определяется стандартным способом:

$$\|s\|_E^* \;\equiv\; \|s\|_{E^*} = \max_x \{ \langle s, x \rangle : \; \|x\|_E = 1 \}, \tag{1.1.1}$$

который обеспечивает выполнение неравенства *Коши – Буняковского*:

$$\langle s, x \rangle \;\leq\; \|s\|_E^* \, \|x\|_E, \quad x \in E, \; s \in E^*. \tag{1.1.2}$$

Если не возникает двусмысленности, то индексы у скалярных произведений и норм обычно опускаются.

Для определения евклидовой нормы в E нам потребуется самосопряженный положительно определенный линейный оператор $B : E \to E^*$. Тогда

$$\|x\|_B = \langle Bx, x\rangle_B^{1/2}, \quad x \in E. \tag{1.1.3}$$

В этом случае $\|s\|_B^* = \langle s, B^{-1}s\rangle^{1/2}$, $s \in E^*$. Если вид оператора B легко определяется из контекста, то соответствующий индекс опускается.

Иногда в наших рассуждениях используется отношение двойственности между выпуклыми функциями, заданными на прямом и двойственном пространствах. Для выпуклой функции $f(x)$, $x \in E$, ее сопряженная (по Фенхелю) функция относительно выпуклого замкнутого множества $Q \subseteq E$ определяется следующим образом:

$$f_Q^*(s) \;=\; \max_{x\in Q}[\langle s, x\rangle - f(x)], \quad s \in E^*.$$

В этом определении допускается случай, когда $\mathrm{dom}\, f_Q^* \neq E^*$. Если $Q \equiv E$, то соответствующий индекс опускается.

Мы часто будем пользоваться следующим результатом..

Лемма 1.1.1 *Для степени $p \geq 1$ рассмотрим функцию $\psi_p(x) = \frac{1}{p}\|x\|^p$. Тогда*

$$\psi_p^*(s) \;=\; \tfrac{1}{q}\left(\|s\|^*\right)^q, \tag{1.1.4}$$

где степень q находится из уравнения $\frac{1}{p} + \frac{1}{q} = 1$.

Доказательство.
Действительно, если $p > 1$, то для всех $s \in E^*$ имеем

$$\max_{x\in Q}[\; \langle s, x\rangle - \psi_p(x)\;] \;\overset{(1.1.1)}{=}\; \max_{t\geq 0}\left[\; \|s\|^* \cdot t - \tfrac{1}{p}t^p \;\right] \;=\; \tfrac{1}{q}\left(\|s\|^*\right)^q.$$

Если же $p = 1$, то искомый результат получается взятием предела при $p \to 1$. В этом случае функция $\psi_1^*(s)$ является индикаторной функцией единичного шара в E^*. □

Если $Q \equiv E$, то утверждение леммы 1.1.1 может быть записано в виде *неравенства Гельдера*:

$$\tfrac{1}{p}\|x\|^p + \tfrac{1}{q}(\|s\|^*)^q \;\geq\; \langle s, x\rangle, \quad x \in E,\; s \in E^*. \tag{1.1.5}$$

Иногда более удобно пользоваться следующим вариантом:

$$\langle s, x\rangle + \tfrac{1}{p}\sigma\|x\|^p \;\geq\; -\tfrac{p-1}{p}\left(\tfrac{1}{\sigma}\right)^{\frac{1}{p-1}}\|s\|_*^{\frac{p}{p-1}}, \quad x \in E,\; s \in E^*. \tag{1.1.6}$$

Норма, используемая для измерения расстояний в E, естественным образом приводит к нормам для производных функций в пространстве E. Так, например, для функции $f(x)$, $x \in E$, можно определить ее k-ю смешанную производную вдоль направлений $h_1, \dots, h_k \in E$:

$$D^k f(x)[h_1, \dots, h_k] \;\overset{\mathrm{def}}{=}\; \tfrac{d}{dt_1}\dots\tfrac{d}{dt_k} f(x + t_1 h_1 + \ldots + t_k h_k)|_{t_1 = \ldots t_k = 0}.$$

Тогда $\|\nabla^k f(x)\| \overset{\rm def}{=} \max\limits_{h_1,\dots,h_k} \{D^k f(x)[h_1,\dots,h_k] : \ \|h_1\| = \ldots = \|h_k\| = 1\}$[1]. В частности, для $k=1$ имеем

$$\max_h \{Df(x)[h] : \ \|h\| = 1\} \ = \ \max_h \{\langle \nabla f(x), h\rangle : \ \|h\| = 1\} \overset{(1.1.1)}{=} \|\nabla f(x)\|^*.$$

Для $k=2$ получаем

$$\begin{aligned} \|\nabla^2 f(x)\| \ &= \ \max_{h_1,h_2} \{D^2 f(x)[h_1,h_2] : \ \|h_1\| = \|h_2\| = 1\} \\ &= \ \max_{h_1,h_2} \{\langle \nabla^2 f(x) h_1, h_2\rangle : \ \|h_1\| = \|h_2\| = 1\} \\ &\overset{(1.1.1)}{=} \ \max_h \{\|\nabla^2 f(x) h\|^* : \ \|h\| = 1\}. \end{aligned}$$

Если оператор $A : E \to E^*$ является неотрицательно определенным, то $2\langle Ah_1, h_2\rangle \leq \langle Ah_1, h_1\rangle + \langle Ah_2, h_2\rangle$. Следовательно, для выпуклых функций

$$\|\nabla^2 f(x)\| \ = \ \max_h \{\langle \nabla^2 f(x) h, h\rangle : \ \|h\| \leq 1\}. \qquad (1.1.7)$$

1.1.2 Классы гладких функций

Классификация дифференцируемых выпуклых функций основывается на уровне гладкости их производных. Будем обозначать через $\mathcal{F}^{k,p,\nu}_{\|\cdot\|,L}(Q)$ класс функций, которые k раз непрерывно дифференцируемы на некотором выпуклом множестве Q и у которых p-я производная удовлетворяет условию Гельдера с параметром $\nu \in [0,1]$:

$$\|\nabla^p f(x) - \nabla^p f(y)\| \ \leq \ L\|x-y\|^\nu, \quad x,y \in Q. \qquad (1.1.8)$$

При этом если $k = p$, то первый индекс опускается. Индекс $\|\cdot\|$ тоже опускается, если смысл нормы ясен из контекста или если это норма общего вида. Указание на множество тоже отсутствует, если $Q \equiv E$. И для $\nu = 0$ мы отождествляем соответствующий класс с его замыканием. Другими словами, это означает, что производная $\nabla^p f(\cdot)$ существует почти всюду и ее вариация ограничена по норме константой L.

При $\nu = 1$ константа L имеет естественную интерпретацию.

Лемма 1.1.2 *Пусть f - $k+1$ раз непрерывно дифференцируемая функция на множестве Q, и $k \geq 0$. Положим*

$$L \ = \ \max_{x\in Q} \|\nabla^{k+1} f(x)\|. \qquad (1.1.9)$$

Тогда $f \in \mathcal{F}^{k,1}_L(Q)$.

---

[1] Для упрощения обозначений мы считаем, что $\nabla^0 f(x) \equiv f(x)$.

Доказательство.

Действительно, $\nabla^k f(y) - \nabla^k f(x) = \int\limits_0^1 \nabla^{k+1} f(x + \tau(y-x))(y-x) d\tau$. Таким образом,

$$\| \nabla^k f(y) - \nabla^k f(x) \| \;\; \leq \;\; \int\limits_0^1 \| \nabla^{k+1} f(x + \tau(y-x))(y-x) \| d\tau \;\leq\; L \| y - x \|.$$

□

Приведем несколько важных примеров. Мы будем в основном рассматривать случай $k = 1$ или 2 и значения параметра гладкости $\nu = 0$ или 1.

- Включение $f \in \mathcal{F}_L^{0,1}(Q)$ означает, что f удовлетворяет условию Липшица с константой L:
$$|f(x) - f(y)| \;\; \leq \;\; L \| x - y \|, \quad x, y \in Q. \qquad (1.1.10)$$
- Включение $f \in \mathcal{F}_L^{1,0}(Q)$ означает, что у функции f ограничено изменение (суб)градиентов:
$$\| \nabla f(x) - \nabla f(y) \|^* \;\; \leq \;\; L, \quad x, y \in Q. \qquad (1.1.11)$$
- Включение $f \in \mathcal{F}_L^{1,1}(Q)$ означает, что у функции f градиенты удовлетворяют условию Липшица:
$$\| \nabla f(x) - \nabla f(y) \|^* \;\; \leq \;\; L \| x - y \|, \quad x, y \in Q. \qquad (1.1.12)$$
- Включение $f \in \mathcal{F}_L^{2,1}(Q)$ означает, что у функции f липшицев гессиан:
$$\| \nabla^2 f(x) - \nabla^2 f(y) \| \;\; \leq \;\; L \| x - y \|, \quad x, y \in Q. \qquad (1.1.13)$$

В качестве иллюстрации рассмотрим функцию $d_3(x) = \frac{1}{3} \| x \|^3$ с евклидовой нормой (см. формулу (1.1.3)).

Лемма 1.1.3 *Для любых $x, y \in E$ выполняется неравенство*

$$\| \nabla^2 d_3(x) - \nabla^2 d_3(y) \| \;\; \leq \;\; 2 \, \| x - y \|. \qquad (1.1.14)$$

Доказательство.

Для $x \in E$ имеем

$$\nabla^2 d_3(x) \;\; = \;\; \| x \| B + \tfrac{1}{\| x \|} B x x^* B.$$

Зафиксируем две точки $x, y \in E$ и произвольное направление $h \in E$. Положим $x(\tau) = x + \tau(y - x)$ и

$$\phi(\tau) \;\; = \;\; \langle \nabla^2 d_3(x(\tau)) h, h \rangle \;\; = \;\; \| x(\tau) \| \cdot \| h \|^2 + \tfrac{1}{\| x(\tau) \|} \langle B x(\tau), h \rangle^2, \quad \tau \in [0, 1].$$

Пусть сначала $0 \notin [x,y]$. Тогда функция $\phi(\tau)$ непрерывно дифференцируема на $[0,1]$ и

$$\begin{array}{rcl} \phi'(\tau) & = & \frac{\langle Bx(\tau), y-x\rangle}{\|x(\tau)\|} \cdot \|h\|^2 + \frac{2\langle Bx(\tau),h\rangle}{\|x(\tau)\|}\langle Bh, y-x\rangle - \frac{\langle Bx(\tau),h\rangle^2}{\|x(\tau)\|^3}\langle Bx(\tau), y-x\rangle \\ & = & \frac{\langle Bx(\tau), y-x\rangle}{\|x(\tau)\|} \cdot \underbrace{\left(\|h\|^2 - \frac{\langle Bx(\tau), h\rangle^2}{\|x(\tau)\|^2}\right)}_{\geq 0} + \frac{2\langle Bx(\tau),h\rangle}{\|x(\tau)\|}\langle Bh, y-x\rangle. \end{array}$$

Обозначим $\alpha = \frac{\langle Bx(\tau),h\rangle}{\|x(\tau)\|\cdot\|h\|} \in [-1,1]$. Тогда

$$|\phi'(\tau)| \;\leq\; \|y-x\| \cdot \|h\|^2 \cdot (1-\alpha^2+2|\alpha|) \;\leq\; 2\,\|y-x\|\cdot\|h\|^2.$$

Поэтому

$$|\langle(\nabla^2 d_3(y) - \nabla^2 d_3(x))h, h\rangle| \;=\; |\phi(1)-\phi(0)| \;\leq\; 2\,\|y-x\|\cdot\|h\|^2,$$

и мы получаем неравенство (1.1.14) из (1.1.7).

Случай $0 \in [x,y]$ тривиален так как $\nabla^2 d_3(0) = 0$. □

Приведем также следующий тривиальный результат.

Лемма 1.1.4 *Пусть $f \in \mathcal{F}_L^{k+1,\nu}(Q)$, $k \geq 0$. Тогда для всех $x, y \in Q$ выполняется неравенство*

$$\|\nabla^k f(y) - \nabla^k f(x) - \nabla^{k+1} f(x)(y-x)\| \;\leq\; \tfrac{L}{1+\nu}\|y-x\|^{1+\nu}. \qquad (1.1.15)$$

Если $f \in \mathcal{F}_L^{k+2,\nu}(Q)$, $k \geq 0$, то

$$\begin{array}{l} \|\nabla^k f(y) - \nabla^k f(x) - \nabla^{k+1} f(x)(y-x) - \frac{1}{2}\nabla^{k+2} f(x)(y-x)^2\| \\ \leq \frac{L}{(1+\nu)(2+\nu)}\|y-x\|^{2+\nu}. \end{array} \qquad (1.1.16)$$

Доказательство.

Действительно, для доказательства неравенства (1.1.15) заметим, что

$$\begin{array}{rl} & \|\nabla^k f(y) - \nabla^k f(x) - \nabla^{k+1} f(x)(y-x)\| \\ = & \|\int\limits_0^1 \nabla^{k+1}(f(x+\tau(y-x)) - \nabla^{k+1} f(x))(y-x)d\tau\| \\ \leq & L\|y-x\|^{1+\nu}\int\limits_0^1 \tau^\nu d\tau \;=\; \frac{L}{1+\nu}\|y-x\|^{1+\nu}. \end{array}$$

Чтобы доказать неравенство (1.1.16), запишем

$$\begin{array}{rl} & \nabla^k f(y) - \nabla^k f(x) - \nabla^{k+1} f(x)(y-x) - \frac{1}{2}\nabla^{k+2} f(x)(y-x)^2 \\ = & \int\limits_0^1\int\limits_0^t [\nabla^{k+2}(f(x+\lambda(y-x)) - \nabla^{k+2} f(x)](y-x)^2 d\lambda d\tau \end{array}$$

и повторим вышеприведенные выкладки. □

В этой работе мы обычно рассматриваем выпуклые функции. Пересечение этого функционального класса с $\mathcal{F}^{k,p,\nu}_{\|\cdot\|,L}(Q)$ обозначается через $\mathcal{C}^{k,p,\nu}_{\|\cdot\|,L}(Q)$, где смысл индексов остается без изменения. Наиболее важным для нас классом является класс $\mathcal{C}^{1,\nu}_{L}(Q)$ состоящий из функций с гельдеровым градиентом.

Лемма 1.1.5 *Пусть* $f \in \mathcal{C}^{1,\nu}_{L}(Q)$. *Тогда для любых* $x, y \in Q$ *и* $\alpha \in [0,1]$ *справедливы неравенства*

$$f(y) \;\le\; f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{L}{1+\nu}\|y-x\|^{1+\nu}, \tag{1.1.17}$$

$$\begin{aligned} 0 \;&\le\; \alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \\ &\le\; \tfrac{L}{1+\nu}\alpha(1-\alpha)(\alpha^{\nu} + (1-\alpha)^{\nu})\|y-x\|^{1+\nu}. \end{aligned} \tag{1.1.18}$$

Если $Q \equiv E$ *и* $\nu > 0$, *то*

$$f(y) \;\ge\; f(x) + \langle \nabla f(x), y-x\rangle + \tfrac{\nu}{L^{1/\nu}}(\|\nabla f(x) - \nabla f(y)\|^{*})^{1+1/\nu}. \tag{1.1.19}$$

Доказательство.

Неравенство (1.1.17) следует из выпуклости функции f и неравенства (1.1.15). Чтобы доказать неравенство (1.1.18), обозначим $x_\alpha = \alpha x + (1-\alpha) y$. Тогда $x - x_\alpha = (1-\alpha)(x-y)$ и $y - x_\alpha = \alpha(y-x)$. Таким образом,

$$f(x) \overset{(1.1.17)}{\le} f(x_\alpha) + (1-\alpha)\langle \nabla f(x_\alpha), x - y\rangle + \tfrac{L}{1+\nu}\|(1-\alpha)(y-x)\|^{1+\nu},$$

$$f(y) \overset{(1.1.17)}{\le} f(x_\alpha) + \alpha\langle \nabla f(x_\alpha), y - x\rangle + \tfrac{L}{1+\nu}\|\alpha(y-x)\|^{1+\nu}.$$

Складывая эти неравенства с коэффициентами α и $1-\alpha$ соответственно, получаем неравенство (1.1.18).

Докажем теперь неравенство (1.1.19). Обозначим $p = 1+\nu$ и $q = \frac{p}{p-1} = 1 + \frac{1}{\nu}$. Рассмотрим функцию $\phi(y) = f(y) - f(x) - \langle \nabla f(x), y - x\rangle \ge 0$, $y \in E$. Заметим, что $\phi \in \mathcal{C}^{1,\nu}_{L}$. Таким образом,

$$\phi(u) \;\le\; \phi(y) + \langle \nabla \phi(y), u - y\rangle + \tfrac{L}{p}\|u-y\|^{p}, \quad \forall u \in E.$$

Минимизируя обе части этого неравенства по u, в силу леммы 1.1.1 получаем

$$0 \;\le\; \phi(y) - \tfrac{L}{q}(\|\tfrac{1}{L}\nabla\phi(y)\|^{*})^{q}.$$

Это в точности неравенство (1.1.19). □

Мы будем часто пользоваться следующим достаточным условием для класса $\mathcal{C}^{2,1,1}_{L}(Q)$.

Теорема 1.1.1 *Пусть выпуклая функция f дважды непрерывно дифференцируема на выпуклом множестве Q. Положим $L = \sup\limits_{x,\,h}\{\langle \nabla^2 f(x)h, h\rangle \;:\; \|h\| \le 1,\; x \in Q\}$. Тогда $f \in \mathcal{C}_L^{2,1,1}(Q)$.*

Доказательство немедленно следует из леммы 1.1.2 и представления (1.1.7).

1.1.3 Равномерно выпуклые функции

Помимо уровня гладкости, нам понадобится как-то охарактеризовать кривизну выпуклых функций. Это можно сделать с помощью понятия *равномерно выпуклых* функций.

Пусть функция $f(x)$ (суб)дифференцируема на выпуклом замкнутом множестве. Она называется *равномерно выпуклой на Q степени $p \ge 2$*, если существует такая константа $\sigma_p = \sigma_p(f) > 0$ что

$$f(y) \;\ge\; f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{1}{p}\sigma_p \|y - x\|^p \quad \forall x, y \in Q. \tag{1.1.20}$$

Константа σ_p называется *параметром выпуклости* этой функции. Прибавляя к такой функции произвольную выпуклую функцию, мы не меняем ни степень, ни параметр. Заметим, что степень $p = 2$ соответствует *сильно выпуклым* функциям.

Любая равномерно выпуклая функция растет быстрее линейной. Таким образом, ее множества уровня всегда ограничены. Из этого следует, что задачи с равномерно выпуклыми целевыми функциями и непустыми допустимыми множествами всегда разрешимы. Более того, их оптимальное решение всегда единственно.

Складывая два неравенства (1.1.20) с переставленными x и y, получаем

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \;\ge\; \tfrac{2}{p}\sigma_p \|x - y\|^p \quad \forall x, y \in Q. \tag{1.1.21}$$

Оказывается это неравенство является также и достаточным условием равномерной выпуклости (правда, с другим значением параметра).

Лемма 1.1.6 *Пусть для некоторых $p \ge 2$ и $\sigma > 0$, и всех $x, y \in Q$ выполняется неравенство*

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \;\ge\; \sigma \|x - y\|^p, \quad x, y \in Q. \tag{1.1.22}$$

Тогда функция f равномерно выпукла на Q со степенью p и параметром σ.

Доказательство.
Действительно,

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y - x\rangle \quad &= \quad \int\limits_0^1 \langle f(x + \tau(y - x)) - \nabla f(x), y - x\rangle d\tau \\ &= \quad \int\limits_0^1 \tfrac{1}{\tau}\langle f(x + \tau(y - x)) - \nabla f(x), \tau(y - x)\rangle d\tau \\ &\overset{(1.1.22)}{\ge} \quad \int\limits_0^1 \sigma \tau^{p-1} \|y - x\|^p d\tau \;=\; \tfrac{1}{p}\sigma \|y - x\|^p. \qquad \Box \end{aligned}$$

Замечание 1 Сопоставляя неравенства (1.1.21) и (1.1.22), легко увидеть, что не бывает равномерно выпуклых функций со степенью $p < 2$. В противном случае, применяя эти утверждения последовательно, мы получим $\sigma_p \to \infty$.

Лемма 1.1.7 *Пусть функция f является равномерно выпуклой на Q со степенью $p \geq 2$. Тогда*

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle \ \leq \ \tfrac{p-1}{p} \left(\tfrac{1}{\sigma_p} \right)^{\frac{1}{p-1}} \| \nabla f(y) - \nabla f(x) \|_*^{\frac{p}{p-1}}, \quad x, y \in Q. \tag{1.1.23}$$

Доказательство.
Предположим что функция f достигает минимума на E в точке $x^* \in Q$. Тогда

$$\begin{aligned} f(x^*) \quad &= \quad \min_{y \in Q} f(y) \overset{(1.1.20)}{\geq} \min_{x \in Q} \left[f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{p} \sigma_p \| y - x \|^p \right] \\ &\geq \quad \min_{x \in E} \left[f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{p} \sigma_p \| y - x \|^p \right] \\ &\overset{(1.1.4)}{=} \quad f(x) - \tfrac{p-1}{p} \left(\tfrac{1}{\sigma} \right)^{\frac{1}{p-1}} \| \nabla f(x) \|_*^{\frac{p}{p-1}}. \end{aligned}$$

Теперь зафиксируем произвольную точку $x \in Q$ и рассмотрим выпуклую функцию $\phi(y) = f(y) - \langle \nabla f(x), y \rangle$. Эта функция будет равномерно выпуклой со степенью p и параметром σ_p. Более того, она достигает минимума в точке $y = x \in Q$. Поэтому, применив предыдущее неравенство к функции $\phi(y)$, мы получим формулу (1.1.23). □

Приведем один важный пример равномерно выпуклой функции. Зафиксируем произвольную точку $x_0 \in E$. Обозначим $d_p(x) = \frac{1}{p} \| x - x_0 \|^p$, где норма выбрана евклидовой (см. формулу (1.1.3)). Тогда

$$\nabla d_p(x) \ = \ \| x - x_0 \|^{p-2} \cdot B(x - x_0), \quad x \in E.$$

Лемма 1.1.8 *Для любых $x, y \in E$ выполнены неравенства*

$$\langle \nabla d_p(x) - \nabla d_p(y), x - y \rangle \ \geq \ \left(\tfrac{1}{2} \right)^{p-2} \| x - y \|^p, \tag{1.1.24}$$

$$d_p(x) - d_p(y) - \langle \nabla d_p(y), x - y \rangle \ \geq \ \tfrac{1}{p} \left(\tfrac{1}{2} \right)^{p-2} \| x - y \|^p. \tag{1.1.25}$$

Доказательство.
Без ограничения общности можем считать, что $x_0 = 0$. Тогда

$$\begin{aligned} \langle \nabla d_p(x) - \nabla d_p(y), x - y \rangle \ &= \ \langle \| x \|^{p-2} \cdot Bx - \| y \|^{p-2} \cdot By, x - y \rangle \\ &= \ \| x \|^p + \| y \|^p - \langle Bx, y \rangle (\| x \|^{p-2} + \| y \|^{p-2}). \end{aligned}$$

Для доказательства неравенства (1.1.24) нам нужно показать, что правая часть последнего неравенства не меньше чем

$$\left(\tfrac{1}{2}\right)^{p-2}\|x-y\|^p = \left(\tfrac{1}{2}\right)^{p-2}\left[\|x\|^2+\|y\|^2-2\langle Bx,y\rangle\right]^{p/2}.$$

Не ограничивая общности, можно считать, что $x \neq 0$ и $y \neq 0$. Тогда обозначив

$$\tau \ = \ \tfrac{\|y\|}{\|x\|}, \quad \alpha = \tfrac{\langle Bx,y\rangle}{\|x\|\cdot\|y\|} \in [-1,1],$$

мы получим неравенство, которое нужно доказать:

$$1+\tau^p \ \geq \ \alpha\tau(1+\tau^{p-2})+\left(\tfrac{1}{2}\right)^{p-2}[1+\tau^2-2\alpha\tau]^{p/2}, \quad \tau\geq 0, \quad |\alpha|\leq 1. \tag{1.1.26}$$

Так как правая часть этого неравенства выпукла по α, нам нужно рассмотреть только два случая:

$$\begin{aligned} \alpha=1: \qquad 1+\tau^p \ &\geq \ \tau(1+\tau^{p-2})+\left(\tfrac{1}{2}\right)^{p-2}|1-\tau|^p, \\ \alpha=-1: \qquad 1+\tau^p \ &\geq \ -\tau(1+\tau^{p-2})+\left(\tfrac{1}{2}\right)^{p-2}(1+\tau)^p \end{aligned} \tag{1.1.27}$$

при всех $\tau \geq 0$.

Второе из неравенств (1.1.27) получается из нижней оценки для дроби

$$\frac{1+\tau^p+\tau(1+\tau^{p-2})}{(1+\tau)^p} \ = \ \frac{1+\tau^{p-1}}{(1+\tau)^{p-1}}, \quad \tau\geq 0,$$

поскольку ее минимум достигается при $\tau = 1$. Для доказательства первого неравенства заметим что оно выполнено при $\tau = 1$. Если $\tau \geq 0$ и $\tau \neq 1$, то нам нужно оценить снизу дробь

$$\frac{1+\tau^p-\tau(1+\tau^{p-2})}{|1-\tau|^p} \ = \ \frac{(1-\tau)(1-\tau^{p-1})}{|1-\tau|^p} \ = \ \frac{1+\tau+\ldots+\tau^{p-2}}{|1-\tau|^{p-2}}.$$

Так как модуль любого коэффициента полинома $(1-\tau)^{p-2}$ не превосходит 2^{p-2}, первое из неравенств (1.1.27) тоже доказано. Это доказывает неравенство (1.1.24), и для доказательства неравенства (1.1.25) остается воспользоваться леммой 1.1.6. □

В заключение этого параграфа остановимся на важной связи между равномерной выпуклостью и гладкостью выпуклых функций.

Теорема 1.1.2 *Пусть функция f является равномерно выпуклой на множестве Q со степенью p и параметром σ_p. Тогда $f_Q^* \in \mathcal{F}_L^{1,\nu}(E^*)$ с параметрами*

$$\nu \ = \ \tfrac{1}{p-1}, \quad L \ = \ \left(\tfrac{p}{2\sigma_p}\right)^{\frac{1}{p-1}}. \tag{1.1.28}$$

Доказательство.
Так как функция f равномерно выпукла, ее сопряженная функция f_Q^* определена для любого $s \in E^*$. Более того, она дифференцируема, и

$$\nabla f_Q^*(s) \;=\; x(s) \overset{\mathrm{def}}{=} \arg\max_{x \in Q}[\langle s, x\rangle - f(x)].$$

Условия оптимальности для этой задачи оптимизации записываются следующим образом:

$$\langle s - \nabla f(x(s)), x - x(s)\rangle \;\leq\; 0, \quad x \in Q. \tag{1.1.29}$$

Выберем произвольные s_1, $s_2 \in E^*$. Обозначим $x_1 = x(s_1)$, $x_2 = x(s_2)$. Тогда в силу неравенства (1.1.29)

$$\langle s_1 - \nabla f(x_1), x_2 - x_1\rangle \;\leq\; 0, \quad \langle s_2 - \nabla f(x_2), x_1 - x_2\rangle \;\leq\; 0.$$

Складывая эти неравенства, получаем

$$\langle s_1 - s_2, x_1 - x_2\rangle \;\geq\; \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle \overset{(1.1.21)}{\geq} \tfrac{2}{p}\sigma_p \|x_1 - x_2\|^p.$$

Таким образом, $\|s_1 - s_2\|^* \geq \frac{2}{p}\sigma_p \|x_1 - x_2\|^{p-1}$. □

Равномерно выпуклые функции обладают следующим важным свойством.

Теорема 1.1.3 *Пусть функция f является равномерно выпуклой на Q со степенью $p \geq 2$ и параметром $\sigma_p > 0$. Обозначим $x^* = \arg\min_{x \in Q} f(x)$. Тогда для всех $x \in Q$ выполнено неравенство*

$$f(x) \;\geq\; f(x^*) + \tfrac{1}{p}\sigma_p \|x - x^*\|^p. \tag{1.1.30}$$

Доказательство.
Действительно, в силу условия оптимальности

$$\langle \nabla f(x^*), x - x^*\rangle \geq 0, \quad x \in Q.$$

Таким образом, неравенство (1.1.30) следует из оценки (1.1.20). □

1.1.4 Сильно выпуклые функции

Сильно выпуклые функции являются одним из самых важных функциональных классов в выпуклой оптимизации. Сильная выпуклость очень важна в следующих ситуациях.

- Присутствие сильно выпуклых функциональных компонент в задаче оптимизации существенно ускоряет скорость сходимости численных методов.

- Сильно выпуклые функции часто используются в схемах с регуляризацией.
- Простые сильно выпуклые функции используются в качестве прокс-функций в методах оптимизации, учитывающих геометрию допустимых множеств.

В этом пункте мы остановимся на наиболее важных свойствах сильно выпуклых функций.

Прежде всего для удобства ссылок перепишем наиболее важные неравенства пункта 1.1.3 для случая $p = 2$. Если функция f является сильно выпуклой, то для всех $x, y \in Q$ выполнены следующие неравенства:

$$f(y) \;\geq\; f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2}\sigma_2 \|y - x\|^2, \tag{1.1.31}$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \;\geq\; \sigma_2 \|x - y\|^2, \tag{1.1.32}$$

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle \;\leq\; \tfrac{1}{2\sigma_2}(\|\nabla f(y) - \nabla f(x)\|^*)^2. \tag{1.1.33}$$

Все эти неравенства являются необходимыми и достаточными условиями для сильной выпуклости функции f с параметром σ_2.

Дважды дифференцируемые сильно выпуклые функции допускают следующую характеризацию:

$$\langle \nabla^2 f(x) h, h \rangle \overset{(1.1.32)}{\geq} \sigma_2 \|h\|^2 \quad \forall x \in Q,\ h \in E. \tag{1.1.34}$$

Иногда это условие легче проверить в его двойственной форме.

Лемма 1.1.9 *Для положительно определенного самосопряженного линейного оператора* $A : E \to E^*$ *условие*

$$\langle Ah, h \rangle \;\geq\; \sigma \|h\|^2 \quad \forall h \in E, \tag{1.1.35}$$

$\sigma > 0$*, эквивалентно следующему неравенству:*

$$\langle s, A^{-1} s \rangle \;\leq\; \tfrac{1}{\sigma}(\|s\|^*)^2 \quad \forall s \in E^*. \tag{1.1.36}$$

Доказательство.
Заметим, что

$$\tfrac{1}{2}\langle Ah, h \rangle - \tfrac{1}{2}\sigma \|h\|^2 \overset{(1.1.4)}{=} \tfrac{1}{2}\langle Ah, h \rangle - \sigma \max_{s \in E^*} [\langle s, h \rangle - \tfrac{1}{2}(\|s\|^*)^2].$$

Таким образом, условие (1.1.35) эквивалентно следующему:

$$\tfrac{1}{2}\langle Ah, h \rangle + \tfrac{1}{2}\sigma(\|s\|^*)^2 - \sigma \langle s, h \rangle \quad \forall h \in E,\ s \in E^*.$$

Минимизируя правую часть этого неравенства по h получаем оценку (1.1.36). □

Если норма $\|\cdot\|$ в E является евклидовой (см. формулу (1.1.3)), то условие (1.1.35) обеспечивает неотрицательную определенность оператора $A - \sigma B$. Таким образом, простейшая квадратичная функция $d_2(x) = \frac{1}{2}\|x\|^2$ является строго выпуклой с параметром $\sigma_2 = 1$. Другая важная ситуация рассмотрена в следующем примере.

Пример 1.1.1 Пусть $E = \mathbb{R}^n$. Рассмотрим функцию *энтропии*

$$\eta(x) \;=\; \sum_{i=1}^{n} x^{(i)} \ln x^{(i)}, \quad x \in \mathbb{R}^n_+. \tag{1.1.37}$$

Определив скалярное произведение $\langle s, x\rangle = \sum\limits_{i=1}^{n} s^{(i)} x^{(i)}$, получаем следующее представление для ее производных:

$$\langle \nabla \eta(x), h\rangle \;=\; \sum_{i=1}^{n} (1 + \ln x^{(i)}) h^{(i)}, \quad \langle \nabla^2 \eta(x) h, h\rangle \;=\; \sum_{i=1}^{n} \frac{1}{x^{(i)}} (h^{(i)})^2, \quad h \in \mathbb{R}^n.$$

Таким образом, $\langle s, [\nabla^2 \eta(x)]^{-1} s\rangle \;=\; \sum\limits_{i=1}^{n} x^{(i)} (h^{(i)})^2$, $s \;\in\; E^* \;=\; \mathbb{R}^n$. Зафиксировав теперь прямую и двойственную норму:

$$\|x\| \;=\; \sum_{i=1}^{n} |x^{(i)}|, \; x \in E, \quad \|s\|^* \;=\; \max_{1 \le i \le n} |x^{(i)}|, \; s \in E^*,$$

мы видим, что $\langle s, [\nabla^2 \eta(x)]^{-1} s\rangle \le (\|s\|^*)^2 \sum\limits_{i=1}^{n} x^{(i)}$. Таким образом, в силу леммы 1.1.9 и условия (1.1.34) мы получаем, что функция энтропии является строго выпуклой относительно ℓ_1-нормы на множестве

$$Q \;=\; \{x \in \mathbb{R}^n_+ : \; \sum_{i=1}^{n} x^{(i)} \le 1\}$$

с параметром выпуклости $\sigma_2 = 1$.

1.2 Методы гладкой минимизации первого порядка

1.2.1 Прямой и двойственный градиентные методы

В этом параграфе мы рассматриваем методы решения следующей оптимизационной задачи:

$$\text{найти } f^* = \min_{x \in Q} f(x), \tag{1.2.1}$$

где Q - выпуклое замкнутое ограниченное множество, и $f \in \mathcal{C}^{1,1}_L(Q)$. Для простоты будем считать, что константа L известна. Позже мы увидим что от этого предположения легко отказаться.

Обозначим через $T_Q(x) \in Q$ оптимальное решение следующей задачи минимизации:

$$\min_y \left\{ \langle \nabla f(x), y - x\rangle + \tfrac{1}{2} L \|y - x\|^2 : \; y \in Q \right\}. \tag{1.2.2}$$

Обычно оно называется *градиентным отображением.* Если норма $\|\cdot\|$ не является строго выпуклой, то задача (1.2.2) может иметь несколько решений. В этом случае обозначение $T_Q(x)$ будет использоваться для любого из них. В силу неравенства (1.1.17) для любого $x \in Q$ имеем

$$f(T_Q(x)) \le f(x) + \min_y \left\{ \langle \nabla f(x), y - x\rangle + \tfrac{1}{2} L \|y - x\|^2 : \; y \in Q \right\}. \tag{1.2.3}$$

Обозначим через $d(x)$ некоторую *прокс-функцию* множества Q. Предполагается что функция $d(x)$ является непрерывной и сильно выпуклой на множестве Q с параметром выпуклости единица. Для простоты предположим что $d(x)$ дифференцируема. Обозначим через x_0 *центр* множества Q:

$$x_0 = \arg\min_x \{d(x) : \ x \in Q\}.$$

Не ограничивая общности, будем считать что $d(x_0) = 0$. Таким образом, при любом $x \in Q$ имеем

$$\begin{aligned} d(x) \ &\geq \ \tfrac{1}{2}\|x - x_0\|^2, \\ \langle \nabla d(x_0), x - x_0 \rangle \ &\geq \ 0. \end{aligned} \tag{1.2.4}$$

Теперь мы можем определить

$$\rho(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \quad x, y \in Q,$$

брегмановское расстояние между точками x и y. Ясно что,

$$\rho(x, y) \geq \tfrac{1}{2}\|x - y\|^2. \tag{1.2.5}$$

С помощью этого расстояния можно вычислить *брегмановскую проекцию*:

$$\mathcal{B}_Q(x, g) = \arg\min_y \{\langle g, y - x \rangle + \rho(x, y) : \ y \in Q\}.$$

Ввиду неравенств (1.2.5) и (1.1.17), у нас всегда будет

$$f(V_Q(x, \tfrac{1}{L}\nabla f(x))) \ \leq \ f(x) + \min_y \{\langle \nabla f(x), y - x \rangle + L\rho(x, y) : \ y \in Q\} \ \leq \ f(x). \tag{1.2.6}$$

Рассмотрим теперь простейший метод решения задачи (1.2.1).

Прямой градиентный метод

$$x_{k+1} = \mathcal{B}_Q(x_k, \tfrac{1}{L}\nabla f(x_k)), \ k = 0, 1, \ldots. \tag{1.2.7}$$

Теорема 1.2.1 *Пусть* $f \in \mathcal{C}_L^{1,1}(Q)$. *Тогда для всех* $k \geq 1$ *выполнено неравенство*

$$f(\bar{x}_k) - f^* \ \leq \ \tfrac{1}{k}\sum_{i=1}^{k}[f(x_i) - f^*] \ \leq \ \tfrac{1}{k}Ld(x^*), \tag{1.2.8}$$

где $\bar{x}_k = \frac{1}{k}\sum\limits_{i=1}^{k} x_i$.

Доказательство.
Условия оптимальности для вспомогательной задачи в методе (1.2.7) имеет следующий вид:

$$\langle \tfrac{1}{L}\nabla f(x_k) + \nabla d(x_{k+1}) - \nabla d(x_k), x - x_{k+1}\rangle \;\; \geq \;\; 0, \quad x \in Q. \tag{1.2.9}$$

Таким образом, для всех $k \geq 0$ имеем

$$\begin{array}{rcl} d(x_{k+1}) & \overset{(1.1.31)}{\geq} & d(x_k) + \langle \nabla d(x_k), x_{k+1} - x_k\rangle + \frac{1}{2}\|x_{k+1} - x_k\|^2, \\ \langle \nabla d(x_{k+1}), x^* - x_{k+1}\rangle & \overset{(1.2.9)}{\geq} & \langle \nabla d(x_k) - \frac{1}{L}\nabla f(x_k), x^* - x_{k+1}\rangle. \end{array}$$

Складывая эти два неравенства, получаем

$$\begin{array}{rl} & [d(x_{k+1}) + \langle \nabla d(x_{k+1}), x^* - x_{k+1}\rangle] - [d(x_k) + \langle \nabla d(x_k), x^* - x_k\rangle] \\ \geq & \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{1}{L}\langle \nabla f(x_k), x_{k+1} - x^*\rangle \\ = & \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{1}{L}\langle \nabla f(x_k), x_{k+1} - x_k\rangle + \frac{1}{L}\langle \nabla f(x_k), x_k - x^*\rangle \\ \overset{(1.1.17)}{\geq} & \frac{1}{L}[f(x_{k+1}) - f(x_k)] + \frac{1}{L}[f(x_k) - f^*] \; = \; \frac{1}{L}[f(x_{k+1}) - f^*]. \end{array}$$

Просуммировав все эти неравенства, получим

$$\begin{array}{rcl} d(x^*) & \geq & d(x_k) + \langle \nabla d(x_k), x^* - x_k\rangle \; \geq \frac{1}{L}\sum\limits_{i=1}^{k}[f(x_i) - f^*] + d(x_0) + \langle \nabla d(x_0), x^* - x_0\rangle \\ & \overset{(1.2.4)}{\geq} & \frac{1}{L}\sum\limits_{i=1}^{k}[f(x_i) - f^*]. \end{array}$$

□

Рассмотрим теперь *двойственный* градиентный метод, который пересчитывает простейшую *модель* целевой функции.

Двойственный градиентный метод

Инициализация Положим $\psi_0(x) = \frac{1}{2}\rho(x_0, x)$.

При $k \geq 0$ **повторяем:** Вычисляем $x_k = \arg\min\limits_{x \in Q} \psi_k(x)$.

Пересчитываем $\psi_{k+1}(x) = \psi_k(x) + \frac{1}{L}[f(x_k) + \langle \nabla f(x_k), x - x_k\rangle]$.

(1.2.10)

Теорема 1.2.2 *Пусть последовательность* $\{x_k\}_{k\geq 0}$ *сформирована методом (1.2.10). Обозначим* $y_k = T_Q(x_k)$ *и* $\bar{y}_k = \frac{1}{k+1}\sum\limits_{i=0}^{k} y_i$, $k \geq 0$. *Тогда при всех* $k \geq 0$ *выполнено неравенство*

$$f(\bar{y}_k) - f^* \;\leq\; \tfrac{1}{k+1}\sum\limits_{i=0}^{k}[f(y_i) - f^*] \;\leq\; \tfrac{L}{k+1}d(x^*). \tag{1.2.11}$$

Доказательство.
Докажем по индукции, что при всех $k \geq 0$ выполнено неравенство

$$\tfrac{1}{L}\sum\limits_{i=0}^{k} f(y_i) \;\leq\; \psi^*_{k+1} \overset{\rm def}{=} \min\limits_{x\in Q}\psi_{k+1}(x). \tag{1.2.12}$$

При $k = 0$ имеем

$$\begin{aligned}\psi_1^* \;&=\; \tfrac{1}{L}\min\limits_{x\in Q}\,[f(x_0) + \langle\nabla f(x_0), x - x_0\rangle + L\rho(x_0,x)]\\ &\overset{(1.2.5)}{\geq} \tfrac{1}{L}\min\limits_{x\in Q}\left[f(x_0) + \langle\nabla f(x_0), x - x_0\rangle + \tfrac{1}{2}L\|x - x_0\|^2\right] \overset{(1.2.3)}{\geq} \tfrac{1}{L}f(y_0).\end{aligned}$$

Предположим, что неравенство (1.2.12) справедливо при некотором $k \geq 0$. Функция $\psi_k(x)$ сильно выпукла с параметром единица. Следовательно,

$$\begin{aligned}\psi^*_{k+1} &\overset{(1.1.30)}{\geq} \min\limits_{x\in Q}\left\{\psi^*_k + \tfrac{1}{2}\|x - x_k\|^2 + \tfrac{1}{L}[f(x_k) + \langle\nabla f(x_k), x - x_k\rangle]\right\}\\ &\overset{(1.2.3)}{\geq} \psi^*_k + \tfrac{1}{L}f(y_k).\end{aligned}$$

Осталось заметить, что $\psi_k(x) \leq \rho(x_0,x) + \frac{k}{L}f(x)$. Таким образом,

$$\psi^*_{k+1} \;\leq\; \min\limits_{x\in Q}\left[\;\rho(x_0,x) + \tfrac{k+1}{L}f(x)\;\right] \overset{(1.2.4)}{\leq} d(x^*) + \tfrac{k+1}{L}f(x^*). \qquad \square$$

Заметим? что двойственный градиентный метод не поддерживает монотонности минимизирующей последовательности $\{y_k\}_{k\geq 0}$. Для ее построения нам не нужны дополнительные вычисления значения функции и градиента.

1.2.2 Быстрый градиентный метод

Напомним что скорость сходимости прямого и двойственного градиентных методов (1.2.8), (1.2.11) не является оптимальной. Более эффективный *быстрый градиентный метод* формирует две последовательности точек $\{x_k\}_{k=0}^{\infty} \subset Q$ и $\{y_k\}_{k=0}^{\infty} \subset Q$ так, чтобы они удовлетворяли следующему условию:

$$A_k f(y_k) \leq \psi_k \equiv \min\limits_{x}\left\{Ld(x) + \sum\limits_{i=0}^{k}\alpha_i[f(x_i) + \langle\nabla f(x_i), x - x_i\rangle] : \; x \in Q\right\}, \tag{$\mathcal{R}_k$}$$

где $A_k = \sum\limits_{i=0}^{k} \alpha_i$ и последовательность $\{\alpha_i\}_{i=0}^{\infty}$ сформирована из положительных шаговых множителей. Покажем? как это можно сделать.

Действительно, для $k = 0$ выберем любое $\alpha_0 \in (0,1]$ и $y_0 = T_Q(x_0)$. Тогда в силу неравенств (1.2.4) и (1.2.3) имеем

$$\min_x \{Ld(x) + \alpha_0[f(x_0) + \langle \nabla f(x_0), x - x_0\rangle] : \; x \in Q\}$$

$$\geq \alpha_0 \min_x \left\{ \tfrac{L}{2\alpha_0}\|x - x_0\|^2 + f(x_0) + \langle \nabla f(x_0), x - x_0\rangle : \; x \in Q \right\} \geq \alpha_0 f(y_0),$$

и это в точности условие $(\mathcal{R}_0)$.

Обозначим

$$z_k = \arg\min_x \left\{ Ld(x) + \sum_{i=0}^{k} \alpha_i [f(x_i) + \langle \nabla f(x_i), x - x_i\rangle] : \; x \in Q \right\}.$$

Лемма 1.2.1 *Пусть последовательность* $\{\alpha_k\}_{k=0}^{\infty}$ *удовлетворяет условию*

$$\alpha_0 \in (0,1], \quad \alpha_{k+1}^2 \leq A_{k+1}, \; k \geq 0. \tag{1.2.13}$$

Предположим, что соотношение $(\mathcal{R}_k)$ *выполнено при некотором* $k \geq 0$. *Выберем* $\tau_k = \frac{\alpha_{k+1}}{A_{k+1}}$ *и*

$$\begin{aligned} x_{k+1} &= \tau_k z_k + (1 - \tau_k) y_k, \\ y_{k+1} &= T_Q(x_{k+1}). \end{aligned} \tag{1.2.14}$$

Тогда соотношение $(\mathcal{R}_{k+1})$ *также будет выполнено.*

Доказательство.

Действительно, пусть выполнено условие $(\mathcal{R}_k)$. Тогда, так как функция $d(x)$ сильно выпукла, получаем

$$\begin{aligned} \psi_{k+1} \quad &= \quad \min_x \left\{ Ld(x) + \sum_{i=0}^{k+1} \alpha_i [f(x_i) + \langle \nabla f(x_i), x - x_i\rangle] : \; x \in Q \right\} \\ &\overset{(1.1.30)}{\geq} \quad \min_x \left\{ \psi_k + \tfrac{1}{2} L\|x - z_k\|^2 + \alpha_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle] : \; x \in Q \right\}. \end{aligned}$$

Далее, в силу соотношения $(\mathcal{R}_k)$ и первого из правил (1.2.14)

$$\begin{aligned} & \psi_k + \alpha_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle] \\ \geq \; & A_k f(y_k) + \alpha_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle] \\ \geq \; & A_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), y_k - x_{k+1}\rangle] \\ & + \alpha_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle] \\ = \; & A_{k+1} f(x_{k+1}) + \alpha_{k+1}\langle \nabla f(x_{k+1}), x - z_k\rangle. \end{aligned} \tag{1.2.15}$$

В силу условия (1.2.13) имеем $A_{k+1}^{-1} \geq \tau_k^2$. Таким образом, мы можем продолжить наши выкладки следующим образом:

$$\psi_{k+1} \geq A_{k+1} f(x_{k+1}) + \min_x \left\{ \tfrac{1}{2} L \|x - z_k\|^2 + \alpha_{k+1} \langle \nabla f(x_{k+1}), x - z_k \rangle : \ x \in Q \right\}$$

$$= A_{k+1} \left[f(x_{k+1}) + \min_x \left\{ \tfrac{L}{2A_{k+1}} \|x - z_k\|^2 + \tau_k \langle \nabla f(x_{k+1}), x - z_k \rangle : \ x \in Q \right\} \right] \qquad (1.2.16)$$

$$\geq A_{k+1} \left[f(x_{k+1}) + \min_x \left\{ \tfrac{1}{2} \tau_k^2 L \|x - z_k\|^2 + \tau_k \langle \nabla f(x_{k+1}), x - z_k \rangle : \ x \in Q \right\} \right].$$

Наконец заметим, что $\tau_k \in [0, 1]$. Для произвольного $x \in Q$ определим

$$y = \tau_k x + (1 - \tau_k) y_k.$$

Тогда в силу первого из соотношений (1.2.14) имеем

$$y - x_{k+1} = \tau_k (x - z_k).$$

Поэтому в силу неравенства (1.2.3) и второго из правил (1.2.14) заключаем, что

$$\min_x \left\{ \tfrac{1}{2} \tau_k^2 L \|x - z_k\|^2 + \tau_k \langle \nabla f(x_{k+1}), x - z_k \rangle : \ x \in Q \right\}$$

$$= \min_y \left\{ \tfrac{1}{2} L \|y - x_{k+1}\|^2 + \langle \nabla f(x_{k+1}), y - x_{k+1} \rangle : \ y \in \tau_k Q + (1 - \tau_k) y_k \right\}$$

$$\geq \min_y \left\{ \tfrac{1}{2} L \|y - x_{k+1}\|^2 + \langle \nabla f(x_{k+1}), y - x_{k+1} \rangle : \ y \in Q \right\}$$

$$\geq f(y_{k+1}) - f(x_{k+1}).$$

Учитывая это неравенство в конечной оценке (1.2.16), мы приходим к требуемому результату. □

Ясно, что имеется много способов удовлетворить условия (1.2.13). Рассмотрим следующий пример.

Лемма 1.2.2 *При* $k \geq 0$ *положим* $\alpha_k = \frac{k+1}{2}$. *Тогда*

$$\tau_k = \tfrac{2}{k+3}, \quad A_k = \tfrac{(k+1)(k+2)}{4} \qquad (1.2.17)$$

и условия (1.2.13) будут выполнены.

Доказательство.
Действительно, $\tau_k^2 = \frac{\alpha_{k+1}^2}{A_{k+1}^2} = \frac{4}{(k+3)^2} \leq \frac{4}{(k+2)(k+3)} = \frac{1}{A_{k+1}}$, и это в точности оценка (1.2.13). □

Теперь мы можем обосновать скорость сходимости следующего метода.

Быстрый градиентный метод

При $k \geq 0$ повторяем следующие операции.

Вычисляем $f(x_k)$ и $\nabla f(x_k)$. Находим $y_k = T_Q(x_k)$.

Находим $z_k = \arg\min\limits_x \left\{ Ld(x) + \sum\limits_{i=0}^{k} \frac{i+1}{2}[f(x_i) + \langle \nabla f(x_i), x - x_i \rangle] : \ x \in Q \right\}$.

Полагаем $x_{k+1} = \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$.

(1.2.18)

Теорема 1.2.3 *Пусть последовательности* $\{x_k\}_{k=0}^{\infty}$ *и* $\{y_k\}_{k=0}^{\infty}$ *сформированы методом (1.2.18). Тогда при всех* $k \geq 0$ *справедлива оценка*

$$\frac{(k+1)(k+2)}{4} f(y_k) \leq \min_x \left\{ Ld(x) + \sum_{i=0}^{k} \frac{i+1}{2}[f(x_i) + \langle \nabla f(x_i), x - x_i \rangle] : \ x \in Q \right\}. \tag{1.2.19}$$

Таким образом,

$$f(y_k) - f(x^*) \leq \frac{4Ld(x^*)}{(k+1)(k+2)}, \tag{1.2.20}$$

где x^* *- оптимальное решение задачи (1.2.1).*

Доказательство.

Действительно, выберем последовательность $\{\alpha_k\}_{k=0}^{\infty}$ в соответствии с рекомендациями леммы 1.2.2. Тогда в силу леммы 1.2.1 и выпуклости функции $f(x)$ имеем

$$A_k f(y_k) \leq \psi_k \leq Ld(x^*) + A_k f(x^*).$$

Осталось воспользоваться соотношениями (1.2.17). □

Заметим что, вообще говоря, метод (1.2.18) не обеспечивает монотонность минимизирующей последовательности. Однако иногда это свойство является весьма желательным. Для его обеспечения можно воспользоваться следующим приемом.

Заметим что в доказательстве леммы 1.2.1 нам нужно, чтобы точка y_{k+1} удовлетворяла условию

$$f(y_{k+1}) \leq f(T_Q(x_{k+1})).$$

Давайте изменим правила формирования точки y_k в методе (1.2.18) следующим образом.

$$\begin{array}{l} \text{Находим } y'_k = T_Q(x_k). \text{ Вычисляем } f(y'_k). \\[2ex] \text{Полагаем } y_k = \arg\min\limits_x \{ f(x) : \ x \in \{y_{k-1}, x_k, y'_k\} \}. \end{array} \tag{1.2.21}$$

Ясно, что в этом случае

$$f(y_k) \leq f(y_{k-1}) \leq \ldots \leq f(x_0). \tag{1.2.22}$$

1.2.3 Минимизация гладких сильно выпуклых функций

Если у функции $f \in \mathcal{C}_L^{1,1}$ параметр выпуклости σ_2 оказывается строго положительным, то соответствующая задача минимизация становится гораздо проще. Ее сложность теперь будет зависеть от *числа обусловленности*

$$\kappa(f) \;=\; \tfrac{L}{\sigma_2} \;\geq\; 1. \tag{1.2.23}$$

Это число корректно определено для произвольных норм. Однако, оказывается для некоторых норм оно не слишком показательно.

Лемма 1.2.3 *Пусть функция f дважды дифференцируема по крайней мере в одной точке. Тогда ее число обусловленности относительно ℓ_1-нормы не может быть меньше, чем размерность пространства переменных.*

Доказательство.
Пусть функция f дважды дифференцируема в точке $x \in Q$. Обозначим $A = \nabla^2 f(x)$. В силу условий (1.1.9) с $k = 1$, (1.1.34) и (1.1.36) получаем

$$\langle Ah, h\rangle \;\leq\; L\|h\|^2, \quad h \in E,$$

$$\langle s, A^{-1}s\rangle \;\leq\; \tfrac{1}{\sigma_2}(\|s\|^*)^2, \quad s \in E^*.$$

Если отождествить E с $\mathbb{R}^n$, то оператор A может быть представлен симметрической положительно определенной матрицей. Мы измеряем расстояния в прямом пространстве с помощью ℓ_1-нормы, а значит

$$L \;=\; \max_{1\leq i\leq n} A_{ii}.$$

С другой стороны, $\frac{1}{\sigma_2} = \max\limits_s\{\langle s, A^{-1}s\rangle : \; |s^{(i)}| \leq 1, \, i = 1,\ldots,n\}$.

Рассмотрим теперь случайный вектор $\tilde{s}$, компоненты которого принимают значения ± 1 с равными вероятностями. Тогда

$$\tfrac{1}{\sigma_2} \;=\; \max_s\{\langle s, A^{-1}s\rangle : \; |s^{(i)}| \leq 1, \, i = 1,\ldots,n\} \;\geq\; E_{\tilde{s}}(\langle \tilde{s}, A^{-1}\tilde{s}\rangle) \;=\; \langle A^{-1}, I\rangle.$$

Таким образом, мы заключаем, что

$$\tfrac{L}{\sigma_2} \;\geq\; \min_{A\succ 0}\{\langle A^{-1}, I\rangle : \; A_{ii} \leq 1, \, i = 1,\ldots,n\}.$$

Матрица A положительно определена, следовательно $A + A^{-1} \succeq 2I$. Таким образом, для любой матрицы A, допустимой для этой задачи, имеем

$$\langle A^{-1}, I\rangle \;\geq\; \langle 2I - A, I\rangle \;=\; 2n - \langle A, I\rangle \;\geq n. \qquad \Box$$

Этот результат подтверждает, что число обусловленности лучше всего определять относительно евклидовых норм. По крайней мере, в этом случае наши оценки сложности могут не зависеть от размерности пространства.

В оставшейся части этого параграфа мы предполагаем? что расстояния в пространстве E измеряются с помощью евклидовой нормы (1.1.3). Выберем $d(x) = d_2(x) = \frac{1}{2}\|x\|^2$. В этом случае брегмановские расстояния имеют очень простое представление:

$$\rho(x,y) \;=\; \tfrac{1}{2}\|y\|^2 - \tfrac{1}{2}\|x\|^2 - \langle Bx, y-x\rangle \;=\; \tfrac{1}{2}\|x-y\|^2. \tag{1.2.24}$$

Давайте оценим в новой ситуации скорость сходимости прямого градиентного метода (1.2.7).

Теорема 1.2.4 *Пусть функция f принадлежит $\mathcal{C}_L^{1,1}(Q)$ и ее параметр выпуклости σ_2 положителен. Если последовательность $\{x_k\}_{k\ge 0}$ сформирована методом (1.2.7), то при всех $k \ge 1$ справедлива оценка*

$$f(x_k) - f^* \;\le\; \frac{1}{\exp(\frac{\sigma_2}{L}k) - 1} \cdot \frac{\sigma_2}{2}\|x_0 - x^*\|^2 \;\le\; \frac{L}{2k}\|x_0 - x^*\|^2. \tag{1.2.25}$$

Доказательство.
Обозначим $r_k = \frac{1}{2}\|x_k - x^*\|^2$. Пользуясь теми же рассуждениями, что и в начале доказательства теоремы 1.2.1, получаем

$$\begin{aligned}
& r_{k+1} - r_k \\
\overset{(1.2.24)}{=}\;& [d(x^*) - d(x_{k+1}) - \langle \nabla d(x_{k+1}), x^* - x_{k+1}\rangle] - [d(x^*) - d(x_k) - \langle \nabla d(x_k), x^* - x_k\rangle] \\
\le\;& -\tfrac{1}{2}\|x_{k+1} - x_k\|^2 - \tfrac{1}{L}\langle \nabla f(x_k), x_{k+1} - x^*\rangle \\
=\;& -\tfrac{1}{2}\|x_{k+1} - x_k\|^2 - \tfrac{1}{L}\langle \nabla f(x_k), x_{k+1} - x_k\rangle - \tfrac{1}{L}\langle \nabla f(x_k), x_k - x^*\rangle \\
\overset{(1.1.17)}{\le}\;& -\tfrac{1}{L}[f(x_{k+1}) - f(x_k)] - \tfrac{1}{L}\langle \nabla f(x_k), x_k - x^*\rangle \overset{(1.1.31)}{\le} -\tfrac{1}{L}[f(x_{k+1}) - f^*] - \tfrac{\sigma_2}{L} r_k.
\end{aligned}$$

Обозначим $\delta_k = \frac{1}{L}[f(x_k) - f^*]$. Так как метод (1.2.7) монотонен, при всех $k \ge 0$ имеем

$$\begin{aligned}
r_k \;&\le\; \left(1 - \tfrac{1}{\kappa(f)}\right)^k r_0 - \delta_k \sum_{i=0}^{k-1}\left(1 - \tfrac{1}{\kappa(f)}\right)^i \\
&=\; \left(1 - \tfrac{1}{\kappa(f)}\right)^k r_0 - \delta_k \kappa(f)\left(1 - \left(1 - \tfrac{1}{\kappa(f)}\right)^k\right).
\end{aligned}$$

Таким образом, $\delta_k \le \frac{\sigma_2 r_0 \left(1 - \frac{1}{\kappa(f)}\right)^k}{L\left[1 - \left(1 - \frac{1}{\kappa(f)}\right)^k\right]} \le \frac{\sigma_2 r_0 e^{-k/\kappa(f)}}{L\left[1 - e^{-k/\kappa(f)}\right]} = \frac{\sigma_2 r_0}{L\left[e^{k/\kappa(f)} - 1\right]}$. □

Отметим, что оценка скорости сходимости метода (1.2.25) непрерывна по параметру σ_2. Если $\sigma_2 \to 0$, то она переходит в оценку (1.2.8). Для того чтобы обеспечить подобное поведение двойственного градиентного метода (1.2.10), нам придется изменить его схему.

Двойственный градиентный метод (строго выпуклые функции)

Инициализация. Положим $\psi_0(x) = \frac{1}{2}\|x - x_0\|^2$, $a_k = \left(\frac{L}{L-\sigma_2}\right)^k$, $k \geq 0$.

При $k \geq 0$ повторяем следующие операции.

Вычисляем $x_k = \arg\min\limits_{x \in Q} \psi_k(x)$. Пересчитываем

$\psi_{k+1}(x) = \psi_k(x) + \frac{a_k}{L-\sigma_2}[f(x_k) + \langle \nabla f(x_k), x - x_k\rangle + \frac{1}{2}\sigma_2\|x - x_k\|^2]$.

(1.2.26)

Обозначим $A_k = \sum\limits_{i=0}^{k} a_i = (\kappa(f) - 1)\left[\left(\frac{L}{L-\sigma_2}\right)^{k+1} - 1\right]$.

Теорема 1.2.5 *Пусть последовательность $\{x_k\}_{k\geq 0}$ сформирована методом (1.2.10). Положим $y_k = T_Q(x_k)$ и $\bar{y}_k = \frac{1}{A_k}\sum\limits_{i=0}^{k} a_i y_i$, $k \geq 0$. Тогда при всех $k \geq 1$ выполняются неравенства*

$$\begin{aligned} f(\bar{y}_k) - f^* \;&\leq\; \tfrac{1}{A_k}\sum_{i=0}^{k} a_i[f(y_i) - f^*] \;\leq\; \tfrac{\sigma_2}{2\left[e^{(k+1)/\kappa(f)} - 1\right]}\|x^* - x_0\|^2 \\ &\leq\; \tfrac{L}{2(k+1)}\|x^* - x_0\|^2. \end{aligned} \tag{1.2.27}$$

Доказательство.

Доказательство очень похоже на доказательство теоремы 1.2.2. Прежде всего докажем по индукции, что при всех $k \geq 0$ выполнено неравенство

$$\tfrac{1}{L-\sigma_2}\sum_{i=0}^{k} a_i f(y_i) \;\leq\; \psi^*_{k+1} \overset{\mathrm{def}}{=} \min_{x \in Q} \psi_{k+1}(x). \tag{1.2.28}$$

При $k = 0$ имеем

$$\psi^*_1 \;=\; \tfrac{1}{L-\sigma_2}\min_{x \in Q}\left[f(x_0) + \langle \nabla f(x_0), x - x_0\rangle + \tfrac{1}{2}L\|x - x_0\|^2\right] \;\overset{(1.2.3)}{\geq}\; \tfrac{1}{L-\sigma_2} f(y_0).$$

Предположим, что неравенство (1.2.28) верно при некотором $k \geq 0$. Заметим, что

$$1 + \tfrac{\sigma_2}{L-\sigma_2}\sum_{i=0}^{k-1} a_i \;=\; 1 + \tfrac{\sigma_2}{L-\sigma_2}(\kappa(f) - 1)\left[\left(\tfrac{L}{L-\sigma_2}\right)^k - 1\right] \;=\; \left(\tfrac{L}{L-\sigma_2}\right)^k \;=\; a_k.$$

Так как функция $\psi_k(x)$ сильно выпукла с параметром единица,

$$\psi^*_{k+1} \overset{(1.1.30)}{\geq} \min_{x\in Q}\left\{\psi^*_k + \tfrac{1}{2}a_k\|x-x_k\|^2 + \tfrac{a_k}{L-\sigma_2}[f(x_k)+\langle\nabla f(x_k), x-x_k\rangle + \tfrac{1}{2}\sigma_2\|x-x_0\|^2]\right\}$$

$$\overset{(1.2.3)}{\geq} \quad \psi^*_k + \tfrac{a_k}{L-\sigma_2}f(y_k).$$

Осталось заметить, что $\psi_{k+1}(x) \leq \frac{1}{2}\|x-x_0\|^2 + \frac{1}{L-\sigma}A_kf(x)$. Таким образом,

$$\psi^*_{k+1} \leq \min_{x\in Q}\left[\ \tfrac{1}{2}\|x-x_0\|^2 + \tfrac{1}{L-\sigma_2}A_kf(x)\ \right] \leq \tfrac{1}{2}\|x^*-x_0\|^2 + \tfrac{1}{L-\sigma_2}A_kf(x^*).$$

Поскольку $\frac{L}{L-\sigma_2} \geq e^{1/\kappa(f)}$, теорема доказана. □

Как и для прямого градиентного метода (1.2.7), оценка скорости сходимости двойственного градиентного метода (1.2.26) непрерывна по σ_2. Когда σ_2 стремится к нулю, правая часть неравенства (1.2.27) монотонно возрастает и достигает уровня (1.2.11).

Наконец, рассмотрим версию быстрого градиентного метода, применяемую для минимизации сильно выпуклых функций. Как и в двойственном градиентном методе, в методе (1.2.18) можно было бы изменить правила пересчета модели целевой функции, добавив соответствующий квадратичный член. Однако существует и другая возможность.

Обозначим через $\mathrm{Opt}_N(x_0)$ точку y_N, полученную с помощью метода (1.2.18) после N итераций, запущенных из точки x_0. Зададим $N =]\kappa^{1/2}(f)[+1$ и рассмотрим следующий процесс:

$$u_0 \in Q, \quad u_{k+1} = \mathrm{Opt}_{4N}(u_k), \quad k \geq 0. \tag{1.2.29}$$

Мы пользуемся евклидовой метрикой, поэтому

$$\tfrac{2L\|u_k-x^*\|^2}{(4N+1)(4N+2)} \overset{(1.2.20)}{\geq} f(u_{k+1}) - f(x^*) \overset{(1.1.30)}{\geq} \tfrac{\sigma_2}{2}\|u_{k+1}-x^*\|^2.$$

Следовательно, ввиду определения числа N мы получим $\|u_{k+1}-x^*\| \leq \frac{1}{2}\|u_k-x^*\|$. Таким образом, метод (1.2.29) достигнет точности ϵ за

$$O(\kappa^{1/2}(f)\ln\tfrac{1}{\epsilon})$$

итераций. Для прямого и двойственного градиентного методов мы можем гарантировать только оценку в $O(\kappa(f)\ln\frac{1}{\epsilon})$ итераций.

1.3 Самосогласованные функции и барьеры

1.3.1 Определение и свойства самосогласованных функций

В этой работе мы не собираемся детально рассматривать полиномиальные методы внутренней точки. Поэтому ниже приводится очень краткая версия соответствующей теории, которая необходима только для описания некоторых свойств выпуклых множеств и для формулировки вспомогательных оптимизационных задач.

Рассмотрим *выпуклую замкнутую* функцию $f(x) \in C^3(\operatorname{dom} f)$ с *открытой* областью определения. Зафиксируем некоторую точку $x \in \operatorname{dom} f$ и направление $u \in \mathbb{R}^n$. Рассмотрим функцию

$$\phi(x;t) = f(x+tu)$$

как функцию одной переменной $t \in \operatorname{dom} \phi(x;\cdot) \subseteq \mathbb{R}$. Обозначим

$$\begin{aligned} Df(x)[u] &= \phi'(x;t) = \langle \nabla f(x), u \rangle, \\ D^2 f(x)[u,u] &= \phi''(x;t) = \langle \nabla^2 f(x) u, u \rangle = \parallel u \parallel^2_{\nabla^2 f(x)}, \\ D^3 f(x)[u,u,u] &= \phi'''(x;t) = \langle D^3 f(x)[u] u, u \rangle. \end{aligned}$$

Определение 1.3.1 Функция f называется *самосогласованной*, если найдется такая константа $M_f \geq 0$, что неравенство

$$D^3 f(x)[u,u,u] \leq M_f \parallel u \parallel^{3/2}_{\nabla^2 f(x)}$$

выполнено при всех $x \in \operatorname{dom} f$, $u \in \mathbb{R}^n$.

Не стоит считать, что таких функций очень много. Однако мы увидим что они могут быть очень полезными, в частности при построении хорошей барьерной модели допустимого множества. Более того, мы скоро увидим, что эти функции легко минимизировать с помощью метода Ньютона.

Приведем одно эквивалентное определение самосогласованной функции.

Лемма 1.3.1 *Функция f является самосогласованной тогда и только тогда, когда при всех $x \in \operatorname{dom} f$ и всех u_1, u_2, $u_3 \in \mathbb{R}^n$ выполнено неравенство*

$$\mid D^3 f(x)[u_1,u_2,u_3] \mid \leq M_f \prod_{i=1}^{3} \parallel u_i \parallel_{\nabla^2 f(x)} . \tag{1.3.1}$$

Мы примем это утверждение без доказательства, так как оно требует привлечения специальны фактов из теории три-линейных форм.

В дальнейшем, определение 1.3.1, как правило, используется для доказательства самосогласованности конкретных функций, а лемма 1.3.1 оказывается полезной для установления их свойств.

Остановимся на нескольких примерах.

Пример 1.3.1 1. *Линейная функция.* Пусть

$$f(x) = \alpha + \langle a, x \rangle, \quad \operatorname{dom} f = \mathbb{R}^n.$$

Тогда

$$\nabla f(x) = a, \quad \nabla^2 f(x) = 0, \quad \nabla^2 f'(x) = 0,$$

и мы заключаем, что $M_f = 0$.

2. *Выпуклая квадратичная функция.* Рассмотрим

$$f(x) = \alpha + \langle a, x \rangle + \tfrac{1}{2}\langle Ax, x \rangle, \quad \operatorname{dom} f = \mathbb{R}^n,$$

где $A = A^T \succeq 0$. Тогда

$$\nabla f(x) = a + Ax, \quad \nabla^2 f(x) = A, \quad \nabla^2 f'(x) = 0,$$

и мы также получаем $M_f = 0$.

3. *Логарифмический барьер для луча.* Рассмотрим

$$f(x) = -\ln x, \quad \operatorname{dom} f = \{x \in \mathbb{R} \mid \ x > 0\}.$$

Тогда

$$\nabla f(x) = -\tfrac{1}{x}, \quad \nabla^2 f(x) = \tfrac{1}{x^2}, \quad \nabla^2 f'(x) = -\tfrac{2}{x^3}.$$

Таким образом, функция $f(x)$ является самосогласованной с константой $M_f = 2$.

4. *Логарифмический барьер для множеств второго порядка.* Пусть $A = A^T \succeq 0$. Рассмотрим *вогнутую* функцию

$$\phi(x) = \alpha + \langle a, x \rangle - \tfrac{1}{2}\langle Ax, x \rangle.$$

Положим $f(x) = -\ln \phi(x)$, $\operatorname{dom} f = \{x \in \mathbb{R}^n \mid \ \phi(x) > 0\}$. Тогда

$$\begin{array}{rcl} Df(x)[u] & = & -\frac{1}{\phi(x)}[\langle a, u \rangle - \langle Ax, u \rangle], \\ \\ D^2 f(x)[u,u] & = & \frac{1}{\phi^2(x)}[\langle a, u \rangle - \langle Ax, u \rangle]^2 + \frac{1}{\phi(x)}\langle Au, u \rangle, \\ \\ D^3 f(x)[u,u,u] & = & -\frac{2}{\phi^3(x)}[\langle a, u \rangle - \langle Ax, u \rangle]^3 \\ \\ & & -\frac{3}{\phi^2(x)}[\langle a, u \rangle - \langle Ax, u \rangle]\langle Au, u \rangle. \end{array}$$

Обозначим $\omega_1 = Df(x)[u]$, $\omega_2 = \frac{1}{\phi(x)}\langle Au, u \rangle$. Тогда

$$\begin{array}{rcl} D^2 f(x)[u,u] & = & \omega_1^2 + \omega_2 \geq 0, \\ \\ \mid D^3 f(x)[u,u,u] \mid & = & \mid 2\omega_1^3 + 3\omega_1\omega_2 \mid . \end{array}$$

Единственным нетривиальным случаем является $\omega_1 \neq 0$. Обозначим $\alpha = \omega_2/\omega_1^2$. Тогда

$$\tfrac{|D^3 f(x)[u,u,u]|}{(D^2 f(x)[u,u])^{3/2}} \leq \tfrac{2|\omega_1|^3 + 3|\omega_1|\omega_2}{(\omega_1^2 + \omega_2)^{3/2}} = \tfrac{2(1+\frac{3}{2}\alpha)}{(1+\alpha)^{3/2}} \leq 2.$$

Таким образом, эта функция является самосогласованной, $M_f = 2$.

5. Легко проверить, что ни одна из следующих функций не является самосогласованной:
$$f(x) = e^x, \quad f(x) = \tfrac{1}{x^p},\ x > 0,\ p > 0, \quad f(x) =\mid x \mid^p,\ p > 2.$$
$\square$

Приведем теперь основные свойства самосогласованных функций.

Теорема 1.3.1 *Пусть функции f_i являются самосогласоваными с константами M_i, $i = 1, 2$, и пусть α, $\beta > 0$. Тогда функция $f(x) = \alpha f_1(x) + \beta f_2(x)$ также будет самосогласованной с константой*
$$M_f = \max\left\{\tfrac{1}{\sqrt{\alpha}} M_1, \tfrac{1}{\sqrt{\beta}} M_2\right\}$$
и областью определения $\operatorname{dom} f = \operatorname{dom} f_1 \bigcap \operatorname{dom} f_2$.

Доказательство.
Функция f является выпуклой и замкнутой как сумма выпуклых и замкнутых функций. Зафиксируем произвольные $x \in \operatorname{dom} f$ и $u \in \mathbb{R}^n$. Тогда
$$\mid D^3 f_i(x)[u,u,u] \mid \leq M_i \left[D^2 f_i(x)[u,u]\right]^{3/2}, \quad i = 1, 2.$$

Обозначим $\omega_i = D^2 f_i(x)[u,u] \geq 0$. Тогда
$$\tfrac{|D^3 f(x)[u,u,u]|}{[D^2 f(x)[u,u]]^{3/2}} \;\leq\; \tfrac{\alpha |D^3 f_1(x)[u,u,u]| + \beta |D^3 f_2(x)[u,u,u]|}{[\alpha D^1 f_1(x)[u,u] + \beta D^2 f_2(x)[u,u]]^{3/2}} \;\leq\; \tfrac{\alpha M_1 \omega_1^{3/2} + \beta M_2 \omega_2^{3/2}}{[\alpha\omega_1 + \beta\omega_2]^{3/2}}.$$

Правая часть этого неравенства не меняется при замене (ω_1, ω_2) на $(t\omega_1, t\omega_2)$, $t > 0$. Таким образом, можно считать, что
$$\alpha\omega_1 + \beta\omega_2 = 1.$$

Обозначим $\xi = \alpha\omega_1$. Тогда правая часть последнего неравенства будет равной
$$\tfrac{M_1}{\sqrt{\alpha}} \xi^{3/2} + \tfrac{M_2}{\sqrt{\beta}} (1 - \xi)^{3/2}, \quad \xi \in [0, 1].$$

Эта функция выпукла по ξ. Она достигает максимума на концах отрезка. $\square$

Следствие 1.3.1 *Пусть функция f является самосогласованной с константой M_f. Если $A = A^T \succeq 0$, то функция*
$$\phi(x) = \alpha + \langle a, x\rangle + \tfrac{1}{2}\langle Ax, x\rangle + f(x)$$
также является самосогласованной с константой $M_\phi = M_f$.

Доказательство.
Действительно, выпуклая квадратичная функция является самосогласованной с нулевой константой. $\square$

Следствие 1.3.2 *Пусть функция f является самосогласованной с константой M_f, и пусть $\alpha > 0$. Тогда функция $\phi(x) = \alpha f(x)$ также является самосогласованной с константой $M_\phi = \frac{1}{\sqrt{\alpha}} M_f$.* □

Покажем что самосогласованность является аффинно-инвариантным свойством.

Теорема 1.3.2 *Рассмотрим линейный оператор $\mathcal{A}(x) = Ax + b$: $\mathbb{R}^n \to \mathbb{R}^m$. Пусть функция $f(y)$ является самосогласованной с константой M_f. Тогда функция $\phi(x) = f(\mathcal{A}(x))$ также является самосогласованной с константой $M_\phi = M_f$.*

Доказательство.
Ясно что функция $\phi(x)$ является выпуклой и замкнутой. Зафиксируем произвольные $x \in \operatorname{dom} \phi = \{x : \mathcal{A}(x) \in \operatorname{dom} f\}$ и $u \in \mathbb{R}^n$. Обозначим $y = \mathcal{A}(x)$, $v = Au$. Тогда

$$\begin{aligned} D\phi(x)[u] &= \langle \nabla f(\mathcal{A}(x)), Au \rangle = \langle \nabla f(y), v \rangle, \\ D^2\phi(x)[u,u] &= \langle \nabla^2 f(\mathcal{A}(x)) Au, Au \rangle = \langle \nabla^2 f(y) v, v \rangle, \\ D^3\phi(x)[u,u,u] &= D^3 f(\mathcal{A}(x))[Au, Au, Au] = D^3 f(y)[v,v,v]. \end{aligned}$$

Таким образом,

$$\begin{aligned} \mid D^3\phi(x)[u,u,u] \mid &= \mid D^3 f(y)[v,v,v] \mid \leq M_f \langle \nabla^2 f(y) v, v \rangle^{3/2} \\ &= M_f (D^2\phi(x)[u,u])^{3/2}. \qquad \square \end{aligned}$$

Следующее утверждение показывает, что локальные характеристики самосогласованной функции отражают глобальные свойства ее области определения.

Теорема 1.3.3 *Пусть функция f является самосогласованной. Если* $\operatorname{dom} f$ *не содержит прямых, то гессиан этой функции $\nabla^2 f(x)$ является невырожденным в любом $x \in \operatorname{dom} f$.*

Доказательство.
Предположим, что $\langle \nabla^2 f(x) u, u \rangle = 0$ при некотором $x \in \operatorname{dom} f$ и $u \in \mathbb{R}^n$, $u \neq 0$. Рассмотрим точку $y_\alpha = x + \alpha u \in \operatorname{dom} f$ и функцию

$$\psi(\alpha) = \langle \nabla^2 f(y_\alpha) u, u \rangle.$$

Заметим, что

$$\psi'(\alpha) = D^3 f(y_\alpha)[u,u,u] \leq 2\psi(\alpha)^{3/2}, \quad \psi(0) = 0.$$

Поскольку $\psi(\alpha) \geq 0$, мы заключаем, что $\psi'(0) = 0$. Таким образом, эта функция является частью решения следующей системы дифференциальных уравнений:

$$\psi(0) = \xi(0) = 0, \qquad \left\{ \begin{aligned} \psi'(\alpha) &= 2\psi(\alpha)^{3/2} - \xi(\alpha), \\ \xi'(\alpha) &= 0. \end{aligned} \right.$$

Однако эта система имеет единственное тривиальное решение. Таким образом $\psi(\alpha) = 0$ для всех допустимых α.

Мы показали, что функция $\phi(\alpha) = f(y_\alpha)$ является линейной:

$$\begin{aligned}\phi(\alpha) &= f(x) + \langle \nabla f(x), y_\alpha - x\rangle + \int\limits_0^\alpha \int\limits_0^\lambda \langle \nabla^2 f(y_\tau)u, u\rangle d\tau d\lambda \\ &= f(x) + \alpha\langle \nabla f(x), u\rangle.\end{aligned}$$

Предположим, что существует такое $\bar{\alpha}$, что $y_{\bar{\alpha}} \in \partial\,(\mathrm{dom}\, f)$. Рассмотрим такую последовательность $\{\alpha_k\}$, что $\alpha_k \uparrow \bar{\alpha}$. Тогда

$$z_k = (y_{\alpha_k}, \phi(\alpha_k)) \to \bar{z} = (y_{\bar{\alpha}}, \phi(\bar{\alpha})).$$

Заметим что $z_k \in \mathrm{epi}\, f$, но $\bar{z} \notin \mathrm{epi}\, f$, поскольку $y_{\bar{\alpha}} \notin \mathrm{dom}\, f$. Это противоречит замкнутости функции f. Рассматривая направление $-u$, и предполагая, что этот луч пересекает границу, мы опять приходим к противоречию. Таким образом, мы заключаем, что $y_\alpha \in \mathrm{dom}\, f$ при всех α. Однако это противоречит предположениям теоремы. □

Опишем поведение самосогласованных функций у границы области определения.

Теорема 1.3.4 *Пусть функция f является самосогласованной. Тогда для любой точки $\bar{x} \in \partial\,(\mathrm{dom}\, f)$ и любой последовательности*

$$\{x_k\} \subset \mathrm{dom}\, f, \quad x_k \to \bar{x},$$

выполняется условие $f(x_k) \to +\infty$.

Доказательство.
Заметим, что последовательность $\{f(x_k)\}$ ограничена снизу:

$$f(x_k) \geq f(x_0) + \langle \nabla f(x_0), x_k - x_0\rangle.$$

Предположим, что она ограничена сверху. Тогда она имеет по крайней мере одну предельную точку $\bar{f}$. Можно считать, что эта точка единственна. Таким образом,

$$z_k = (x_k, f(x_k)) \to \bar{z} = (\bar{x}, \bar{f}).$$

Заметим, что $z_k \in \mathrm{epi}\, f$, но $\bar{z} \notin \mathrm{epi}\, f$, поскольку $\bar{x} \notin \mathrm{dom}\, f$. Это противоречит замкнутости функции f. □

Таким образом, мы доказали, что $f(x)$ является *барьерной функцией* для множества $\mathrm{Cl}\,(\mathrm{dom}\, f)$.

Основные неравенства

Зафиксируем некоторую самосогласованную функцию $f(x)$. Мы будем считать, что $M_f = 2$ (в противном случае можно умножить $f(x)$ на соответствующую константу, см. следствие 1.3.2). Такие функции называются *стандартно* самосогласованными. Мы также будем предполагать, что область определения $\operatorname{dom} f$ не содержит прямых (в силу теоремы 1.3.3 это обеспечивает невырожденность всех гессианов $\nabla^2 f(x)$).

Обозначим

$$\begin{aligned}\| u \|_x &= \langle \nabla^2 f(x) u, u \rangle^{1/2}, \\ \| v \|_x^* &= \langle [\nabla^2 f(x)]^{-1} v, v \rangle^{1/2}, \\ \lambda_f(x) &= \langle [\nabla^2 f(x)]^{-1} \nabla f(x), \nabla f(x) \rangle^{1/2}.\end{aligned}$$

Ясно, что $| \langle v, u \rangle | \leq \| v \|_x^* \cdot \| u \|_x$. Назовем $\| u \|_x$ *локальной нормой* направления u относительно точки x, а $\lambda_f(x) = \| \nabla f(x) \|_x^*$ - *локальной нормой градиента*[2] $\nabla f(x)$.

Зафиксируем произвольные $x \in \operatorname{dom} f$ и $u \in \mathbb{R}^n$, $u \neq 0$. Рассмотрим следующую функцию одной переменной:

$$\phi(t) = \frac{1}{\langle \nabla^2 f(x+tu) u, u \rangle^{1/2}}$$

с областью определения $\operatorname{dom} \phi = \{ t \in \mathbb{R} : \ x + tu \in \operatorname{dom} f \}$.

Лемма 1.3.2 *При всех допустимых t выполняется неравенство* $| \phi'(t) | \leq 1$.

Доказательство.
Действительно,

$$\phi'(t) = -\frac{\nabla^2 f'(x+tu)[u,u,u]}{2 \langle \nabla^2 f(x+tu) u, u \rangle^{3/2}}.$$

Таким образом, $| \phi'(t) | \leq 1$ в силу Определения 1.3.1. □

Следствие 1.3.3 *Область определения функции $\phi(t)$ содержит интервал*

$$(-\phi(0), \phi(0)).$$

Доказательство.
Поскольку $f(x + tu) \to \infty$ при приближении точки $x + tu$ к границе области $\operatorname{dom} f$ (см. теорему 1.3.4), функция $\langle \nabla^2 f(x + tu) u, u \rangle$ не может быть ограниченной. Таким образом, $\operatorname{dom} \phi \equiv \{ t \mid \ \phi(t) > 0 \}$. Остается заметить, что

$$\phi(t) \geq \phi(0) - | t |$$

в силу леммы 1.3.2. □

---

[2] Иногда $\lambda_f(x)$ называется *ньютоновским убыванием* функции f в точке x.

Рассмотрим следующий эллипсоид:

$$\begin{aligned} W^0(x;r) &= \{y \in \mathbb{R}^n \mid \| y-x \|_x < r\}, \\ W(x;r) &= \mathrm{Cl}\,(W^0(x;r)) \equiv \{y \in \mathbb{R}^n \mid \| y-x \|_x \leq r\}. \end{aligned}$$

Он называется эллипсоидом *Дикина* функции f в точке x.

Теорема 1.3.5 *1. При всех $x \in \mathrm{dom}\, f$ выполнено включение $W^0(x;1) \subseteq \mathrm{dom}\, f$.*
2. При всех $x,\, y \in \mathrm{dom}\, f$ выполнено неравенство

$$\| y-x \|_y \geq \tfrac{\|y-x\|_x}{1+\|y-x\|_x}. \tag{1.3.2}$$

3. Если $\| y-x \|_x < 1$, то

$$\| y-x \|_y \leq \tfrac{\|y-x\|_x}{1-\|y-x\|_x}. \tag{1.3.3}$$

Доказательство.
1. В силу следствия 1.3.3 область определения $\mathrm{dom}\, f$ содержит множество

$$\{y = x + tu \mid \; t^2 \| u \|_x^2 < 1\}$$

(поскольку $\phi(0) = 1/ \| u \|_x$). Это множество есть в точности $W^0(x;1)$.

2. Выберем $u = y - x$. Тогда

$$\phi(1) = \tfrac{1}{\|y-x\|_y}, \quad \phi(0) = \tfrac{1}{\|y-x\|_x},$$

и $\phi(1) \leq \phi(0) + 1$ в силу леммы 1.3.2. Это в точности неравенство (1.3.2).

3. Если $\| y-x \|_x < 1$, то $\phi(0) > 1$ и в силу леммы 1.3.2 имеем $\phi(1) \geq \phi(0) - 1$. Это и есть неравенство (1.3.3). □

Теорема 1.3.6 *Пусть $x \in \mathrm{dom}\, f$. Тогда для любого $y \in W^0(x;1)$ выполнены соотношения*

$$(1 - \| y-x \|_x)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \tfrac{1}{(1-\|y-x\|_x)^2} \nabla^2 f(x). \tag{1.3.4}$$

Доказательство.
Зафиксируем произвольное направление $u \in \mathbb{R}^n$, $u \neq 0$. Рассмотрим функцию

$$\psi(t) = \langle \nabla^2 f(x + t(y-x))u, u \rangle, \quad t \in [0,1].$$

Обозначим $y_t = x + t(y-x)$. Тогда в силу леммы 1.3.1 и неравенства (1.3.3) имеем

$$\begin{aligned} \mid \psi'(t) \mid &= \mid D^3 f(y_t)[y-x,u,u] \mid \leq 2 \| y-x \|_{y_t} \| u \|_{y_t}^2 \\ &= \tfrac{2}{t} \| y_t - x \|_{y_t} \psi(t) \leq \tfrac{2}{t} \cdot \tfrac{\|y_t-x\|_x}{1-\|y_t-x\|_x} \cdot \psi(t) \\ &= \tfrac{2\|y-x\|_x}{1-t\|y-x\|_x} \cdot \psi(t). \end{aligned}$$

Таким образом,

$$2(\ln(1-t \parallel y-x \parallel_x))' \leq (\ln \psi(t))' \leq -2(\ln(1-t \parallel y-x \parallel_x))'.$$

Проинтегрировав это неравенство по $t \in [0,1]$, получаем

$$(1- \parallel y-x \parallel_x)^2 \leq \frac{\psi(1)}{\psi(0)} \leq \frac{1}{(1-\|y-x\|_x)^2}.$$

Это в точности соотношение (1.3.4). □

Следствие 1.3.4 *Пусть* $x \in \operatorname{dom} f$ *и* $r = \parallel y-x \parallel_x < 1$. *Тогда матрицу*

$$G = \int\limits_0^1 \nabla^2 f(x+\tau(y-x))d\tau$$

можно оценить сверху и снизу следующим образом:

$$(1-r+\tfrac{r^2}{3})\nabla^2 f(x) \preceq G \preceq \tfrac{1}{1-r}\nabla^2 f(x).$$

Доказательство.

Действительно в силу теоремы 1.3.6 имеем

$$\begin{aligned} G &= \int\limits_0^1 \nabla^2 f(x+\tau(y-x))d\tau \succeq \nabla^2 f(x) \cdot \int\limits_0^1 (1-\tau r)^2 d\tau \\ &= (1-r+\tfrac{1}{3}r^2)\nabla^2 f(x), \\ G &\preceq \nabla^2 f(x) \cdot \int\limits_0^1 \tfrac{d\tau}{(1-\tau r)^2} = \tfrac{1}{1-r}\nabla^2 f(x). \end{aligned}$$ □

Остановимся еще раз на наиболее важных свойствах самосогласованных функций.

- В любой точке $x \in \operatorname{dom} f$ можно указать *эллипсоид*

  $$W^0(x;1) = \{x \in \mathbb{R}^n \mid \langle \nabla^2 f(x)(y-x), y-x \rangle < 1\},$$

  принадлежащий $\operatorname{dom} f$.

- Внутри эллипсоида $W(x;r)$ с радиусом $r \in [0,1)$ функция f почти квадратична:

  $$(1-r)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \tfrac{1}{(1-r)^2}\nabla^2 f(x)$$

  для всех $y \in W(x;r)$. Выбирая r достаточно малым, можно довести качество квадратичной аппроксимации до любого нужного нам уровня.

Мы завершаем этот пункт оценками уклонения самосогласованной функции от линейной модели.

Теорема 1.3.7 *При всех x, $y \in \operatorname{dom} f$ справедливы неравенства*

$$\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq \tfrac{\|y-x\|_x^2}{1+\|y-x\|_x}, \tag{1.3.5}$$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \omega(\| y - x \|_x), \tag{1.3.6}$$

где $\omega(t) = t - \ln(1+t)$.

Доказательство.
Обозначим $y_\tau = x + \tau(y-x)$, $\tau \in [0,1]$, и $r = \| y - x \|_x$. Тогда в силу (1.3.2) имеем

$$\begin{aligned} \langle \nabla f(y) - \nabla f(x), y - x\rangle &= \int\limits_0^1 \langle \nabla^2 f(y_\tau)(y-x), y-x\rangle d\tau \\ &= \int\limits_0^1 \tfrac{1}{\tau^2} \| y_\tau - x \|_{y_\tau}^2 \, d\tau \\ &\geq \int\limits_0^1 \tfrac{r^2}{(1+\tau r)^2} d\tau = r \int\limits_0^r \tfrac{1}{(1+t)^2} dt = \tfrac{r^2}{1+r}. \end{aligned}$$

Далее, пользуясь неравенством (1.3.5), получаем

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y - x\rangle &= \int\limits_0^1 \langle \nabla f(y_\tau) - \nabla f(x), y - x\rangle d\tau \\ &= \int\limits_0^1 \tfrac{1}{\tau}\langle \nabla f(y_\tau) - \nabla f(x), y_\tau - x\rangle d\tau \\ &\geq \int\limits_0^1 \tfrac{\|y_\tau - x\|_x^2}{\tau(1+\|y_\tau - x\|_x)} d\tau = \int\limits_0^1 \tfrac{\tau r^2}{1+\tau r} d\tau \\ &= \int\limits_0^r \tfrac{tdt}{1+t} = \omega(r). \end{aligned}$$

□

Теорема 1.3.8 *Пусть* $x \in \operatorname{dom} f$ *and* $\| y - x \|_x < 1$. *Тогда*

$$\langle \nabla f(y) - \nabla f(x), y - x\rangle \leq \tfrac{\|y-x\|_x^2}{1-\|y-x\|_x}, \tag{1.3.7}$$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \omega_*(\| y - x \|_x), \tag{1.3.8}$$

где $\omega_*(t) = -t - \ln(1-t)$.

Доказательство.

Обозначим $y_\tau = x + \tau(y - x)$, $\tau \in [0,1]$ и $r = \| \, y - x \, \|_x$. Поскольку $\| \, y_\tau - x \, \| < 1$, в силу неравенства (1.3.3) имеем

$$\begin{aligned}
\langle \nabla f(y) - \nabla f(x), y - x \rangle &= \int\limits_0^1 \langle \nabla^2 f(y_\tau)(y - x), y - x \rangle d\tau \\
&= \int\limits_0^1 \tfrac{1}{\tau^2} \| \, y_\tau - x \, \|^2_{y_\tau} \, d\tau \\
&\leq \int\limits_0^1 \tfrac{r^2}{(1-\tau r)^2} d\tau = r \int\limits_0^r \tfrac{1}{(1-t)^2} dt = \tfrac{r^2}{1-r}.
\end{aligned}$$

Далее, пользуясь неравенством (1.3.7), получаем

$$\begin{aligned}
f(y) - f(x) - \langle \nabla f(x), y - x \rangle &= \int\limits_0^1 \langle \nabla f(y_\tau) - \nabla f(x), y - x \rangle d\tau \\
&= \int\limits_0^1 \tfrac{1}{\tau} \langle \nabla f(y_\tau) - \nabla f(x), y_\tau - x \rangle d\tau \\
&\leq \int\limits_0^1 \tfrac{\|y_\tau - x\|_x^2}{\tau(1 - \|y_\tau - x\|_x)} d\tau = \int\limits_0^1 \tfrac{\tau r^2}{1 - \tau r} d\tau \\
&= \int\limits_0^r \tfrac{t dt}{1-t} = \omega_*(r). \qquad \square
\end{aligned}$$

Теорема 1.3.9 *Неравенства (1.3.2), (1.3.3), (1.3.5), (1.3.6), (1.3.7) и (1.3.8) являются необходимыми и достаточными характеристиками самосогласованных функций.*

Доказательство.

Мы уже доказали следующие импликации:

$$\text{определение } 1.3.1 \;\Rightarrow\; (1.3.2) \;\Rightarrow\; (1.3.5) \;\Rightarrow\; (1.3.6),$$

$$\text{определение } 1.3.1 \;\Rightarrow\; (1.3.3) \;\Rightarrow\; (1.3.7) \;\Rightarrow\; (1.3.8).$$

Докажем теперь импликацию (1.3.6) $\Rightarrow$ определение 1.3.1. Пусть $x \in \operatorname{dom} f$ и $x - \alpha u \in \operatorname{dom} f$ для $\alpha \in [0, \epsilon)$. Рассмотрим функцию

$$\psi(\alpha) = f(x - \alpha u), \quad \alpha \in [0, \epsilon).$$

Обозначим $r = \|u\|_x \equiv [\phi''(0)]^{1/2}$. Предполагая что неравенство (1.3.6) выполняется для всех x и y из $\operatorname{dom} f$, получаем

$$\psi(\alpha) - \psi(0) - \psi'(0)\alpha - \tfrac{1}{2}\psi''(0)\alpha^2 \geq \omega(\alpha r) - \tfrac{1}{2}\alpha^2 r^2.$$

Таким образом,

$$
\begin{aligned}
\tfrac{1}{6}\psi'''(0) &= \lim_{\alpha\downarrow 0}\left[\psi(\alpha)-\psi(0)-\psi'(0)\alpha-\tfrac{1}{2}\psi''(0)\alpha^2\right]\\
&\geq \lim_{\alpha\downarrow 0}\tfrac{1}{\alpha^3}\left[\omega(\alpha r)-\tfrac{1}{2}\alpha^2 r^2\right]=\lim_{\alpha\downarrow 0}\tfrac{r}{3\alpha^2}\left[\omega'(\alpha r)-\alpha r\right]\\
&= \lim_{\alpha\downarrow 0}\tfrac{r}{3\alpha^2}\left[\tfrac{\alpha r}{1+\alpha r}-\alpha r\right]=-\tfrac{r^3}{3}.
\end{aligned}
$$

Следовательно, $D^3f(x)[u,u,u]=-\psi''(0)\leq\psi'''(0)\leq 2[\psi''(0)]^{3/2}$, и это есть определение 1.3.1 с $M_f=2$. Импликация (1.3.8) $\Rightarrow$ определение 1.3.1 доказывается аналогично. □

В вышеприведенных теоремах используются две вспомогательные функции $\omega(t)=t-\ln(1+t)$ и $\omega_*(\tau)=-\tau-\ln(1-\tau)$. Заметим, что

$$
\begin{aligned}
\omega'(t) &= \tfrac{t}{1+t}\geq 0, \qquad & \omega''(t) &= \tfrac{1}{(1+t)^2}>0,\\
\omega_*'(\tau) &= \tfrac{\tau}{1-\tau}\geq 0, \qquad & \omega_*''(\tau) &= \tfrac{1}{(1-\tau)^2}>0.
\end{aligned}
$$

Таким образом, функции $\omega(t)$ и $\omega_*(\tau)$ являются выпуклыми. Мы будем часто пользоваться разными соотношениями между этими функциями. Ниже они представлены в явном виде.

Лемма 1.3.3 *При всех $t\geq 0$ и $\tau\in[0,1)$*

$$\omega'(\omega_*'(\tau))=\tau,\quad \omega_*'(\omega'(t))=t,$$

$$\omega(t)=\max_{0\leq\xi<1}[\xi t-\omega_*(\xi)],\quad \omega_*(\tau)=\max_{\xi\geq 0}[\xi\tau-\omega(\xi)],$$

$$\omega(t)+\omega_*(\tau)\geq\tau t,$$

$$\omega_*(\tau)=\tau\omega_*'(\tau)-\omega(\omega_*'(\tau)),\quad \omega(t)=t\omega'(t)-\omega_*(\omega'(t)).$$

Мы оставляем эти тривиальные факты без доказательства. Основной причиной всех этих соотношений является тот факт, что функции $\omega(t)$ и $\omega_*(t)$ являются *взаимно сопряженными.*

В дальнейшем нам потребуются еще два неравенства.

Теорема 1.3.10 *При всех $x,y\in Q$ справедливо неравенство*

$$f(y)\geq f(x)+\langle\nabla f(x),y-x\rangle+\omega(\|\nabla f(y)-\nabla f(x)\|_y^*). \tag{1.3.9}$$

Если к тому же $\|\nabla f(y)-\nabla f(x)\|_y^*<1$, то

$$f(y)\leq f(x)+\langle\nabla f(x),y-x\rangle+\omega_*(\|\nabla f(y)-\nabla f(x)\|_y^*). \tag{1.3.10}$$

Доказательство.

Зафиксируем произвольные $x, y \in Q$. Рассмотрим функцию

$$\phi(z) = f(z) - \langle \nabla f(x), z \rangle, \quad z \in Q.$$

Эта функция является самосогласованной, и $\phi'(x) = 0$. Таким образом, в силу неравенства (1.3.8) получаем

$$\begin{aligned} f(x) - \langle \nabla f(x), x \rangle &= \phi(x) = \min_{z \in Q} \phi(z) \\ &\leq \min_{z \in Q} [\phi(y) + \langle \phi'(y), z - y \rangle + \omega_*(\|z - y\|_y)] \\ &= \phi(y) - \omega(\|\phi'(y)\|_y^*) \\ &= f(y) - \langle \nabla f(x), y \rangle - \omega(\|\nabla f(y) - \nabla f(x)\|_y^*). \end{aligned}$$

А это и есть неравенство (1.3.9). Неравенство (1.3.10) доказывается аналогичным образом с помощью оценки (1.3.6). □

1.3.2 Минимизация самосогласованных функций

Рассмотрим следующую задачу минимизации:

$$\min\{f(x) \mid x \in \operatorname{dom} f\}. \tag{1.3.11}$$

Следующая теорема дает нам достаточное условие существования ее решения. Напомним, что мы считаем функцию f стандартно самосогласованной на множестве $\operatorname{dom} f$, которое не содержит прямых.

Теорема 1.3.11 *Пусть $\lambda_f(x) < 1$ для некоторого $x \in \operatorname{dom} f$. Тогда решение x_f^* задачи (1.3.11) существует и единственно.*

Доказательство.

Действительно, в силу неравенства (1.3.6) для любого $y \in \operatorname{dom} f$ имеем

$$\begin{aligned} f(y) &\geq f(x) + \langle \nabla f(x), y - x \rangle + \omega(\| y - x \|_x) \\ &\geq f(x) - \| \nabla f(x) \|_x^* \cdot \| y - x \|_x + \omega(\| y - x \|_x) \\ &= f(x) - \lambda_f(x) \cdot \| y - x \|_x + \omega(\| y - x \|_x). \end{aligned}$$

Таким образом, для любого $y \in \mathcal{L}_f(f(x)) = \{y \in \mathbb{R}^n \mid f(y) \leq f(x)\}$ получаем

$$\tfrac{1}{\|y-x\|_x} \omega(\| y - x \|_x) \leq \lambda_f(x) < 1.$$

Заметим, что функция $\frac{1}{t}\omega(t) = 1 - \frac{1}{t}\ln(1+t)$ строго возрастает по t. Поэтому $\| y - x \|_x \leq \bar{t}$, где $\bar{t}$ является единственным положительным корнем уравнения

$$(1 - \lambda_f(x))t = \ln(1+t).$$

Таким образом, $\mathcal{L}_f(f(x))$ ограничено, и поэтому решение x_f^* существует. Оно единственно, поскольку в силу оценки (1.3.6) для всех $y \in \operatorname{dom} f$ выполнено неравенство

$$f(y) \geq f(x_f^*) + \omega(\| y - x_f^* \|_{x_f^*}). \qquad \Box$$

Таким образом, мы показали, что локальное условие $\lambda_f(x) < 1$ дает нам некоторую глобальную информацию о функции f, а именно гарантирует существование точки минимума x_f^*. Заметим, что результат теоремы 1.3.11 не может быть усилен.

Пример 1.3.2 Зафиксируем произвольное $\epsilon > 0$. Рассмотрим функцию одной переменной

$$f_\epsilon(x) = \epsilon x - \ln x, \quad x > 0.$$

Как мы видели в примере 1.3.1 и следствии 1.3.1, эта функция самосогласована. Заметим, что

$$\nabla f_\epsilon(x) = \epsilon - \tfrac{1}{x}, \quad \nabla^2 f_\epsilon = \tfrac{1}{x^2}.$$

Таким образом, $\lambda_{f_\epsilon}(x) = \mid 1 - \epsilon x \mid$. Следовательно, при $\epsilon = 0$ получим $\lambda_{f_0}(x) = 1$ для всех $x > 0$. При этом f_0 не ограничена снизу.

Если $\epsilon > 0$, то $x_{f_\epsilon}^* = \frac{1}{\epsilon}$. Заметим, что мы можем гарантировать существование минимума проверкой в точке $x = 1$, даже если ϵ очень мало. $\Box$

Рассмотрим теперь схему *редуцированного метода Ньютона.*

Редуцированный метод Ньютона
Выбираем $x_0 \in \operatorname{dom} f$. Вычисляем $x_{k+1} = x_k - \frac{1}{1+\lambda_f(x_k)}[\nabla^2 f(x_k)]^{-1}\nabla f(x_k),\ k \geq 0.$

(1.3.12)

Теорема 1.3.12 *При всех $k \geq 0$ справедлива оценка*

$$f(x_{k+1}) \leq f(x_k) - \omega(\lambda_f(x_k)). \tag{1.3.13}$$

Доказательство.
Обозначим $\lambda = \lambda_f(x_k)$. Тогда $\| x_{k+1} - x_k \|_x = \frac{\lambda}{1+\lambda} = \omega'(\lambda)$. Таким образом, в силу неравенства (1.3.8) и леммы 1.3.3 имеем

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \omega_*(\| x_{k+1} - x_k \|_x) \\ &= f(x_k) - \tfrac{\lambda^2}{1+\lambda} + \omega_*(\omega'(\lambda)) \\ &= f(x_k) - \lambda\omega'(\lambda) + \omega_*(\omega'(\lambda)) = f(x_k) - \omega(\lambda). \end{aligned}$$

□

Итак, для всех таких $x \in \operatorname{dom} f$, что $\lambda_f(x) \geq \beta > 0$, один шаг редуцированного метода Ньютона уменьшает значение функции $f(x)$ по крайней мере на константу $\omega(\beta) > 0$. Результат теоремы 1.3.12 может быть использован для получения глобальной оценки эффективности этого процесса.

Опишем теперь *локальную* сходимость *стандартного* метода Ньютона.

Стандартный метод Ньютона

(1.3.14)

0. Выбираем $x_0 \in \operatorname{dom} f$.

1. Вычисляем $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$, $k \geq 0$.

Заметим, что скорость сходимости этого процесса может измеряться разными способами. Мы можем смотреть на *невязку по функции* $f(x_k) - f(x_f^*)$, или на локальную норму градиента $\lambda_f(x_k) = \| \nabla f(x_k) \|_{x_k}^*$, или на *локальную оценку расстояния до минимума* $\| x_k - x_f^* \|_{x_k}$. В конце концов, можно смотреть и на расстояние до минимума в фиксированной метрике

$$r_*(x_k) \equiv \|x_k - x_f^*\|_{x_f^*},$$

задаваемой самой точкой минимума. Докажем, что все эти меры локально эквивалентны.

Теорема 1.3.13 *Пусть* $\lambda_f(x) < 1$. *Тогда*

$$\omega(\lambda_f(x)) \leq f(x) - f(x_f^*) \leq \omega_*(\lambda_f(x)), \tag{1.3.15}$$

$$\omega'(\lambda_f(x)) \leq \| x - x_f^* \|_x \leq \omega_*'(\lambda_f(x)), \tag{1.3.16}$$

$$\omega(r_*(x)) \leq f(x) - f(x_f^*) \leq \omega_*(r_*(x)), \tag{1.3.17}$$

где последнее неравенство справедливо при $r_*(x) < 1$.

Доказательство.
Обозначим $r =\| x - x^*_f \|_x$ и $\lambda = \lambda_f(x)$. Неравенства (1.3.15) следуют из теоремы 1.3.10. Далее, в силу неравенства (1.3.5) имеем

$$\tfrac{r^2}{1+r} \leq \langle \nabla f(x), x - x^*_f \rangle \leq \lambda r.$$

Это и есть правая часть неравенства (1.3.16). Если $r \geq 1$, то левая часть этого неравенства тривиальна. Предоложим, что $r < 1$. Тогда $\nabla f(x) = G(x - x^*_f)$, где

$$G = \int\limits_0^1 \nabla^2 f(x^*_f + \tau(x - x^*_f)) d\tau,$$

и

$$\lambda^2_f(x) = \langle [\nabla^2 f(x)]^{-1} G(x - x^*_f), G(x - x^*_f) \rangle \leq \| H \|^2 r^2,$$

где $H = [\nabla^2 f(x)]^{-1/2} G [\nabla^2 f(x)]^{-1/2}$. В силу следствия 1.3.4 имеем

$$G \preceq \tfrac{1}{1-r} \nabla^2 f(x).$$

Следовательно, $\| H \| \leq \frac{1}{1-r}$, и мы заключаем что

$$\lambda^2_f(x) \leq \tfrac{r^2}{(1-r)^2} = (\omega'_*(r))^2.$$

Таким образом, $\lambda_f(x) \leq \omega'_*(r)$. Применив $\omega'(\cdot)$ к обеим частям этого неравенства, мы получим оставшуюся часть неравенства (1.3.16).

Наконец, неравенства (1.3.17) следуют из оценок (1.3.6) и (1.3.8). □

Оценим теперь локальную скорость сходимости стандартного метода Ньютона (1.3.14). Это будет удобнее сделать в терминах локальной нормы градиента $\lambda_f(x)$.

Теорема 1.3.14 *Пусть* $x \in \mathrm{dom}\, f$ *и* $\lambda_f(x) < 1$. *Тогда точка*

$$x_+ = x - [\nabla^2 f(x)]^{-1} \nabla f(x)$$

принадлежит $\mathrm{dom}\, f$. *При этом*

$$\lambda_f(x_+) \leq \left(\tfrac{\lambda_f(x)}{1-\lambda_f(x)} \right)^2.$$

Доказательство.
Обозначим $p = x_+ - x$, $\lambda = \lambda_f(x)$. Тогда $\| p \|_x = \lambda < 1$. Таким образом, $x_+ \in \mathrm{dom}\, f$ (см. теорему 1.3.5). В силу теоремы 1.3.6

$$\begin{aligned} \lambda_f(x_+) &= \langle [\nabla^2 f(x_+)]^{-1} \nabla f(x_+), \nabla f(x_+) \rangle^{1/2} \\ &\leq \tfrac{1}{1-\|p\|_x} \| \nabla f(x_+) \|_x = \tfrac{1}{1-\lambda} \| \nabla f(x_+) \|_x . \end{aligned}$$

Далее,

$$\nabla f(x_+) = \nabla f(x_+) - \nabla f(x) - \nabla^2 f(x)(x_+ - x) = Gp,$$

где $G = \int\limits_0^1 [\nabla^2 f(x + \tau p) - \nabla^2 f(x)] d\tau$. Следовательно,

$$\| \nabla f(x_+) \|_x^2 = \langle [\nabla^2 f(x)]^{-1} Gp, Gp \rangle \leq \| H \|^2 \cdot \| p \|_x^2,$$

где $H = [\nabla^2 f(x)]^{-1/2} G [\nabla^2 f(x)]^{-1/2}$. В силу следствия 1.3.4 имеем

$$(-\lambda + \tfrac{1}{3}\lambda^2) \nabla^2 f(x) \preceq G \preceq \tfrac{\lambda}{1-\lambda} \nabla^2 f(x).$$

Таким образом, $\| H \| \leq \max \left\{ \frac{\lambda}{1-\lambda}, \lambda - \frac{1}{3}\lambda^2 \right\} = \frac{\lambda}{1-\lambda}$, и мы заключаем, что

$$\lambda_f^2(x_+) \leq \tfrac{1}{(1-\lambda)^2} \| \nabla f(x_+) \|_x^2 \leq \tfrac{\lambda^4}{(1-\lambda)^4}. \qquad \Box$$

Теорема 1.3.14 дает нам следующее описание области квадратичной сходимости метода (1.3.14):

$$\lambda_f(x) < \bar{\lambda} = \tfrac{3-\sqrt{5}}{2} = 0.3819\ldots,$$

где $\bar{\lambda}$ является корнем уравнения $\frac{\lambda}{(1-\lambda)^2} = 1$. В этом случае мы можем гарантировать, что $\lambda_f(x_+) < \lambda_f(x)$.

Таким образом, наши результаты обосновывают следующую стратегию решения исходной задачи (1.3.11).

- *Первая стадия*: $\lambda_f(x_k) \geq \beta$, где $\beta \in (0, \bar{\lambda})$. На этой стадии мы применяем редуцированный метод Ньютона. На каждой итерации этого метода выполнено неравенство

  $$f(x_{k+1}) \leq f(x_k) - \omega(\beta).$$

  Таким образом, число шагов на этой стадии ограничено следующим образом:

  $$N \leq \tfrac{1}{\omega(\beta)} [f(x_0) - f(x_f^*)].$$

- *Вторая стадия*: $\lambda_f(x_k) \leq \beta$. На этой стадии мы применяем стандартный метод Ньютона. Этот процесс сходится квадратично:

  $$\lambda_f(x_{k+1}) \leq \left(\tfrac{\lambda_f(x_k)}{1-\lambda_f(x_k)} \right)^2 \leq \tfrac{\beta \lambda_f(x_k)}{(1-\beta)^2} < \lambda_f(x_k).$$

Можно показать, что локальная сходимость редуцированного метода Ньютона (1.3.12) тоже квадратична:

$$x_+ = x - \tfrac{[\nabla^2 f(x)]^{-1} \nabla f(x)}{1+\lambda_f(x)} \quad \Rightarrow \quad \lambda_f(x_+) \leq 2\lambda_f^2(x). \tag{1.3.18}$$

Однако мы предпочитаем переключающуюся стратегию, поскольку она дает лучшие оценки эффективности. Соотношение (1.3.18) может быть обосновано так же, как это сделано в теореме 1.3.14.

1.3.3 Самосогласованные барьеры и метод отслеживания траектории

Мотивировка

В предыдущем пункте мы показали, что метод Ньютона очень эффективен для минимизации стандартной *самосогласованной* функции. Такая функция всегда является барьером для своей области определения. Давайте посмотрим, как эти функции могут быть использованы в схемах последовательной безусловной минимизации.

В этом пункте мы будем работать с задачами минимизации специального вида. Обозначим $\mathrm{dom\,f} = \mathrm{Cl\,(dom\,f)}$.

Определение 1.3.2 Задача минимизации называется *стандартной*, если она представлена в форме

$$\min\{\langle c, x\rangle \mid \ x \in Q\}, \tag{1.3.19}$$

где Q - выпуклое замкнутое множество. Предполагается, что известна самосогласованная функция f, у которой $\mathrm{dom\,f} = \mathrm{Q}$.

Рассмотрим параметрическую штрафную функцию

$$f(t; x) = t\langle c, x\rangle + f(x)$$

с параметром $t \geq 0$. Заметим, что функция $f(t; x)$ самосогласована по x (см. следствие 1.3.1). Обозначим

$$x^*(t) = \arg \min_{x \in \mathrm{dom}\, f} f(t; x).$$

Эта траектория называется *центральной* для задачи (1.3.19). Можно ожидать, что $x^*(t) \to x^*$ при $t \to \infty$. Нашей целью является отслеживание этой траектории.

Напомним, что стандартный метод Ньютона, применяемый к функции $f(t; x)$, сходится квадратично (см. теорему 1.3.14). Более того, у нас имеется явное описание области квадратичной сходимости:

$$\lambda_{f(t;\cdot)}(x) \leq \beta < \bar{\lambda} = \tfrac{3-\sqrt{5}}{2}.$$

Давайте посмотрим на наши возможности, предположив, что нам известна точка $x = x^*(t)$ при некотором $t > 0$.

Мы собираемся увеличить t:

$$t_+ = t + \Delta, \quad \Delta > 0.$$

При этом мы хотим сохранить x в области квадратичной сходимости метода Ньютона для функции $f(t + \Delta; \cdot)$:

$$\lambda_{f(t+\Delta;\cdot)}(x) \leq \beta < \bar{\lambda}.$$

Заметим, что пересчет $t \to t_+$ не меняет гессиан барьерной функции:

$$\nabla^2 f(t+\Delta; x) = \nabla^2 f(t; x).$$

Таким образом, нетрудно понять, насколько большим может быть шаг Δ. Действительно, из условия оптимальности первого порядка можно выписать *уравнение центральной траектории:*

$$tc + \nabla f(x^*(t)) = 0. \tag{1.3.20}$$

Поскольку $tc + \nabla f(x) = 0$, получаем

$$\lambda_{f(t+\Delta;\cdot)}(x) =\parallel t_+c + \nabla f(x) \parallel_x = \Delta \parallel c \parallel_x = \tfrac{\Delta}{t} \parallel \nabla f(x) \parallel_x \leq \beta.$$

Поэтому если мы хотим увеличивать параметр t с *линейной скоростью*, то необходимо предположить, что величина

$$\lambda_f^2(x) =\parallel \nabla f(x) \parallel_x^2 \equiv \langle [\nabla^2 f(x)]^{-1} \nabla f(x), \nabla f(x) \rangle$$

равномерно ограничена на $\operatorname{dom} f$.

Таким образом, мы приходим к определению *самосогласованного барьера.*

Определение самосогласованного барьера

Определение 1.3.3 Пусть функция $F(x)$ является стандартной самосогласованной. Мы назовем ее *ν-самосогласлванным барьером* для множества $\operatorname{dom} \mathrm{F}$, если

$$\sup_{u \in \mathbb{R}^n} [2\langle \nabla F(x), u \rangle - \langle \nabla^2 F(x) u, u \rangle] \leq \nu \tag{1.3.21}$$

для всех $x \in \operatorname{dom} F$. Число ν называется *параметром* этого барьера.

Заметим, что в этом определении не требуется невырожденность гессиана $\nabla^2 F(x)$. Однако если это так, то неравенство (1.3.21) может быть записано в эквивалентной форме

$$\langle [\nabla^2 F(x)]^{-1} \nabla F(x), \nabla F(x) \rangle \leq \nu. \tag{1.3.22}$$

Иногда удобна следующая форма неравенства (1.3.21):

$$\langle \nabla F(x), u \rangle^2 \leq \nu \langle \nabla^2 F(x) u, u \rangle \quad \forall u \in \mathbb{R}^n. \tag{1.3.23}$$

(Чтобы ее получить для направления u удовлетворяюшего условию $\langle \nabla^2 F(x) u, u \rangle > 0$, заменим u в неравенстве (1.3.21) на λu и найдем максимум левой части по λ.) Заметим, что условие (1.3.23) может быть записано в матричной форме:

$$\nabla^2 F(x) \succeq \tfrac{1}{\nu} \nabla F(x) \nabla F(x)^T. \tag{1.3.24}$$

Давайте проверим, какие из самосогласованных функций примера 1.3.1 являются также и самосогласованными барьерами.

Пример 1.3.3 1. *Линейная функция*: $f(x) = \alpha + \langle a, x\rangle$, $\mathrm{dom}\, f = \mathbb{R}^n$. Ясно, что при $a \neq 0$ эта функция не будет самосогласованным барьером, поскольку $\nabla^2 f(x) = 0$.

2. *Выпуклая квадратичная функция*. Пусть $A = A^T \succ 0$. Рассмотрим функцию

$$f(x) = \alpha + \langle a, x\rangle + \tfrac{1}{2}\langle Ax, x\rangle, \quad \mathrm{dom}\, f = \mathbb{R}^n.$$

Тогда $\nabla f(x) = a + Ax$ и $\nabla^2 f(x) = A$. Таким образом,

$$\begin{aligned} \langle [f(x)]^{-1}\nabla f(x), \nabla f(x)\rangle &= \langle A^{-1}(Ax - a), Ax - a\rangle \\ &= \langle Ax, x\rangle - 2\langle a, x\rangle + \langle A^{-1}a, a\rangle. \end{aligned}$$

Ясно, что это выражение неограничено сверху на $\mathbb{R}^n$. Таким образом, квадратичная функция не является самосогласованным барьером.

3. *Логарифмический барьер для луча*. Рассмотрим следующую функцию одной переменной:

$$F(x) = -\ln x, \quad \mathrm{dom}\, F = \{x \in \mathbb{R} \mid \ x > 0\}.$$

Тогда $\nabla F(x) = -\frac{1}{x}$ и $\nabla^2 F(x) = \frac{1}{x^2} > 0$. Таким образом,

$$\tfrac{(\nabla F(x))^2}{\nabla^2 F(x)} = \tfrac{1}{x^2} \cdot x^2 = 1.$$

Следовательно, функция $F(x)$ является ν-самосогласованным барьером для множества $\{x > 0\}$ с параметром $\nu = 1$.

4. *Логарифмический барьер для выпуклой области второго порядка*. Пусть $A = A^T \succeq 0$. Рассмотрим *вогнутую* квадратичную функцию

$$\phi(x) = \alpha + \langle a, x\rangle - \tfrac{1}{2}\langle Ax, x\rangle.$$

Положим $F(x) = -\ln \phi(x)$, $\mathrm{dom}\, f = \{x \in \mathbb{R}^n \mid \ \phi(x) > 0\}$. Тогда

$$\begin{aligned} \langle \nabla F(x), u\rangle = &\ -\tfrac{1}{\phi(x)}[\langle a, u\rangle - \langle Ax, u\rangle], \\ \langle \nabla^2 F(x)u, u\rangle = &\ \tfrac{1}{\phi^2(x)}[\langle a, u\rangle - \langle Ax, u\rangle]^2 + \tfrac{1}{\phi(x)}\langle Au, u\rangle. \end{aligned}$$

Обозначим $\omega_1 = \langle \nabla F(x), u\rangle$ и $\omega_2 = \frac{1}{\phi(x)}\langle Au, u\rangle$. Тогда

$$\langle \nabla^2 F(x)u, u\rangle = \omega_1^2 + \omega_2 \geq \omega_1^2.$$

Таким образом, $2\langle \nabla F(x), u\rangle - \langle \nabla^2 F(x)u, u\rangle \leq 2\omega_1 - \omega_1^2 \leq 1$. Следовательно, $F(x)$ является ν-самосогласованным барьером с параметром $\nu = 1$. □

Приведем несколько простейших свойств самосогласованных барьеров.

Теорема 1.3.15 *Пусть функция $F(x)$ является самосогласованным барьером. Тогда функция $\langle c, x\rangle + F(x)$ самосогласованна на множестве* $\operatorname{dom} F$.

Доказательство.
Так как $F(x)$ - самосогласованная функция, можно воспользоваться Следствием 1.3.1. □

Вышеупомянутое свойство является важным для методов отслеживания траектории.

Теорема 1.3.16 *Пусть функции F_i являются ν_i-самосогласованными барьерами, $i = 1, 2$. Тогда функция*

$$F(x) = F_1(x) + F_2(x)$$

является самосогласованным барьером для выпуклого множества $\mathrm{dom\,F = dom\,F_1 \bigcap dom\,F_2}$ *с параметром $\nu = \nu_1 + \nu_2$.*

Доказательство.
В силу теоремы 1.3.1 функция F будет самосогласованной. Зафиксируем произвольную точку $x \in \operatorname{dom} F$. Тогда

$$\begin{aligned}
& \max_{u \in \mathbb{R}^n} \left[2\langle \nabla F(x)u, u\rangle - \langle \nabla^2 F(x)u, u\rangle\right] \\
= \ & \max_{u \in \mathbb{R}^n} \left[2\langle \nabla F_1(x)u, u\rangle - \langle \nabla^2 F_1(x)u, u\rangle + 2\langle \nabla F_1(x)u, u\rangle - \langle \nabla^2 F_1(x)u, u\rangle\right] \\
\leq \ & \max_{u \in \mathbb{R}^n} \left[2\langle \nabla F_1(x)u, u\rangle - \langle \nabla^2 F_1(x)u, u\rangle\right] \\
& + \max_{u \in \mathbb{R}^n} \left[2\langle \nabla F_2(x)u, u\rangle - \langle \nabla^2 F_2(x)u, u\rangle\right] \leq \nu_1 + \nu_2. \qquad \Box
\end{aligned}$$

Наконец, покажем, что значение параметра самосогласованного барьера является аффинно-инвариантной характеристикой.

Теорема 1.3.17 *Рассмотрим линейный оператор $\mathcal{A}(x) = Ax + b$, $\mathcal{A}(x) : \mathbb{R}^n \to \mathbb{R}^m$. Пусть функция $F(y)$ является ν-самосогласованным барьером. Тогда функция $\Phi(x) = F(\mathcal{A}(x))$ также является ν-самосогласованным барьером, но для множества*

$$\operatorname{dom} \Phi = \{\mathrm{x} \in \mathbb{R}^{\mathrm{n}} \mid \ \mathcal{A}(\mathrm{x}) \in \mathrm{dom\,F}\}.$$

Доказательство.
Функция $\Phi(x)$ является стандартной самосогласованной функцией в силу теоремы 1.3.2. Зафиксируем произвольную точку $x \in \operatorname{dom} \Phi$. Тогда $y = \mathcal{A}(x) \in \operatorname{dom} F$. При всех $u \in \mathbb{R}^n$ имеем

$$\langle \Phi'(x), u\rangle = \langle \nabla F(y), Au\rangle, \quad \langle \Phi''(x)u, u\rangle = \langle \nabla^2 F(y)Au, Au\rangle.$$

Таким образом,

$$\max_{u\in\mathbb{R}^n}\left[2\langle\Phi'(x),u\rangle-\langle\Phi''(x)u,u\rangle\right]$$

$$=\max_{u\in\mathbb{R}^n}\left[2\langle\nabla F(y),Au\rangle-\langle\nabla^2F(y)Au,Au\rangle\right]$$

$$\le\max_{v\in\mathbb{R}^m}\left[2\langle\nabla F(y),v\rangle-\langle\nabla^2F(y)v,v\rangle\right]\le\nu. \qquad \square$$

Основные неравенства

Покажем, что локальные характеристики самосогласованного барьера (градиент и гессиан) обеспечивают нас *глобальной* информацией о структуре его области определения.

Теорема 1.3.18 *1. Пусть функция $F(x)$ является ν-самосогласованным барьером. Тогда для любых $x, yin\mathrm{dom}\,F$ выполняется неравенство*

$$\langle\nabla F(x),y-x\rangle<\nu. \tag{1.3.25}$$

Более того, если $\langle\nabla F(x),y-x\rangle\ge0$, то

$$\langle\nabla F(y)-\nabla F(x),y-x\rangle\ge\tfrac{\langle\nabla F(x),y-x\rangle^2}{\nu-\langle\nabla F(x),y-x\rangle}. \tag{1.3.26}$$

2. Стандартная самосогласованная функция $F(x)$ является ν-самосогласованным барьером тогда и только тогда, когда

$$F(y)\ge F(x)-\nu\ln\left(1-\tfrac{1}{\nu}\langle\nabla F(x),y-x\rangle\right)\quad\forall x,y\in\mathrm{dom}\,F. \tag{1.3.27}$$

Доказательство.
1. Пусть $x,y\in\mathrm{dom}\,F$. Рассмотрим функцию

$$\phi(t)=\langle\nabla F(x+t(y-x)),y-x\rangle,\quad t\in[0,1].$$

Если $\phi(0)\le0$, то неравенство (1.3.25) тривиально. Если $\phi(0)=0$, то неравенство (1.3.26) тоже тривиально. Предположим, что $\phi(0)>0$. В силу определения (1.3.23) имеем

$$\begin{aligned}\phi'(t)&=\langle\nabla^2F(x+t(y-x))(y-x),y-x\rangle\\ &\ge\tfrac{1}{\nu}\langle\nabla F(x+t(y-x)),y-x\rangle^2=\tfrac{1}{\nu}\phi^2(t).\end{aligned}$$

Таким образом, функция $\phi(t)$ возрастает и является положительной при $t\in[0,1]$. Более того, для любых $t\in[0,1]$ имеем

$$-\tfrac{1}{\phi(t)}+\tfrac{1}{\phi(0)}\ge\tfrac{1}{\nu}t.$$

Это означает, что $\langle \nabla F(x), y-x\rangle = \phi(0) < \frac{\nu}{t}$ при всех $t \in [0,1]$. Таким образом, неравенство (1.3.25) доказано. Более того,

$$\phi(t) - \phi(0) \geq \tfrac{\nu\phi(0)}{\nu - t\phi(0)} - \phi(0) = \tfrac{t\phi(0)^2}{\nu - t\phi(0)}, \quad t \in [0,1].$$

Выбирая $t = 1$, получаем неравенство (1.3.26).

2. Обозначим $\psi(x) = e^{-\frac{1}{\nu}F(x)}$. Тогда

$$\begin{array}{rcl} \psi'(x) & = & -\frac{1}{\nu}e^{-\frac{1}{\nu}F(x)} \cdot \nabla F(x), \\ \\ \psi''(x) & = & -\frac{1}{\nu}e^{-\frac{1}{\nu}F(x)} \left[\nabla^2 F(x) - \frac{1}{\nu}\nabla F(x) \nabla F(x)^T\right]. \end{array}$$

Таким образом, в силу определения (1.3.24) функция $\psi(x)$ является вогнутой в том и только том случае, когда функция $F(x)$ является ν-самосогласованным барьером. Осталось заметить что неравенство (1.3.27) эквивалентно неравенству

$$\psi(y) \leq \psi(x) + \langle \psi'(x), y - x\rangle$$

с точностью до взятия логарифма от обеих его частей. □

Докажем теперь *теорему о рецессивном направлении.*

Теорема 1.3.19 *Пусть функция F является ν-самосогласованным барьером, а направление h явялется рецессивным для области* $\operatorname{dom} F$. *Тогда для любых $x \in \operatorname{dom} F$ выполнено неравенство*

$$\langle \nabla^2 F(x)h, h\rangle \;\; \leq \;\; \langle \nabla F(x), h\rangle^2. \tag{1.3.28}$$

Доказательство.

Единственный нетривиальный случай есть $\langle \nabla^2 F(x)h, h\rangle > 0$. Рассмотрим функцию $\phi(t) = F(x + th)$. Это самосогласованный барьер. Таким образом, в силу теоремы 1.3.11 условие $\lambda_\phi^2(1) = \frac{\langle \nabla F(x), h\rangle^2}{\langle \nabla^2 F(x)h, h\rangle} < 1$ означает, что функция $\phi(t)$ достигает своего минимума. Однако в силу неравенства (1.3.25) $\langle \nabla F(u), h\rangle \leq 0$ при любом $u \in \operatorname{dom} F$. Таким образом, если минимум функции ϕ достигается при некотором $\bar{t} > 0$, то $\phi(t) \equiv \text{const}$ при всех $t \geq \bar{t}$. Но для самосогласованного барьера это невозможно (см., например, неравенство (1.3.5)). □

Следствие 1.3.5 *Пусть $x \in \operatorname{dom} F$ и $y \in \operatorname{Cl}(\operatorname{dom} F)$. Тогда выполняется неравенство*

$$F(x + \alpha(y - x)) \;\; \leq \;\; F(x) - \nu \ln(1 - \alpha) \quad \forall \alpha \in [0, 1). \tag{1.3.29}$$

Доказательство.

Обозначим $x(\alpha) = x + \alpha(y - x)$ и $\phi(\alpha) = F(x(\alpha))$. Тогда

$$\phi'(\tau) \;\; = \;\; \tfrac{1}{1-\tau}\langle \nabla F(x(\tau)), y - x(\tau)\rangle \overset{(1.3.25)}{\leq} \tfrac{\nu}{1-\tau}.$$

Остается проинтегрировать это неравенство по $\tau \in [0, \alpha]$. □

Теорема 1.3.20 *Пусть функция $F(x)$ является ν-самосогласованным барьером. Тогда для всех таких $x \in \operatorname{dom} F$ и $y \in \operatorname{dom} \mathrm{F}$, что*

$$\langle \nabla F(x), y - x \rangle \geq 0, \tag{1.3.30}$$

выполнено неравенство

$$\| y - x \|_x \leq \nu + 2\sqrt{\nu}. \tag{1.3.31}$$

Доказательство.
Обозначим $r = \| y - x \|_x$. Пусть $r > \sqrt{\nu}$. Рассмотрим точку $y_\alpha = x + \alpha(y-x)$, где $\alpha = \frac{\sqrt{\nu}}{r} < 1$. В силу предположения (1.3.30) и неравенства (1.3.5) имеем

$$\begin{aligned} \omega \equiv \langle \nabla F(y_\alpha), y - x \rangle &\geq \langle \nabla F(y_\alpha) - \nabla F(x), y - x \rangle \\ &= \tfrac{1}{\alpha} \langle \nabla F(y_\alpha) - \nabla F(x), y_\alpha - x \rangle \\ &\geq \tfrac{1}{\alpha} \cdot \tfrac{\|y_\alpha - x\|_x^2}{1 + \|y_\alpha - x\|_x^2} = \tfrac{\alpha \|y-x\|_x^2}{1+\alpha\|y-x\|_x} = \tfrac{r\sqrt{\nu}}{1+\sqrt{\nu}}. \end{aligned}$$

С другой стороны, в силу неравенства (1.3.25) получаем

$$(1-\alpha)\omega = \langle \nabla F(y_\alpha), y - y_\alpha \rangle \leq \nu.$$

Таким образом,

$$\left(1 - \tfrac{\sqrt{\nu}}{r}\right) \tfrac{r\sqrt{\nu}}{1+\sqrt{\nu}} \leq \nu,$$

а это и есть неравенство (1.3.31). □

В завершение этого пункта, изучим свойства одной важной точки, связанной с областью определения самосогласованного барьера.

Определение 1.3.4 Пусть функция $F(x)$ является ν-самосогласованным барьером для множества $\operatorname{dom} \mathrm{F}$. Точка

$$x_F^* = \arg \min_{x \in \operatorname{dom} F} F(x),$$

называется *аналитическим центром* выпуклого множества $\operatorname{dom} \mathrm{F}$, порождаемым барьером $F(x)$.

Теорема 1.3.21 *Предположим, что аналитический центр ν-самосогласованного барьера $F(x)$ существует. Тогда для любого $x \in \operatorname{dom} \mathrm{F}$ выполнено неравенство*

$$\| x - x_F^* \|_{x_F^*} \leq \nu + 2\sqrt{\nu}.$$

С другой стороны, для любого такого $x \in \mathbb{R}^n$, что $\| x - x_F^* \|_{x_F^*} \leq 1$, верно включение $x \in \operatorname{dom} \mathrm{F}$.

Доказательство.

Первое утверждение следует из теоремы 1.3.20 ввиду равенства $\nabla F(x_F^*) = 0$. Второе утверждение следует из теоремы 1.3.5. $\square$

Таким образом, *асферичность* множества $\mathrm{dom\,F}$ относительно точки x_F^*, измеренная в метрике $\| \cdot \|_{x_F^*}$, не может быть больше $\nu + 2\sqrt{\nu}$. Хорошо известно, что для любого выпуклого множества в $\mathbb{R}^n$ существует метрика, в которой его асферичность не превосходит n (теорема Джона). Однако нам удалось оценить асферичность в терминах *параметра барьера*. Это число не связано непосредственно с размерностью пространства.

Заметим также что для множества $\mathrm{dom\,F}$, не содержащего прямых, существование точки x_F^* означает его ограниченность (поскольку тогда гессиан $\nabla^2 F(x_F^*)$ невырожден, см. теорему 1.3.3).

Следствие 1.3.6 *Пусть множество* $\mathrm{dom\,F}$ *ограничено. Тогда для любых* $x \in \mathrm{dom}\,F$ *и* $v \in \mathbb{R}^n$ *выполняется неравенство*

$$\| v \|_x^* \leq (\nu + 2\sqrt{\nu}) \| v \|_{x_F^*}^* .$$

Доказательство.

Воспользуемся следующим представлением:

$$\| v \|_x^* \equiv \langle [\nabla^2 F(x)]^{-1} v, v\rangle^{1/2} = \max\{\langle v, u\rangle \mid \ \langle \nabla^2 F(x) u, u\rangle \leq 1\}.$$

С другой стороны, в силу теорем 1.3.5 и 1.3.21 имеем

$$\begin{aligned} B \ &\equiv \{y \in \mathbb{R}^n \mid \| y - x \|_x \leq 1\} \subseteq \mathrm{dom\,F} \\ &\subseteq \{y \in \mathbb{R}^n \mid \| y - x_F^* \|_x \leq \nu + 2\sqrt{\nu}\} \equiv B_*. \end{aligned}$$

Таким образом, воспользовавшись опять теоремой 1.3.21, получаем

$$\begin{aligned} \| v \|_x^* \ &= \max\{\langle v, y - x\rangle \mid \ y \in B\} \leq \max\{\langle v, y - x\rangle \mid \ y \in B_*\} \\ &= \langle v, x_F^* - x\rangle + (\nu + 2\sqrt{\nu}) \| v \|_{x_F^*}^* . \end{aligned}$$

Заметим, что $\| v \|_x^* = \| -v \|_x^*$. Таким образом, мы можем считать, что $\langle v, x_F^* - x\rangle \leq 0$. $\square$

Метод отслеживания траектории

Мы готовы описать *барьерную модель* нашей задачи минимизации. Это *стандартная* задача минимизации

$$\min\{\langle c, x\rangle \mid \ x \in Q\}, \tag{1.3.32}$$

у которой *ограниченное* выпуклое замкнутое допустимое множество $Q \equiv \mathrm{dom\,F}$ имеет непустую внутренность и снабжено ν-самосогласованным барьером $F(x)$.

Напомним, что мы собираемся решать задачу (1.3.32) с помощью отслеживания *центральной траектории*:

$$x^*(t) = \arg\min_{x \in \mathrm{dom}\, F} f(t; x), \tag{1.3.33}$$

где $f(t; x) = t\langle c, x\rangle + F(x)$ и $t \geq 0$. Ввиду условия оптимальности первого порядка любая точка этой траектории удовлетворяет уравнению

$$tc + \nabla F(x^*(t)) = 0. \tag{1.3.34}$$

Поскольку множество Q ограничено, его *аналитический центр* x_F^* существует и лежит на центральной траектории:

$$x^*(0) = x_F^*. \tag{1.3.35}$$

Мы будем отслеживать центральную траекторию, пересчитывая точки, удовлетворяющие *приближенному условию центрирования*:

$$\lambda_{f(t;\cdot)}(x) \equiv \| \nabla f(t; x) \|_x^* = \| tc + \nabla F(x) \|_x^* \leq \beta, \tag{1.3.36}$$

где *параметр центрирования* β достаточно мал.

Покажем, что такая стратегия приводит нас к цели.

Теорема 1.3.22 *Для любого* $t > 0$ *выполняется неравенство*

$$\langle c, x^*(t)\rangle - c^* \leq \tfrac{\nu}{t}, \tag{1.3.37}$$

где c^* *- оптимальное значение задачи (1.3.32). Если точка* x *удовлетворяет условию центрироания (1.3.36), то*

$$\langle c, x\rangle - c^* \leq \tfrac{1}{t}\left(\nu + \tfrac{(\beta+\sqrt{\nu})\beta}{1-\beta}\right). \tag{1.3.38}$$

Доказательство.

Пусть x^* является решением задачи (1.3.32). В силу неравенств (1.3.34) и (1.3.25) имеем

$$\langle c, x^*(t) - x^*\rangle = \tfrac{1}{t}\langle \nabla F(x^*(t)), x^* - x^*(t)\rangle \leq \tfrac{\nu}{t}.$$

Далее, пусть x удовлетворяет условию (1.3.36). Обозначим $\lambda = \lambda_{f(t;\cdot)}(x)$. Тогда

$$\begin{aligned} t\langle c, x - x^*(t)\rangle &= \langle \nabla f(t; x) - \nabla F(x), x - x^*(t)\rangle \\ &\leq (\lambda + \sqrt{\nu}) \| x - x^*(t) \|_x \\ &\leq (\lambda + \sqrt{\nu})\tfrac{\lambda}{1-\lambda} \leq \tfrac{(\beta+\sqrt{\nu})\beta}{1-\beta} \end{aligned}$$

в силу неравенства (1.3.22), теоремы 1.3.13 и неравенства (1.3.36). □

Проведем анализ одного шага метода отслеживания траектории. Пусть $x \in \operatorname{dom} F$. Рассмотрим следующую итерацию:

$$\boxed{\begin{aligned} t_+ &= t + \tfrac{\gamma}{\|c\|^*_x}, \\ x_+ &= x - [\nabla^2 F(x)]^{-1}(t_+ c + \nabla F(x)). \end{aligned}} \tag{1.3.39}$$

Теорема 1.3.23 *Если точка x удовлетворяет условию (1.3.36):*

$$\| \, tc + \nabla F(x) \, \|^*_x \leq \beta,$$

где $\beta < \bar{\lambda} = \frac{3-\sqrt{5}}{2}$, то для множителя γ, удовлеворяющего условию

$$| \, \gamma \, | \leq \tfrac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta, \tag{1.3.40}$$

выполняется неравенство $\| \, t_+ c + \nabla F(x_+) \, \|^*_x \leq \beta$.

Доказательство.
Обозначим $\lambda_0 = \| \, tc + \nabla F(x) \, \|^*_x \leq \beta$, $\lambda_1 = \| \, t_+ c + \nabla F(x) \, \|^*_x$ и $\lambda_+ = \| \, t_+ c + \nabla F(x_+) \, \|^*_{x_+}$. Тогда

$$\lambda_1 \leq \lambda_0 + | \, \gamma \, | \leq \beta + | \, \gamma \, |$$

и в силу теоремы 1.3.14 имеем

$$\lambda_+ \leq \left(\tfrac{\lambda_1}{1-\lambda_1} \right)^2 \equiv [\omega'_*(\lambda_1)]^2.$$

Осталось заметить, что условие (1.3.40) эквивалентно следующему:

$$\omega'_*(\beta + | \, \gamma \, |) \leq \sqrt{\beta}$$

(напомним, что $\omega'(\omega'_*(\tau)) = \tau$, см. лемму 1.3.3). □

Покажем, что изменение параметра t в схеме (1.3.39) будет достаточно большим.

Лемма 1.3.4 *Пусть x удовлетворяет условию (1.3.36). Тогда*

$$\| \, c \, \|^*_x \leq \tfrac{1}{t}(\beta + \sqrt{\nu}). \tag{1.3.41}$$

Доказательство.
Действительно, в силу неравенств (1.3.36) и (1.3.22) имеем

$$t \, \| \, c \, \|^*_x = \| \, \nabla f(t;x) - \nabla F(x) \, \|^*_x \leq \| \, \nabla f(t;x) \, \|^*_x + \| \, \nabla F(x) \, \|^*_x \leq \beta + \sqrt{\nu}.$$ □

Выпишем подходящие значения параметров в схеме (1.3.39). В оставшейся части этого пункта мы предполагаем, что

$$\beta = \tfrac{1}{9}, \quad \gamma = \tfrac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta = \tfrac{5}{36}. \tag{1.3.42}$$

Мы показали, что центральная траектория может отслеживаться с помощью правил (1.3.39). При этом возможно как увеличение, так и уменьшение параметра t. Нижняя оценка на скорость *увеличения* параметра t дается неравенством

$$t_+ \geq \left(1 + \tfrac{5}{4+36\sqrt{\nu}}\right) \cdot t,$$

а верхняя оценка на скорость *уменьшения* параметра такова:

$$t_+ \leq \left(1 - \tfrac{5}{4+36\sqrt{\nu}}\right) \cdot t.$$

Итак, общая схема решения задачи (1.3.32) выглядит следующим образом.

Основная схема отслеживания траектории

0. Положим $t_0 = 0$. Выберем такую точность $\epsilon > 0$ и точку $x_0 \in \mathrm{dom}\, F$, что

$$\| \nabla F(x_0) \|^*_{x_0} \leq \beta.$$

1. k-я итерация ($k \geq 0$). Положим

$$\begin{aligned} t_{k+1} &= t_k + \tfrac{\gamma}{\|c\|^*_{x_k}}, \\ x_{k+1} &= x_k - [\nabla^2 F(x_k)]^{-1}(t_{k+1}c + \nabla F(x_k)). \end{aligned}$$

2. Останавливаемся, если $\epsilon\, t_k \geq \nu + \frac{(\beta+\sqrt{\nu})\beta}{1-\beta}$.

(1.3.43)

Оценим трудоемкость вышеприведенной схемы.

Теорема 1.3.24 *Схема (1.3.43) останавливается не более чем через N шагов, где*

$$N \leq O\left(\sqrt{\nu} \ln \frac{\nu \|c\|^*_{x^*_F}}{\epsilon}\right).$$

Более того, в момент остановки выполнено неравенство $\langle c, x_N \rangle - c^* \leq \epsilon$.

Доказательство.
Заметим, что $r_0 \equiv \| x_0 - x_F^* \|_{x_0} \le \frac{\beta}{1-\beta}$ (см. теорему 1.3.13). Таким образом, в силу теоремы 1.3.6 имеем

$$\tfrac{\gamma}{t_1} = \| c \|_{x_0}^* \le \tfrac{1}{1-r_0} \| c \|_{x_F^*}^* \le \tfrac{1-\beta}{1-2\beta} \| c \|_{x_F^*}^* .$$

Следовательно, $t_k \ge \frac{\gamma(1-2\beta)}{(1-\beta)\|c\|_{x_F^*}^*} \left(1 + \frac{\gamma}{\beta+\sqrt{\nu}}\right)^{k-1}$ для всех $k \ge 1$. □

Обсудим вышеприведенную оценку сложности. Ее основным членом является

$$7,2\sqrt{\nu} \ln \frac{\nu\|c\|_{x_F^*}^*}{\epsilon}.$$

Заметим, что число $\nu \| c \|_{x_F^*}^*$ оценивает сверху изменение линейной функции на множестве $\mathrm{dom\,F}$ (см. теорему 1.3.21). Таким образом, отношение

$$\frac{\epsilon}{\nu\|c\|_{x_F^*}^*}$$

может интерпретироваться как *относительная точность* решения.

В процессе (1.3.43) есть один серьезный недостаток. Иногда бывает трудно обеспечить выполнение начального условия

$$\| \nabla F(x_0) \|_{x_0}^* \le \beta.$$

В таком случае приходится прибегать к вспомогательному процессу *вычисления* подходящей стартовой точки. Соответствующая стратегия анализируется в следующем пункте.

Нахождение аналитического центра

Таким образом, нам необходимо найти приближение к *аналитическому центру* множества $\mathrm{dom\,F}$. Рассмотрим следующую задачу минимизации:

$$\min\{F(x) \mid \ x \in \mathrm{dom}\, F\}, \qquad (1.3.44)$$

где функция F является ν-самосогласованным барьером. Для запуска метода из предыдущего пункта, нам необходимо найти приближенное решение $\bar{x} \in \mathrm{dom}\, F$ этой задачи, удовлетворяющее условию

$$\| \nabla F(\bar{x}) \|_{\bar{x}}^* \le \beta$$

при некотором $\beta \in (0, 1)$.

Для этого можно применить две различные стратегии. Первая из них – это обычный редуцированный метод Ньютона. Вторая стратегия основана на отслеживании вспомогательной траектории.

Рассмотрим сначала первую стратегию.

Редуцированый метод Ньютона для аналитического центра

0. Выбираем $y_0 \in \operatorname{dom} F$.

1. k-я итерация ($k \geq 0$). Полагаем

$$y_{k+1} = y_k - \frac{[\nabla^2 F(y_k)]^{-1} \nabla F(y_k)}{1 + \| \nabla F(y_k) \|^*_{y_k}}.$$

2. Останавливаемся, когда $\| \nabla F(y_k) \|^*_{y_k} \leq \beta$.

(1.3.45)

Теорема 1.3.25 *Процесс (1.3.45) прерывается не позднее чем через* $\frac{1}{\omega(\beta)}(F(y_0) - F(x^*_F))$ *итераций.*

Доказательство.

Действительно, в силу теоремы 1.3.12 имеем

$$F(y_{k+1}) \leq F(y_k) - \omega(\lambda_F(y_k)) \leq F(y_k) - \omega(\beta).$$

Таким образом, $F(y_0) - k\,\omega(\beta) \geq F(y_k) \geq F(x^*_F)$. □

Схема метода отслеживания траектории немного сложнее. Пусть у нас есть точка $y_0 \in \operatorname{dom} F$. Определим *вспомогательную центральную траекторию* следующим образом:

$$y^*(t) = \arg \min_{y \in \operatorname{dom} F} [-t \langle \nabla F(y_0), y \rangle + F(y)],$$

где $t \geq 0$. Эта траектория удовлетворяет уравнению

$$\nabla F(y^*(t)) = t \nabla F(y_0). \qquad (1.3.46)$$

Таким образом, она связывает две точки: стартовую точку y_0 и аналитический центр x^*_F:

$$y^*(1) = y_0, \quad y^*(0) = x^*_F.$$

Эту траекторию можно отслеживать с помощью процесса (1.3.39) с *уменьшающимся* параметром t.

Оценим скорость приближения вспомогательной центральной траектории $y^*(t)$ к аналитическому центру.

Лемма 1.3.5 *При всех* $t \geq 0$ *выполняется неравенство*

$$\| \nabla F(y^*(t)) \|^*_{y^*(t)} \leq (\nu + 2\sqrt{\nu}) \| \nabla F(x_0) \|^*_{x^*_F} \cdot t.$$

Доказательство.

Эта оценка вытекает из формулы (1.3.46) и следствия 1.3.6. □

Выпишем соответствующую итеративную процедуру.

Отслеживание вспомогательной центральной траектории

0. Выбираем $y_0 \in \mathrm{dom\,F}$. Положим $t_0 = 1$.

1. k-я итерация ($k \geq 0$). Положим

$$\begin{aligned} t_{k+1} &= t_k - \tfrac{\gamma}{\|\nabla F(y_0)\|^*_{y_k}}, \\ y_{k+1} &= y_k - [\nabla^2 F(y_k)]^{-1}(t_{k+1}\nabla F(y_0) + \nabla F(y_k)). \end{aligned} \tag{1.3.47}$$

2. Останавливаемся, если $\| \nabla F(y_k) \|_{y_k} \leq \frac{\sqrt{\beta}}{1+\sqrt{\beta}}$.
Положим $\bar{x} = y_k - [\nabla^2 F(y_k)]^{-1}\nabla F(y_k)$.

Эта схема отслеживает вспомогательную траекторию $y^*(t)$ при $t_k \to 0$. Она пересчитывает точки $\{y_k\}$, удовлетворяющие приближенному условию центрирования

$$\| t_k \nabla F(y_0) + \nabla F(y_k) \|_{y_k} \leq \beta.$$

Критерий остановки этого процесса:

$$\lambda_k = \| \nabla F(y_k) \|_{y_k} \leq \tfrac{\sqrt{\beta}}{1+\sqrt{\beta}}.$$

Он гарантирует, что $\| \nabla F(\bar{x}) \|_{\bar{x}} \leq \left(\frac{\lambda_k}{1-\lambda_k}\right)^2 \leq \beta$ (см. теорему 1.3.14).

Выпишем оценки сложности этого процесса.

Теорема 1.3.26 *Процесс (1.3.47) останавливается не более чем через*

$$\tfrac{1}{\gamma}(\beta + \sqrt{\nu}) \ln \left[\tfrac{1}{\gamma}(\nu + 2\sqrt{\nu}) \parallel \nabla F(x_0) \parallel^*_{x^*_F} \right]$$

итераций.

Доказательство.

Напомним, что мы зафиксировали следующие значения параметров:

$$\beta = \tfrac{1}{9}, \quad \gamma = \tfrac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta = \tfrac{5}{36}.$$

Мы выбрали $t_0 = 1$. Таким образом, в силу теоремы 1.3.23 и леммы 1.3.4 имеем

$$t_{k+1} \leq \left(1 - \tfrac{\gamma}{\beta+\sqrt{\nu}}\right) t_k \leq \exp\left(-\tfrac{\gamma(k+1)}{\beta+\sqrt{\nu}}\right).$$

Далее, в силу леммы 1.3.5 получаем

$$\begin{aligned} \| \nabla F(y_k) \|^*_{y_k} &= \| (t_k \nabla F(x_0) + \nabla F(y_k)) - t_k \nabla F(x_0) \|^*_{y_k} \\ &\leq \beta + t_k \| \nabla F(x_0) \|^*_{y_k} \\ &\leq \beta + t_k(\nu + 2\sqrt{\nu}) \| \nabla F(x_0) \|^*_{x^*_F}. \end{aligned}$$

Таким образом, процесс прервется по крайней мере после выполнения следующего неравенства:

$$t_k(\nu + 2\sqrt{\nu}) \| \nabla F(x_0) \|^*_{x^*_F} \leq \tfrac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta = \gamma. \qquad \Box$$

Теперь можно обсудить сложность обеих стратегий. Основной член в оценке сложности вспомогательной схемы отслеживания траектории выглядит так:

$$7,2\sqrt{\nu}[\ln \nu + \ln \| \nabla F(x_0) \|^*_{x^*_F}].$$

Для вспомогательного редуцированного метода Ньютона это $O(F(y_0) - F(x^*_F))$. Мы не можем напрямую сравнить эти оценки. Однако более точный анализ показывает преимущество метода отслеживания траектории. Также важно, что его оценка сложности подобна оценке сложности основного метода отслеживания траектории. Действительно, если применить схему (1.3.43) вместе с (1.3.47), то мы получим следующую оценку для всего процесса:

$$7,2\sqrt{\nu}\left[2\ln \nu + \ln \| \nabla F(x_0) \|^*_{x^*_F} + \ln \| c \|^*_{x^*_F} + \ln \tfrac{1}{\epsilon}\right].$$

Логарифмически однородные барьеры

Очень часто множество стандартной задачи минимизации представляется как пересечение аффинного многообразия с выпуклым конусом. Такие *конические* задачи стимулировали интерес к специальным самосогласованным барьерам для выпуклых конусов.

Определение 1.3.5 Пусть выпуклый замкнутый конус K имеет непустую внутренность. Выпуклая барьерная функция $F(x)$, $\operatorname{dom} F = \operatorname{int} K$, называется *логарифмически однородной*, если для любых $x \in \operatorname{int} K$ и $\tau > 0$ выполнено тождество

$$F(\tau x) = F(x) - \nu \ln \tau. \tag{1.3.48}$$

Контанта $\nu > 0$ называется *параметром* этого барьера.

Тождество (1.3.48) приводит ко многим интересным следствиям. Дифференцируя его по x, получаем

$$\nabla F(\tau x) \;=\; \tfrac{1}{\tau}\nabla F(x), \quad \nabla^2 F(\tau x) \;=\; \tfrac{1}{\tau^2}\nabla^2 F(x), \quad \tau > 0. \tag{1.3.49}$$

Дифференцируя теперь первое из равенств (1.3.49) по τ в точке $\tau = 1$, получаем

$$\nabla^2 F(x)x \;=\; -\nabla F(x). \tag{1.3.50}$$

Более того, дифференцируя равенство (1.3.48) по τ в точке $\tau = 1$, получаем

$$\langle \nabla F(x), x\rangle \;=\; -\nu. \tag{1.3.51}$$

Это равенство имеет два важных следствия. Предположим, что конус K точечный (не содержит прямых). Тогда в силу теоремы 1.3.3 гессиан $\nabla^2 F(x)$ невырожден. Таким образом,

$$\langle \nabla F(x), [\nabla^2 F(x)]^{-1}\nabla F(x)\rangle \;\overset{(1.3.50)}{=}\; \nu, \quad x \in \operatorname{int} K. \tag{1.3.52}$$

Следовательно, ν является параметром барьера F. В то же время это неравенство может быть записано как

$$\langle \nabla^2 F(x)x, x\rangle \;=\; \nu, \quad x \in \operatorname{int} K. \tag{1.3.53}$$

Этот результат приводит к полезному следствию.

Лемма 1.3.6 *Пусть* $x_0 \in \operatorname{int} K$ *и* $\bar{x} \in \{x \in K : \langle \nabla F(x_0), x - x_0\rangle \geq 0\}$. *Тогда*

$$\|x - x_0\|^2_{x_0} \;\leq\; \nu. \tag{1.3.54}$$

Доказательство.
Действительно,

$$\begin{aligned}
\|x - x_0\|^2_{x_0} \;&=\; \langle \nabla^2 F(x_0)(x - x_0), x - x_0\rangle \\
&\overset{(1.3.50)}{=}\; \|x\|^2_{x_0} - 2\langle \nabla F(x_0), x\rangle + \nu \\
&\overset{(1.3.28)}{\leq}\; \langle \nabla F(x_0), x\rangle^2 - 2\langle \nabla F(x_0), x\rangle + \nu.
\end{aligned}$$

Осталось заметить, что $0 \geq \langle \nabla F(x_0), x\rangle \geq \langle \nabla F(x_0), x_0\rangle \overset{(1.3.51)}{=} -\nu$. □

1.3.4 Конструирование самосогласованных барьеров

Границы на параметр

В предыдущем пункте мы рассмотрели метод отслеживания траектории для задачи

$$\min_{x\in Q} \langle c, x\rangle, \tag{1.3.55}$$

где Q – выпуклое замкнутое множество с непустой внутренностью, для кторого известен ν-самосогласованный барьер $F(x)$. С помощью этого барьера мы можем решить задачу (1.3.55) за $O\left(\sqrt{\nu}\cdot\ln\frac{\nu}{\epsilon}\right)$ итераций. Напомним, что наиболее сложная часть каждой итерации состоит в решении системы линейных уравнений.

В этом пункте мы изучим пределы применимости данного подхода. Мы получим нижние и верхние оценки для параметра самосогласованного барьера. Мы также обсудим классы выпуклых задач, для которых постановка (1.3.55) может быть обеспечена вычислимыми барьерами.

Начнем с нижней оценки на параметр барьера.

Лемма 1.3.7 *Пусть функция $f(t)$ является ν-самосогласованным барьером для интервала $(\alpha,\beta)\subset\mathbb{R}$, $\alpha<\beta<\infty$. Тогда*

$$\nu \geq \kappa \equiv \sup_{t\in(\alpha,\beta)} \frac{(\nabla f(t))^2}{\nabla^2 f(t)} \geq 1.$$

Доказательство.
Заметим, что $\nu\geq\kappa$ в силу своего определения. Предположим, что $\kappa<1$. Поскольку функция $f(t)$ является барьером для (α,β), найдется такое число $\bar{\alpha}\in(\alpha,\beta)$, что $\nabla f(t)>0$ при всех $t\in[\bar{\alpha},\beta)$.

Рассмотрим функцию $\phi(t)=\frac{(\nabla f(t))^2}{\nabla^2 f(t)}$, $t\in[\bar{\alpha},\beta)$. Поскольку $\nabla f(t)>0$ и функция $f(t)$ является самосогласованной, в силу $\phi(t)\leq\kappa<1$ имеем

$$\begin{aligned}
\phi'(t) &= 2\nabla f(t) - \left(\frac{\nabla f(t)}{\nabla^2 f(t)}\right)^2 \nabla^2 f'(t)\\
&= \nabla f(t)\left(2-\frac{\nabla f(t)}{\sqrt{\nabla^2 f(t)}}\cdot\frac{\nabla^2 f'(t)}{[\nabla^2 f(t)]^{3/2}}\right)\\
&\geq 2(1-\sqrt{\kappa})\nabla f(t).
\end{aligned}$$

Поэтому при всех $t\in[\bar{\alpha},\beta)$ получаем $\phi(t)\geq\phi(\bar{\alpha})+2(1-\sqrt{\kappa})(f(t)-f(\bar{\alpha}))$. Это невозможно, так как функция $f(t)$ является барьером, а функция $\phi(t)$ ограничена сверху. □

Следствие 1.3.7 *Пусть функция $F(x)$ является ν-самосогласованным барьером для множества $Q\subset\mathbb{R}^n$. Тогда $\nu\geq 1$.*

Доказательство.

Действительно, пусть $x \in \operatorname{int} Q$. Поскольку $Q \subset \mathbb{R}^n$, существует такое направление $u \in \mathbb{R}^n$, что прямая $\{y = x + tu,\ t \in \mathbb{R}\}$ пересекает границу множества Q. Таким образом, рассматривая функцию $f(t) = F(x + tu)$ и пользуясь леммой 1.3.7, мы получаем нужное утверждение. □

Докажем теперь простую нижнюю границу на параметр самосогласованного барьера для неограниченных множеств.

Пусть выпуклое замкнутое множество Q имеет непустую внутренность. Зафиксируем точку $\bar{x} \in \operatorname{int} Q$. Предположим, что существуют *рецессивные направления* $\{p_1, \ldots, p_k\}$ множества Q:

$$\bar{x} + \alpha p_i \in Q \quad \forall \alpha \geq 0.$$

Теорема 1.3.27 *Предположим, что положительные коэффициенты* $\{\beta_i\}_{i=1}^k$ *удовлетворяют условию*

$$\bar{x} - \beta_i\, p_i \notin \operatorname{int} Q, \quad i = 1 \ldots k,$$

и пусть для некоторых положительных $\alpha_1, \ldots, \alpha_k$ *выполнено равенство* $\bar{y} = \bar{x} - \sum\limits_{i=1}^{k} \alpha_i p_i \in Q$. *Тогда параметр* ν *любого самосогласованного барьера для множества* Q *удовлетворяет неравенству*

$$\nu \geq \sum_{i=1}^{k} \frac{\alpha_i}{\beta_i}.$$

Доказательство.

Пусть функция $F(x)$ является ν-самосогласованным барьером для множества Q. Поскольку p_i – рецессивное направление, по теореме 1.3.11 о рецессивном направлении имеем

$$\langle \nabla F(\bar{x}), -p_i \rangle \geq \langle \nabla^2 F(\bar{x}) p_i, p_i \rangle^{1/2} \equiv \| p_i \|_{\bar{x}} \,.$$

Заметим, что $\bar{x} - \beta_i\, p_i \notin Q$. Таким образом, в силу теоремы 1.3.5 норма направления p_i будет достаточно большой: $\beta_i \| p_i \|_{\bar{x}} \geq 1$. Поэтому в силу теоремы 1.3.18 получаем

$$\begin{aligned} \nu &\geq \langle \nabla F(\bar{x}), \bar{y} - \bar{x} \rangle \\ &= \langle \nabla F(\bar{x}), -\sum_{i=1}^{k} \alpha_i p_i \rangle \\ &\geq \sum_{i=1}^{k} \alpha_i \| p_i \|_{\bar{x}} \\ &\geq \sum_{i=1}^{k} \frac{\alpha_i}{\beta_i}. \end{aligned}$$

□

Приведем теперь теорему существования самосогласованного барьера. Рассмотрим выпуклое замкнутое множество Q, $\operatorname{int} Q \neq \emptyset$, и предположим, что Q не содержит прямых. Определим *поляру* множества Q относительно точки $\bar{x} \in \operatorname{int} Q$:

$$P(\bar{x}) = \{s \in \mathbb{R}^n \mid \langle s, x - \bar{x} \rangle \leq 1 \quad \forall x \in Q\}.$$

Можно показать, что для любого $x \in \operatorname{int} Q$ множество $P(x)$ является выпуклым замкнутым и ограниченным множеством с непустой внутренностью. Обозначим $V(x) = \operatorname{vol} P(x)$.

Теорема 1.3.28 *Существуют такие абсолютные константы c_1 и c_2, что функция*

$$U(x) = c_1 \cdot \ln V(x)$$

является $(c_2 \cdot n)$-самосогласованным барьером для Q. □

Функция $U(x)$ называется *универсальным барьером* множества Q. Заметим, что аналитическая сложность задачи (1.3.55), снабженной универсальным барьером, составляет $O\left(\sqrt{n} \cdot \ln \frac{n}{\epsilon}\right)$ обращений к оракулу. Такая эффективность *невозможна* в рамках модели черного ящика.

Вышеприведенный результат имеет в основном теоретический интерес. В общем случае универсальный барьер $U(x)$ не может быть вычислен эффективно. Однако теорема 1.3.28 показывает, что в принципе такие барьеры существуют для *каждого* выпуклого множества. Таким образом, применимость нашего подхода ограничивается только нашими возможностями построения *вычислимых* самосогласованных барьеров с как можно меньшим значением параметра. Процесс построения *барьерной модели* исходной задачи трудно формализуем. Для каждой конкретной задачи могут существовать разные барьерные модели. Мы должны выбрать лучшую из них, учитывая значения параметров барьеров, сложность вычисления градиента и гессиана и трудоемкость решения Ньютоновской системы. В оставшейся части этого параграфа мы покажем как это можно сделать для важных конкретных примеров.

Положительный ортант

Для *положительного ортанта* $\mathbb{R}^n_+$ мы можем выбрать следующий самосогласованный барьер:

$$F(x) = -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu = n,$$

(см. Пример 1.3.3 и теорему 1.3.16). Этот барьер называется *стандартным логарифмическим барьером* для $\mathbb{R}^n_+$. Покажем, что этот барьер оптимален.

Лемма 1.3.8 *Параметр ν любого самосогласованного барьера для $\mathbb{R}^n_+$ удовлетворяет неравенству $\nu \geq n$.*

Доказательство.
Выберем

$$\bar{x} \quad = e \equiv (1, \ldots, 1)^T \in \operatorname{int} \mathbb{R}^n_+,$$

$$p_i \quad = e_i, \quad i = 1 \ldots n,$$

где e_i – i-й базисный вектор в $\mathbb{R}^n$. Тогда условия теоремы 1.3.27 выполняются с $\alpha_i = \beta_i = 1$, $i = 1 \ldots n$. Таким образом

$$\nu \geq \sum_{i=1}^{n} \frac{\alpha_i}{\beta_i} = n.$$

□

Конус Лоренца

Во многих приложениях функциональные компоненты задачи содержат негладкие квадратичные члены вида $\| \ Ax - b \ \|$. Покажем, как эти элементы могут быть описаны с помощью самосогласованных барьеров.

Лемма 1.3.9 *Функция*

$$F(x,t) = -\ln(t^2 - \| \ x \ \|^2)$$

является 2-самосогласованным барьером для выпуклого конуса[3]

$$K_2 = \{(x,t) \in \mathbb{R}^{n+1} \mid \ t \geq \| \ x \ \|\}.$$

Доказательство.
Зафиксируем точку $z = (x,t) \in \operatorname{int} K_2$ и ненулевое направление $u = (h, \tau) \in \mathbb{R}^{n+1}$. Обозначим $\xi(\alpha) = (t + \alpha\tau)^2 - \| \ x + \alpha h \ \|^2$. Нам нужно сравнить производные функции

$$\phi(\alpha) = F(z + \alpha u) = -\ln \xi(\alpha)$$

в точке $\alpha = 0$. Обозначим $\phi^{(\cdot)} = \phi^{(\cdot)}(0)$, $\xi^{(\cdot)} = \xi^{(\cdot)}(0)$. Тогда

$$\xi' = 2(t\tau - \langle x, h \rangle), \quad \xi'' = 2(\tau^2 - \| \ h \ \|^2),$$

$$\phi' = -\frac{\xi'}{\xi}, \quad \phi'' = \left(\frac{\xi'}{\xi}\right)^2 - \frac{\xi''}{\xi}, \quad \phi''' = 3\frac{\xi'\xi''}{\xi^2} - 2\left(\frac{\xi'}{\xi}\right)^3.$$

---

[3] В зависимости от области приложений, это множество имеет разные названия: конус Лоренца, конус второго порядка, и т. д.

Заметим, что неравенство $2\phi'' \geq (\phi')^2$ эквивалентно тому, что $(\xi')^2 \geq 2\xi\xi''$. Таким образом, нужно доказать, что для направления (h, τ) справедливо неравенство

$$(t\tau - \langle x, h\rangle)^2 \geq (t^2 - \| x \|^2)(\tau^2 - \| h \|^2). \tag{1.3.56}$$

Ясно, что можно ограничиться случаем $| \tau | > \| h \|$ (в противном случае правая часть неравенства неположительна). Более того, для минимизации левой части неравенства нужно обеспечить равенство $\operatorname{sign} \tau = \operatorname{sign} \langle x, h\rangle$ (поэтому пусть $\tau > 0$) и $\langle x, h\rangle = \| x \| \cdot \| h \|$. Подставляя эти значения в формулу (1.3.56), получаем правильное неравенство.

Наконец, так как $0 \leq \frac{\xi\xi''}{(\xi')^2} \leq \frac{1}{2}$ и $[1 - \xi]^{3/2} \geq 1 - \frac{3}{2}\xi$, получаем

$$\frac{|\phi'''|}{(\phi'')^{3/2}} = 2\frac{|\xi'| \cdot |(\xi')^2 - \frac{3}{2}\xi\xi''|}{[(\xi')^2 - \xi\xi'']^{3/2}} \leq 2. \qquad \square$$

Покажем, что полученный барьер оптимален.

Лемма 1.3.10 *Параметр ν любого самосогласованного барьера для конуса K_2 удовлетворяет неравенству $\nu \geq 2$.*

Доказательство.
Выберем $\bar{z} = (0, 1) \in \operatorname{int} K_2$ и некоторое $h \in \mathbb{R}^n$, $\| h \| = 1$. Обозначим

$$p_1 = (h, 1), \quad p_2 = (-h, 1), \quad \alpha_1 = \alpha_2 = \tfrac{1}{2}, \quad \beta_1 = \beta_2 = \tfrac{1}{2}.$$

Заметим, что при всех $\gamma \geq 0$ выполняются соотношения $\bar{z} + \gamma p_i = (\pm\gamma h, 1 + \gamma) \in K_2$ и

$$\bar{z} - \beta_i p_i = (\pm\tfrac{1}{2}h, \tfrac{1}{2}) \notin \operatorname{int} K_2,$$

$$\bar{z} - \alpha_1 p_1 - \alpha_2 p_2 = (-\tfrac{1}{2}h + \tfrac{1}{2}h, 1 - \tfrac{1}{2} - \tfrac{1}{2}) = 0 \in K_2.$$

Таким образом, условия теоремы 1.3.27 выполнены, и

$$\nu \geq \frac{\alpha_1}{\beta_1} + \frac{\alpha_2}{\beta_2} = 2. \qquad \square$$

Полуопределенная оптимизация

В полуопределенной оптимизации переменными являются матрицы. Пусть матрица $X = \{X^{(i,j)}\}_{i,j=1}^n$ является симметричной (обозначение: $X \in S^{n\times n}$). В линейном пространстве $S^{n\times n}$ введем следующее скалярное произведение: для произвольных $X, Y \in S^{n\times n}$ положим

$$\langle X, Y\rangle_F = \sum_{i=1}^n \sum_{j=1}^n X^{(i,j)} Y^{(i,j)}, \quad \| X \|_F = \langle X, X\rangle_F^{1/2}.$$

Иногда норма $\| X \|_F$ называется *фробениусовой нормой* матрицы X. Для симметрических матриц X и Y справедливо следующее тождество:

$$\begin{aligned}
\langle X, Y \cdot Y \rangle_F &= \sum_{i=1}^{n} \sum_{j=1}^{n} X^{(i,j)} \sum_{k=1}^{n} Y^{(i,k)} Y^{(j,k)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} X^{(i,j)} Y^{(i,k)} Y^{(j,k)} \\
&= \sum_{k=1}^{n} \sum_{j=1}^{n} Y^{(k,j)} \sum_{i=1}^{n} X^{(j,i)} Y^{(i,k)} \qquad (1.3.57) \\
&= \sum_{k=1}^{n} \sum_{j=1}^{n} Y^{(k,j)} (XY)^{(j,k)} \\
&= \sum_{k=1}^{n} (YXY)^{(k,k)} \; = \; \mathrm{Trace}\,(YXY) \; = \; \langle YXY, I_n \rangle_F .
\end{aligned}$$

В полуопределенной оптимизации нетривиальная часть ограничений задается в форме конуса *неотрицательно определенных* $n \times n$-матриц $\mathcal{P}_n \subset S^{n \times n}$. Напомним, что $X \in \mathcal{P}_n$ тогда и только тогда, когда $\langle Xu, u \rangle \geq 0$ при всех $u \in \mathbb{R}^n$. Если $\langle Xu, u \rangle > 0$ для всех ненулевых u, матрица X называется *положительно определенной*. Такие матрицы составляют внутренность конуса $\mathcal{P}_n$. Заметим, что конус $\mathcal{P}_n$ является выпуклым и замкнутым.

Общая формулировка задачи полуопределенной оптимизации выглядит следующим образом:

$$\begin{aligned}
&\min \langle C, X \rangle_F, \\
\text{при ограничениях:} \quad &\langle A_i, X \rangle_F = b_i, \; i = 1 \ldots m, \qquad (1.3.58) \\
&X \in \mathcal{P}_n,
\end{aligned}$$

где матрицы C и A_i лежат в $S^{n \times n}$. Для применения к этой задаче метода отслеживания траектории нам нужно построить самосогласованный барьер для конуса $\mathcal{P}_n$.

Пусть матрица X принадлежит $\mathrm{int}\, \mathcal{P}_n$. Обозначим $F(X) = -\ln \det X$. Ясно, что

$$F(X) = -\ln \prod_{i=1}^{n} \lambda_i(X),$$

где $\{\lambda_i(X)\}_{i=1}^{n}$ являются собственными значениями матрицы X.

Лемма 1.3.11 *Функция $F(X)$ является выпуклой, и $\nabla F(X) = -X^{-1}$. Для любого на-*

правления $\Delta \in S^{n\times n}$ *выполняются равенства*

$$\begin{aligned}
\langle \nabla^2 F(X)\Delta, \Delta\rangle_F &= \| X^{-1/2}\Delta X^{-1/2} \|_F^2 \\
&= \langle X^{-1}\Delta X^{-1}, \Delta\rangle_F \\
&= \mathrm{Trace}\left([X^{-1/2}\Delta X^{-1/2}]^2\right), \\
D^3F(x)[\Delta,\Delta,\Delta] &= -2\langle I_n, [X^{-1/2}\Delta X^{-1/2}]^3\rangle_F \\
&= -2\mathrm{Trace}\left([X^{-1/2}\Delta X^{-1/2}]^3\right).
\end{aligned}$$

Доказательство.
Зафиксируем такие $\Delta \in S^{n\times n}$ и $X \in \operatorname{int}\mathcal{P}_n$, что $X + \Delta \in \mathcal{P}_n$. Тогда

$$\begin{aligned}
F(X+\Delta) - F(X) &= -\ln\det(X+\Delta) - \ln\det X \\
&= -\ln\det(I_n + X^{-1/2}\Delta X^{-1/2}) \\
&\geq -\ln\left[\tfrac{1}{n}\mathrm{Trace}\,(I_n + X^{-1/2}\Delta X^{-1/2})\right]^n \\
&= -n\ln\left[1 + \tfrac{1}{n}\langle I_n, X^{-1/2}\Delta X^{-1/2}\rangle_F\right] \\
&\geq -\langle I_n, X^{-1/2}\Delta X^{-1/2}\rangle_F = -\langle X^{-1}, \Delta\rangle_F.
\end{aligned}$$

Таким образом, $-X^{-1} \in \partial F(X)$. Следовательно, функция F выпукла, и $\nabla F(x) = -X^{-1}$.

Далее, рассмотрим функцию $\phi(\alpha) \equiv \langle \nabla F(X + \alpha\Delta), \Delta\rangle_F$, $\alpha \in [0,1]$. Имеем

$$\begin{aligned}
\phi(\alpha) - \phi(0) &= \langle X^{-1} - (X+\alpha\Delta)^{-1}, \Delta\rangle_F \\
&= \langle (X+\alpha\Delta)^{-1}[(X+\alpha\Delta) - X]X^{-1}, \Delta\rangle_F \\
&= \alpha\langle (X+\alpha\Delta)^{-1}\Delta X^{-1}, \Delta\rangle_F.
\end{aligned}$$

Таким образом, $\phi'(0) = \langle \nabla^2 F(X)\Delta, \Delta\rangle_F = \langle X^{-1}\Delta X^{-1}, \Delta\rangle_F$.

Последнее равенство обосновывается аналогичным образом с помощью вычисления производной функции $\psi(\alpha) = \langle (X+\alpha\Delta)^{-1}\Delta(X+\alpha\Delta)^{-1}, \Delta\rangle_F$. □

Теорема 1.3.29 *Функция* $F(X)$ *является* n*-самосогласованным барьером для конуса* $\mathcal{P}_n$.

Доказательство.

Зафиксируем $X \in \operatorname{int} \mathcal{P}_n$ и $\Delta \in S^{n\times n}$. Обозначим $Q = X^{-1/2}\Delta X^{-1/2}$ и $\lambda_i = \lambda_i(Q)$, $i = 1 \dots n$. Тогда в силу леммы 1.3.11 имеем

$$\begin{aligned} \langle \nabla F(X), \Delta\rangle_F &= \sum_{i=1}^{n} \lambda_i, \\ \langle \nabla^2 F(X)\Delta, \Delta\rangle_F &= \sum_{i=1}^{n} \lambda_i^2, \\ D^3 F(X)[\Delta,\Delta,\Delta] &= -2\sum_{i=1}^{n} \lambda_i^3. \end{aligned}$$

Используя два стандартных неравенства

$$\begin{aligned} \left(\sum_{i=1}^{n} \lambda_i\right)^2 &\le n\sum_{i=1}^{n} \lambda_i^2, \\ \left|\sum_{i=1}^{n} \lambda_i^3\right| &\le \left(\sum_{i=1}^{n} \lambda_i^2\right)^{3/2}, \end{aligned}$$

получаем

$$\begin{aligned} \langle \nabla F(X), \Delta\rangle_F^2 &\le n\langle \nabla^2 F(X)\Delta, \Delta\rangle_F, \\ \mid D^3 F(X)[\Delta,\Delta,\Delta] \mid &\le 2\langle \nabla^2 F(X)\Delta, \Delta\rangle_F^{3/2}. \end{aligned}$$

□

Покажем, что барьер $F(X) = -\ln\det X$ является оптимальным для конуса $\mathcal{P}_n$.

Лемма 1.3.12 *Параметр ν любого самосогласованного барьера для конуса $\mathcal{P}_n$ не может быть меньше чем n.*

Доказательство.

Выберем $\bar{X} = I_n \in \operatorname{int}\mathcal{P}_n$ и направления $P_i = e_i e_i^T$, $i = 1 \dots n$, где e_i - это i-й базисный вектор в $\mathbb{R}^n$. Заметим, что для любого $\gamma \ge 0$ выполняется включение $I_n + \gamma P_i \in \operatorname{int}\mathcal{P}_n$. Более того,

$$I_n - e_i e_i^T \notin \operatorname{int}\mathcal{P}_n, \quad I_n - \sum_{i=1}^{n} e_i e_i^T = 0 \in \mathcal{P}_n.$$

Таким образом, условия теоремы 1.3.27 выполняются с $\alpha_i = \beta_i = 1$, $i = 1 \dots n$, и мы получаем $\nu \ge \sum\limits_{i=1}^{n} \frac{\alpha_i}{\beta_i} = n$. □

Надграфики функций одной переменной

Приведем самосогласованные барьеры для надграфиков нескольких важных функций от одной переменной.

Логарифм и экспонента. Функция $F_1(x,t) = -\ln x - \ln(\ln x + t)$ является 2-самосогласованным барьером для множества

$$Q_1 = \{(x,t) \in \mathbb{R}^2 \mid \ x > 0, \ t \geq -\ln x\},$$

а функция $F_2(x,t) = -\ln t - \ln(\ln t - x)$ является 2-самосогласованным барьером для множества

$$Q_2 = \{(x,t) \in \mathbb{R}^2 \mid \ t \geq e^x\}.$$

Функция энтропии. Функция $F_3(x,t) = -\ln x - \ln(t - x\ln x)$ является 2-самосогласованным барьером для множества

$$Q_3 = \{(x,t) \in \mathbb{R}^2 \mid \ x \geq 0, \ t \geq x\ln x\}.$$

Возрастающая степенная функция. Функция $F_4(x,t) = -2\ln t - \ln(t^{2/p} - x^2)$ является 4-самосогласованным барьером для множества

$$Q_4 = \{(x,t) \in \mathbb{R}^2 \mid \ t \geq \mid x \mid^p\}, \quad p \geq 1,$$

а функция $F_5(x,t) = -\ln x - \ln(t^p - x)$ является 2-самосогласованным барьером для множества

$$Q_5 = \{(x,t) \in \mathbb{R}^2 \mid \ x \geq 0, \ t^p \geq x\}, \quad 0 < p \leq 1.$$

Убывающая степенная функция. Функция $F_6(x,t) = -\ln t - \ln(x - t^{-1/p})$ является 2-самосогласованным барьером для множества

$$Q_6 = \left\{(x,t) \in \mathbb{R}^2 \mid \ x > 0, \ t \geq \tfrac{1}{x^p}\right\}, \quad p \geq 1,$$

а функция $F_7(x,t) = -\ln x - \ln(t - x^{-p})$ является 2-самосогласованным барьером для множества

$$Q_7 = \left\{(x,t) \in \mathbb{R}^2 \mid \ x > 0, \ t \geq \tfrac{1}{x^p}\right\}, \quad 0 < p < 1.$$

Мы опускаем соответствующие доказательства, так как они основаны на рутинных вычислениях. Можно показать, что приведенные барьеры для всех этих множеств оптимальны (возможно, за исключением Q_4). Докажем соответствующее утверждение для барьеров Q_6 и Q_7.

Лемма 1.3.13 *Параметр любого самосогласованного барьера для множества*

$$Q = \left\{(x^{(1)}, x^{(2)}) \in \mathbb{R}^2 \mid \ x^{(1)} > 0, \ x^{(2)} \geq \tfrac{1}{(x^{(1)})^p}\right\},$$

$p > 0$*, не может быть меньше двух.*

Доказательство.
Зафиксируем $\gamma > 1$ и выберем $\bar{x} = (\gamma, \gamma) \in \operatorname{int} Q$. Обозначим

$$p_1 = e_1, \quad p_2 = e_2, \quad \beta_1 = \beta_2 = \gamma, \quad \alpha_1 = \alpha_2 = \alpha \equiv \gamma - 1.$$

Тогда $\bar{x} + \xi e_i \in Q$ для любого $\xi \geq 0$ и

$$\bar{x} - \beta e_1 = (0, \gamma) \notin Q, \quad \bar{x} - \beta e_2 = (\gamma, 0) \notin Q,$$

$$\bar{x} - \alpha(e_1 + e_2) = (\gamma - \alpha, \gamma - \alpha) = (1, 1) \in Q.$$

Таким образом, условия теоремы 1.3.27 выполнены, и

$$\nu \geq \tfrac{\alpha_1}{\beta_1} + \tfrac{\alpha_2}{\beta_2} = 2\tfrac{\gamma - 1}{\gamma}.$$

Это доказывает нужное неравенство, так как γ может быть произвольно большим. $\square$

Глава 2

Современные субградиентные методы

2.1 Прямо-двойственные методы для негладких задач

2.1.1 Введение

Мотивировка

Исторически субградиентный метод с *постоянным шагом* был первым численным методом для решения задачи

$$\min_x \{ f(x) : \ x \in \mathbb{R}^n \} \tag{2.1.1}$$

с негладкой выпуклой целевой функцией f (подробные ссылки см. в работе [117]). Для его сходящегося варианта, изложенного в работах [50, 109], необходимо заранее выбрать последовательность шагов $\{\lambda_k\}_{k=0}^{\infty}$, удовлетворяющую условию *расходящегося ряда*:

$$\lambda_k > 0, \quad \lambda_k \to 0, \quad \sum_{k=0}^{\infty} \lambda_k = \infty. \tag{2.1.2}$$

Тогда для $k \geq 0$ можно применять правило

$$x_{k+1} \ = \ x_k - \lambda_k g_k, \quad k \geq 0, \tag{2.1.3}$$

с некоторыми $g_k \in \partial f(x_k)$. В другом варианте этой схемы направление движения сначала нормализуется:

$$x_{k+1} \ = \ x_k - \lambda_k g_k / \|g_k\|_2, \quad k \geq 0, \tag{2.1.4}$$

где $\| \cdot \|_2$ обозначает стандартную евклидову норму в $\mathbb{R}^n$, задаваемую скалярным произведением

$$\langle x, y \rangle \ = \ \sum_{j=1}^{n} x^{(j)} y^{(j)}, \quad x, y \in \mathbb{R}^n.$$

Поскольку целевая функция f является негладкой, мы не можем ожидать, что в окрестности решения задачи (2.1.1) субградиенты будут стремиться к нулю. Поэтому условие $\lambda_k \to 0$ необходимо для сходимости процессов (2.1.3) и (2.1.4). С другой стороны, предполагая что $\|g_k\|_2 \leq L$, для всех $x \in \mathbb{R}^n$, получаем

$$\begin{aligned} \|x - x_{k+1}\|_2^2 &\overset{(2.1.3)}{=} \|x - x_k\|_2^2 + 2\lambda_k \langle g_k, x - x_k \rangle + \lambda_k^2 \|g_k\|_2^2 \\ &\leq \|x - x_k\|_2^2 + 2\lambda_k \langle g_k, x - x_k \rangle + \lambda_k^2 L^2. \end{aligned}$$

Поэтому для любого $x \in \mathbb{R}^n$, $\frac{1}{2}\|x - x_0\|_2^2 \leq D$, имеем

$$\begin{aligned} f(x) &\geq l_k(x) \overset{\text{def}}{=} \sum_{i=0}^{k} \lambda_i [f(x_i) + \langle g_i, x - x_i \rangle] / \sum_{i=0}^{k} \lambda_i \\ &\geq \left\{ \sum_{i=0}^{k} \lambda_i f(x_i) - D - \tfrac{1}{2} L^2 \sum_{i=0}^{k} \lambda_i^2 \right\} / \sum_{i=0}^{k} \lambda_i. \end{aligned} \tag{2.1.5}$$

Таким образом, обозначив $f_D^* = \min\limits_x \{ f(x) : \frac{1}{2}\|x - x_0\|_2^2 \leq D \}$,

$$\bar{f}_k = \frac{\sum\limits_{i=0}^{k} \lambda_i f(x_i)}{\sum\limits_{i=0}^{k} \lambda_i}, \quad \omega_k = \frac{2D + L^2 \sum\limits_{i=0}^{k} \lambda_i^2}{2 \sum\limits_{i=0}^{k} \lambda_i},$$

мы получим, что $\bar{f}_k - f_D^* \leq \omega_k$. Заметим, что условие (2.1.2) необходимо и достаточно для сходимости $\omega_k \to 0$.

В приведенном анализе сходимость процесса (2.1.3) основывается на том факте, что производная функции $l_k(x)$, *нижней линейной модели* целевой функции стремится к нулю. Эта модель чрезвычайно важна, поскольку она дает проверяемый критерий остановки. Однако, посмотрев на структуру линейной функции $l_k(\cdot)$, можно заметить одну странную вещь.

Новые субградиенты входят в модель с *убывающими* *весами.* (2.1.6)

Это обстоятельство противоречит здравому смыслу и общим принципам итеративных методов, согласно которым новая информация более важна, чем старая. Таким образом, в стандартной схеме что-то делается не так. К сожалению, простая коррекция этой схемы, по-видимому, невозможна: мы видели, что уменьшающиеся веса ($\equiv$ шаги в прямом пространстве) *необходимы* для сходимости последовательности $\{x_k\}_{k=0}^{\infty}$.

Отмеченное противоречие послужило отправной точкой для результатов, приводимых в этом пункте. Предлагаемая альтернатива выглядит вполне естественно. Действительно, мы видели, что в прямом пространстве нам нужна убывающая последовательность шагов. Но в двойственном пространстве (пространстве линейных функций) нам нужны неубывающие веса. Следовательно, нам необходимы *две разные* последовательности параметров, каждая из которых ответственна за процессы в прямом и двойственном пространствах. Идея связать минимизирующую последовательность в прямом пространстве с основным процессом, формируемым в двойственном пространстве, не является новой. Впервые она была реализована в методах *зеркального спуска* (см. [79]; новые версии и исторический комментарий см. в работах [33, 38]). Однако идея расходящегося ряда каким-то образом просочилась и туда. В результате в евклидовой ситуации метод зеркального спуска, совпадающий с обычным субградиентным методом, разделяет недостаток (2.1.6). В нашем подходе ситуацию удается улучшить. Конечно же, зависимость оценок сложности от точности ϵ остается такой же. Однако улучшение константы оказывается существенным: *наилучшая* стратегия выбора шаговых множителей в работах [33], [38] дает такую же (бесконечную) константу в оценке скорости сходимости, что и *худшая* среди новых схем (см. пункт 2.1.8).

В этом параграфе мы рассматриваем *прямо-двойственные* субградиентные методы. По-видимому, эта естественная особенность *всех* субградиентных методов до сих пор не привлекала к себе внимания. Рассмотрим, например метод (2.1.3). Из неравенства (2.1.5) следует, что

$$f_D^* \quad \geq \quad \hat{f}_k(D) \;\stackrel{\mathrm{def}}{=}\; \min_x \{ l_k(x) : \; \tfrac{1}{2}\|x - x_0\|_2^2 \leq D \}.$$

Заметим, что число $\hat{f}_k(D)$ может быть легко вычислено. С другой стороны, в выпуклой оптимизации существует только один способ получить нижнюю оценку для *оптимального* значения некоторой задачи оптимизации. Для этого необходимо получить *допустимую* точку в некоторой двойственной задаче [1]. Таким образом, вычислимость значения $\hat{f}_k(D)$ означает, что у нас имеется допустимое двойственное решение, и сходимость двойственного зазора $\bar{f}_k - \hat{f}_k(D)$ к нулю означает, что двойственное решение сходится к оптимальному. Ниже мы подробно обсудим разные возможности для формирования двойственных решений с помощью предлагаемых методов.

Наконец, субградиентные методы, описываемые в этом пункте, сильно отличаются от стандартных "поисковых" методов [2]. Например, для задачи (2.1.1) предлагается следующий метод:

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ \tfrac{1}{k+1} \sum_{i=0}^{k} [f(x_i) + \langle g_i, x - x_i \rangle] + \mu_k \|x - x_0\|_2^2 \right\}, \qquad (2.1.7)$$

где $\mu_k = O(\frac{1}{\sqrt{k}}) \to 0$. Заметим что в этой схеме отсутствуют какие-либо искусственные элементы, призванные искусственно ограничить последовательность тестовых точек.

Содержание

В параграфе 2.1.2 мы рассматриваем общую схему метода двойственных усреднений (ДУ) и доказываем верхние оценки для соответствующих функций зазора. Эти оценки будут далее уточняться для соответствующих классов задач. Мы приведем два основных варианта ДУ-методов, простые двойственные усреднения (ПДУ) и взвешенные двойственные усреднения (ВДУ).

---

[1] В зависимости от имеющейся информации о *структуре* задачи, эта двойственная задача может быть поставлена в разном виде. Поэтому иногда лучше называть эту постановку *сопряженной*.

[2] Имеются в виду методы описываемые в терминах шагов в прямом пространстве. Альтернативой являются методы *зеркального спуска* [79], в которых основной процесс идет в двойственном пространстве.

В параграфе 2.1.3 мы воспользуемся ДУ-методами для решения задач минимизации на простых множествах. Основной целью этого параграфа является демонстрация того факта что все ДУ-методы являются прямо-двойственными. Для этого мы всегда формулируем некоторую сопряженную задачу и указываем правила формирования последовательности, сходящейся к ее решению. В пункте 2.1.3(A) мы рассматриваем различные формы ДУ-методов, применяемых к общим задачам минимизации. В пункте 2.1.3(B) это делается для общих задач безусловной минимизации. В пункте 2.1.3(C) это сделано для минимаксных задач. В этом случае, аппроксимация двойственных множителей соответствует частотам, с которыми соответствующие функциональные компоненты оказываются активными в функции максимума (ср. с [28]). Наконец, в пункте 2.1.3(D) мы рассматриваем прямо-двойственные методы для решения задач минимизации с простыми функциональными ограничениями. Для таких задач удается построить приближенное прямо-двойственное решение, несмотря на тот факт, что обычно даже вычисление значения двойственной функции по своей трудоемкости сравнимо с трудоемкостью решения исходной задачи.

В параграфе 2.1.4 мы применяем ПДУ-метод к задаче нахождения седловой точки, а в параграфе 2.1.5 он применяется к вариационным неравенствам. Важной особенностью этой схемы является то, что для разрешимых задач она генерирует ограниченную последовательность точек даже в случае неограниченного допустимого множества. В параграфе 2.1.6 рассматриваются методы решения задачи стохастической оптимизации. Эти методы генерируют случайное приближенное решение (т.е. повторные запуски метода дадут разные результаты). Однако мы доказываем, что среднее значение целевой функции на выходе сходится к оптимальному значению с оптимальной скоростью.

Наконец, в параграфе 2.1.7 мы рассматриваем применение ДУ-методов в моделировании. В пункте 2.1.7(A) показано, что естественная поведенческая стратегия производства (которую мы называем схемой *балансированного развития*) может рассматриваться как реализация ДУ-метода. В результате удается доказать, что в пределе мы получаем оптимальное решение некоторой оптимизационной задачи. В пункте 2.1.7(B) для некоторой многостадийной задачи минимизации мы сравниваем эффективность предварительного планирования и некоторой динамической стратегии настройки (основанной на ПДУ-методе). Динамическая стратегия очень проста в реализации. В то же время, оказывается, что она приводит к результатам сравнимым по своей эффективности с результатами предварительного планирования, даже когда последняя стратегия основана на точном знании будущего.

Общие факты и обозначения

Для измерения расстояний в пространстве E мы пользуемся некоторой (прямой) нормой $\|\cdot\|$, определяющей систему шаров

$$B_r(x) = \{y \in E : \; \|y - x\| \le r\}.$$

Пусть множество Q является выпуклым и замкнутым в E. Предположим, что нам известна некоторая *прокс-функция* $d(x)$ множества Q. Это означает, что $d(x)$ непрерывна и определена всюду на Q. Более того, она сильно выпукла на Q: для всех $x, y \in Q$ и $\alpha \in [0,1]$ выполнено неравенство

$$d(\alpha x + (1-\alpha)y) \ \leq \ \alpha d(x) + (1-\alpha)d(y) - \tfrac{1}{2}\sigma\alpha(1-\alpha)\|x-y\|^2, \tag{2.1.8}$$

где $\sigma > 0$ является ее *параметром выпуклости*. Обозначим через x_0 *прокс-центр* множества Q:

$$x_0 = \arg\min_x \{d(x) : \ x \in Q\}. \tag{2.1.9}$$

Не ограничивая общности, можно считать, что $d(x_0) = 0$. В силу теоремы 1.1.3 прокс-центр корректно определен и

$$d(x) \geq \tfrac{1}{2}\sigma\|x - x_0\|^2, \quad x \in Q.$$

Важным примером прокс-функции является квадрат евклидовой нормы:

$$\|x\| \ = \ \|x\|_2 \ \equiv \ \langle x, x\rangle^{1/2}, \quad d(x) = \tfrac{1}{2}\|x - x_0\|_2^2, \quad \sigma \ = \ 1,$$

для некоторой точки $x_0 \in Q$. В качестве другого примера рассмотрим l_1-норму:

$$\|x\| \ = \ \|x\|_1 \ \equiv \ \sum_{i=1}^{n} |x^{(i)}|. \tag{2.1.10}$$

Положим $Q = \Delta_n \stackrel{\mathrm{def}}{=} \{x \in \mathbb{R}^n_+ : \ \sum\limits_{i=1}^{n} x^{(i)} = 1\}$. Тогда *функция энтропии*

$$d(x) \ = \ \ln n + \sum_{i=1}^{n} x^{(i)} \ln x^{(i)}.$$

будет сильно выпуклой на Q с параметром $\sigma = 1$ и $x_0 = (\frac{1}{n}, \dots \frac{1}{n})^T$ (см. пример 1.1.1).

2.1.2 Основные алгоритмические схемы

Пусть выпуклое замкнутое множество Q снабжено прокс-функцией $d(x)$. Это множество может быть неограниченным (например, $Q \equiv E$). Нам потребуются две вспомогательные функции для множества Q опорного типа:

$$\begin{aligned} \xi_D(s) \ &= \ \max_{x \in Q}\{\langle s, x - x_0\rangle : \ d(x) \leq D\}, \\ V_\beta(s) \ &= \ \max_{x \in Q}\{\langle s, x - x_0\rangle - \beta d(x)\}, \end{aligned} \tag{2.1.11}$$

где $D \geq 0$ и $\beta > 0$ являются параметрами. Первая функция является обычной опорной функцией множества

$$\mathcal{F}_D = \{x \in Q : \ d(x) \leq D\}.$$

Вторая функция представляет собой сглаженный вариант первой. Поскольку функция $d(\cdot)$ строго выпукла, при всех положительных D и β имеем $\operatorname{dom}\xi_D = \operatorname{dom} V_\beta = E^*$. Заметим, что обе функции положительны.

Перечислим важные свойства функции $V(\cdot)$. Если $\beta_2 \geq \beta_1 > 0$, то для любых $s \in E^*$ имеем

$$V_{\beta_2}(s) \leq V_{\beta_1}(s). \tag{2.1.12}$$

Заметим, что степень гладкости функции $V_\beta(\cdot)$ контролируется параметром β.

Лемма 2.1.1 *Функция $V_\beta(\cdot)$ является выпуклой и дифференцируемой всюду на E^*. Более того, ее градиент удовлетворяет условию Липшица с константой $\frac{1}{\beta\sigma}$:*

$$\|\nabla V_\beta(s_1) - \nabla V_\beta(s_2)\| \leq \tfrac{1}{\beta\sigma}\|s_1 - s_2\|_*, \quad \forall s_1, s_2 \in E^*. \tag{2.1.13}$$

При всех $s \in E_*$ вектор $\nabla V_\beta(s)$ принадлежит множеству Q:

$$\nabla V_\beta(s) \ = \ \pi_\beta(s) - x_0, \quad \pi_\beta(s) \ \overset{\text{def}}{=} \ \arg\min_{x\in Q}\{-\langle s, x\rangle + \beta d(x)\}. \tag{2.1.14}$$

(это следует из теоремы 1.1.2.)

В качестве тривиального следствия неравенства (2.1.13) получаем следующее неравенство:

$$V_\beta(s+\delta) \leq V_\beta(s) + \langle \delta, \nabla V_\beta(s)\rangle + \tfrac{1}{2\sigma\beta}\|\delta\|_*^2 \quad \forall s, \delta \in E^*. \tag{2.1.15}$$

Из определения (2.1.9) следует, что $\pi_\beta(0) = x_0$. Поэтому $V_\beta(0) = 0$ и $\nabla V_\beta(0) = 0$. Таким образом, в этом случае из неравенства (2.1.15) с $s = 0$ получаем

$$V_\beta(\delta) \leq \tfrac{1}{2\sigma\beta}\|\delta\|_*^2 \quad \forall \delta \in E^*. \tag{2.1.16}$$

В последующем мы предполагаем, что множество Q является простым, т.е. вычисление вектора $\pi_\beta(s)$ возможно в явном виде. Нам потребуется следующее соотношение между функциями (2.1.11).

Лемма 2.1.2 *При всех $s \in E^*$ и $\beta \geq 0$ выполняется неравенство*

$$\xi_D(s) \ \leq \ \beta D + V_\beta(s). \tag{2.1.17}$$

Доказательство.
Действительно,

$$\begin{aligned} \xi_D(s) \ &= \ \max_{x\in Q}\{\langle s, x - x_0\rangle : \ d(x) \leq D\} \\ &= \ \max_{x\in Q}\ \min_{\beta\geq 0}\ \{\ \langle s, x - x_0\rangle + \beta\,[D - d(x)]\ \} \\ &\leq \ \min_{\beta\geq 0}\ \max_{x\in Q}\ \{\ \langle s, x - x_0\rangle + \beta\,[D - d(x)]\ \} \\ &\leq \ \beta D + V_\beta(s). \qquad\qquad \square \end{aligned}$$

Рассмотрим следующие последовательности:

$$X_k = \{x_i\}_{i=0}^k \subset Q, \quad G_k = \{g_i\}_{i=0}^k \subset E^*, \quad \Lambda_k = \{\lambda_i\}_{i=0}^k \subset R_+.$$

Обычно тестовые точки x_i и веса λ_i генерируются некоторым методом, а векторы (субградиенты) g_i вычисляются оракулом $\mathcal{G}(\cdot)$ типа “черный ящик”,

$$g_i = \mathcal{G}(x_i), \quad i \geq 0,$$

связанным с конкретной выпуклой задачей. В этом параграфе мы рассматриваем задачи, для которых решение $x^* \in Q$ существует и удовлетворяет условию

$$\langle g, x - x^* \rangle \geq 0, \quad g \in \mathcal{G}(x), \ x \in Q. \tag{2.1.18}$$

Назовем x^* the *прямым решением* нашей задачи. В то же время, мы собираемся строить приближения к некоторому *двойственному* (или сопряженному) решению с помощью наблюдаемых субградиентов. В последующем конкретный смысл двойственного объекта всегда будет ясен из контекста.

Мы собираемся приближать прямые и двойственные решения нашей задачи с помощью следующих усредненных объектов:

$$\begin{array}{rclcrcl} S_k & = & \sum\limits_{i=0}^k \lambda_i, & \qquad & \hat{x}_{k+1} & = & \frac{1}{S_k} \sum\limits_{i=0}^k \lambda_i x_i, \\ s_{k+1} & = & \sum\limits_{i=0}^k \lambda_i g_i, & \qquad & \hat{s}_{k+1} & = & \frac{1}{S_k} s_{k+1}, \end{array} \tag{2.1.19}$$

где $\hat{x}_0 = x_0$ и $s_0 = 0$.

Как будет видно из дальнейшего, качество последовательности тестовых точек X_k естественным образом описывается следующей *функцией зазора*:

$$\delta_k(D) = \max_x \left\{ \sum_{i=0}^k \lambda_i \langle g_i, x_i - x \rangle : \ x \in \mathcal{F}_D, \right\}, \quad D \geq 0. \tag{2.1.20}$$

С помощью обозначений (2.1.19) получаем явную формулу для этого зазора:

$$\delta_k(D) = \sum_{i=0}^k \lambda_i \langle g_i, x_i - x_0 \rangle + \xi_D(-s_{k+1}). \tag{2.1.21}$$

Иногда оказывается полезной *сглаженная функция зазора*

$$\begin{array}{rcl} \Delta_k(\beta, D) & = & \beta D + \sum\limits_{i=0}^k \lambda_i \langle g_i, x_i - x_0 \rangle + V_\beta(-s_{k+1}) \\ & = & \sum\limits_{i=0}^k \lambda_i \langle g_i, x_i - \pi_\beta(-s_{k+1}) \rangle + \beta \cdot \left(D - d(\pi_\beta(-s_{k+1})) \right). \end{array} \tag{2.1.22}$$

В силу соотношений (2.1.17) и (2.1.21) для всех неотрицательных D и β имеем

$$\delta_k(D) \leq \Delta_k(\beta, D). \tag{2.1.23}$$

Поскольку множеств Q простое, значения функций зазора можно легко вычислить. Заметим, что при некоторых D эти значения могут быть отрицательными. Однако если решение x^* нашей задачи существует (в смысле (2.1.18)), то для

$$D \geq d(x^*) \quad (\Rightarrow \; x^* \in \mathcal{F}_D)\,,$$

значение $\delta_k(D)$ неотрицательно независимо от последовательностей X_k, Λ_k и G_k, используемых для его определения.

Рассмотрим теперь общую схему *двойственных усреднений* (ДУ).

Инициализация Положим $s_0 = 0 \in E^*$. Выберем $\beta_0 > 0$.

Итерация $(k \geq 0)$

1. Вычисляем $g_k = \mathcal{G}(x_k)$.

2. Выбираем $\lambda_k > 0$. Полагаем $s_{k+1} = s_k + \lambda_k g_k$.

3. Выбираем $\beta_{k+1} \geq \beta_k$. Полагаем $x_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1})$.

(2.1.24)

Теорема 2.1.1 *Пусть последовательности X_k, G_k и Λ_k построены по схеме (2.1.24). Тогда*

1. *При всех $k \geq 0$ и $D \geq 0$*

$$\delta_k(D) \;\leq\; \Delta_k(\beta_{k+1}, D) \;\leq\; \beta_{k+1}D + \tfrac{1}{2\sigma}\sum_{i=0}^{k} \tfrac{\lambda_i^2}{\beta_i}\|g_i\|_*^2. \tag{2.1.25}$$

2. *Предположим, что решение x^* в смысле неравенства (2.1.18) существует. Тогда*

$$\tfrac{1}{2}\sigma\|x_{k+1} - x^*\|^2 \leq d(x^*) + \tfrac{1}{2\sigma\beta_{k+1}}\sum_{i=0}^{k} \tfrac{\lambda_i^2}{\beta_i}\|g_i\|_*^2. \tag{2.1.26}$$

3. *Предположим, что x^* является внутренним решением: $B_r(x^*) \subseteq \mathcal{F}_D$ для некоторых положительных r и D. Тогда*

$$\|\hat{s}_{k+1}\|_* \leq \tfrac{1}{rS_k}\left[\beta_{k+1}D + \tfrac{1}{2\sigma}\sum_{i=0}^{k} \tfrac{\lambda_i^2}{\beta_i}\|g_i\|_*^2\right]. \tag{2.1.27}$$

Доказательство.

1. В силу оценки (2.1.23), нам нужно доказать только второе из неравенств (2.1.25). Ввиду правил пересчета в схеме (2.1.24) при всяком $i \geq 1$ имеем

$$\begin{aligned} V_{\beta_{i+1}}(-s_{i+1}) &\overset{(2.1.12)}{\leq} V_{\beta_i}(-s_{i+1}) \\ &\overset{(2.1.15)}{\leq} V_{\beta_i}(-s_i) - \lambda_i \langle g_i, \nabla V_{\beta_i}(-s_i)\rangle + \tfrac{\lambda_i^2}{2\sigma\beta_i}\|g_i\|_*^2 \\ &\overset{(2.1.14)}{=} V_{\beta_i}(-s_i) + \lambda_i\langle g_i, x_0 - x_i\rangle + \tfrac{\lambda_i^2}{2\sigma\beta_i}\|g_i\|_*^2. \end{aligned}$$

Таким образом,

$$\lambda_i\langle g_i, x_i - x_0\rangle \;\leq\; V_{\beta_i}(-s_i) - V_{\beta_{i+1}}(-s_{i+1}) + \tfrac{\lambda_i^2}{2\sigma\beta_i}\|g_i\|_*^2, \quad i = 1, \ldots, k.$$

Просуммировав эти неравенства, получим

$$\sum_{i=0}^{k} \lambda_i\langle g_i, x_i - x_0\rangle \leq V_{\beta_1}(-s_1) - V_{\beta_{k+1}}(-s_{k+1}) + \tfrac{1}{2\sigma}\sum_{i=1}^{k}\tfrac{\lambda_i^2}{\beta_i}\|g_i\|_*^2. \tag{2.1.28}$$

Но в силу оценки (2.1.16) имеем $V_{\beta_1}(-s_1) \leq \frac{\lambda_0^2}{2\sigma\beta_1}\|g_0\|_*^2 \leq \frac{\lambda_0^2}{2\sigma\beta_0}\|g_0\|_*^2$. Таким образом, неравенство (2.1.28) приводит к оценке (2.1.25).

2. Предположим, что решение x^* существует. Тогда

$$\begin{aligned} 0 &\overset{(2.1.18)}{\leq} \sum_{i=0}^{k}\lambda_i\langle g_i, x_i - x^*\rangle \\ &\overset{(2.1.28)}{\leq} \langle s_{k+1}, x_0 - x^*\rangle - V_{\beta_{k+1}}(-s_{k+1}) + \tfrac{1}{2\sigma}\sum_{i=0}^{k}\tfrac{\lambda_i^2}{\beta_i}\|g_i\|_*^2 \\ &\overset{(2.1.11)}{=} \langle s_{k+1}, x_{k+1} - x^*\rangle + \beta_{k+1}d(x_{k+1}) + \tfrac{1}{2\sigma}\sum_{i=0}^{k}\tfrac{\lambda_i^2}{\beta_i}\|g_i\|_*^2 \\ &\overset{(1.1.30)}{\leq} \beta_{k+1}d(x^*) - \tfrac{1}{2}\beta_{k+1}\sigma\|x_{k+1} - x^*\|^2 + \tfrac{1}{2\sigma}\sum_{i=0}^{k}\tfrac{\lambda_i^2}{\beta_i}\|g_i\|_*^2, \end{aligned}$$

и это в точности неравенство (2.1.26).

3. Предположим теперь, что x^* является внутренним решением. Тогда

$$\begin{aligned} \delta_k(D) &= \max_x\left\{\sum_{i=0}^{k}\lambda_i\langle g_i, x_i - x\rangle : \; x \in \mathcal{F}_D\right\} \\ &\geq \max_x\left\{\sum_{i=0}^{k}\lambda_i\langle g_i, x_i - x\rangle : \; x \in B_r(x^*)\right\} \geq r\|s_{k+1}\|_*. \end{aligned}$$

Таким образом, неравенство (2.1.27) следует из оценки (2.1.25). □

Форма неравенств (2.1.25) и (2.1.26) приводит к некоторым естественным стратегиям выбора параметров β_i в схеме (2.1.24). Определим следующую последовательность:

$$\hat{\beta}_0 = \hat{\beta}_1 = 1, \quad \hat{\beta}_{i+1} = \hat{\beta}_i + \tfrac{1}{\hat{\beta}_i}, \; i \geq 1. \tag{2.1.29}$$

Разумность ее определения подтверждается следующим соотношением:

$$\hat{\beta}_{k+1} = \sum_{i=0}^{k} \frac{1}{\hat{\beta}_i}, \quad k \geq 0.$$

Таким образом, эта последовательность может использоваться для балансировки двух членов, возникающих в правой части неравенства (2.1.25). Заметим, что скорость роста этой последовательности оценивается следующим образом.

Лемма 2.1.3 *Справедливы неравенства*

$$\sqrt{2k-1} \leq \hat{\beta}_k \leq \frac{1}{1+\sqrt{3}} + \sqrt{2k-1}, \quad k \geq 1. \tag{2.1.30}$$

Доказательство.

Из соотношений (2.1.29) имеем $\hat{\beta}_1 = 1$ и $\hat{\beta}_{k+1}^2 = \hat{\beta}_k^2 + \hat{\beta}_k^{-2} + 2$ for $k \geq 1$. Из этого следует первое неравенство (2.1.30). С другой стороны, поскольку функция $\beta + \frac{1}{\beta}$ возрастает при $\beta \geq 1$ и $\hat{\beta}_k \geq 1$, второе неравенство легко доказывается по индукции. □

Рассмотрим две основные стратегии для выбора λ_i в методе (2.1.24):

- *простые усреднения:* $\lambda_k = 1$;
- *взвешенные усреднения:* $\lambda_k = \frac{1}{\|g_k\|_*}$.

Давайте запишем соответствующие схемы в явном виде. Оба приведенных ниже утверждения непосредственно следуют из теоремы 2.1.1.

Метод простых усреднений

Инициализация Set $s_0 = 0 \in E^*$. Выбираем $\gamma > 0$.

Итерация ($k \geq 0$): (2.1.31)

1. Вычисляем $g_k = \mathcal{G}(x_k)$. Полагаем $s_{k+1} = s_k + g_k$.

2. Выбираем $\beta_{k+1} = \gamma \hat{\beta}_{k+1}$. Полагаем $x_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1})$.

Теорема 2.1.2 *Предположим, что* $\|g_k\|_* \leq L$, $k \geq 0$. *Для метода (2.1.31) справедливо* $S_k = k+1$ *и*

$$\delta_k(D) \leq \hat{\beta}_{k+1} \left(\gamma D + \frac{L^2}{2\sigma\gamma} \right).$$

Более того, если решение x^* *в смысле (2.1.18) существует, то схема (2.1.31) генерирует ограниченную последовательность:*

$$\|x_k - x^*\|^2 \leq \frac{2}{\sigma} d(x^*) + \frac{L^2}{\sigma^2 \gamma^2}, \quad k \geq 0.$$

Метод взвешенных усреднений

Инициализация Полагаем $s_0 = 0 \in E^*$. Выбираем $\rho > 0$.

Итерация ($k \geq 0$): (2.1.32)

1. Вычисляем $g_k = \mathcal{G}(x_k)$. Полагаем $s_{k+1} = s_k + g_k / \|g_k\|_*$.

2. Выбираем $\beta_{k+1} = \frac{\hat{\beta}_{k+1}}{\rho\sqrt{\sigma}}$. Полагаем $x_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1})$.

Теорема 2.1.3 *Предположим, что* $\|g_k\|_* \leq L$, $k \geq 0$. *Для метода (2.1.32) выполнены неравенства* $S_k \geq \frac{k+1}{L}$ *и*

$$\delta_k(D) \leq \frac{\hat{\beta}_{k+1}}{\sqrt{\sigma}} \left(\frac{D}{\rho} + \frac{1}{2}\rho \right).$$

Более того, если решение x^* *в смысле (2.1.18) существует, то метод (2.1.32) генерирует ограниченную последовательность:*

$$\|x_k - x^*\|^2 \leq \frac{1}{\sigma}(2d(x^*) + \rho^2).$$

В следующих пунктах мы покажем как можно применять вышеприведенные результаты к различным классам задач, обладающих выпуклой структурой.

2.1.3 Минимизация на простых множествах

A. Общая задача минимизации

Рассмотрим следующую задачу минимизации:

$$\min_x \{ f(x) : \ x \in Q \}, \tag{2.1.33}$$

где f – выпуклая на E функция и множество Q является выпуклым и замкнутым. Напомним что мы предполагаем, что множество Q простое, а это означает возможность решения вспомогательных задач минимизации (2.1.11) в явном виде.

Для решения задачи (2.1.33), нам нужен оракул типа "черный ящик", который может вычислять субградиент целевой функции в тестовых точках[3]:

$$\mathcal{G}(x) \in \partial f(x), \quad x \in E.$$

---

[3] Обычно такой оракул может вычислять и значение целевой функции. Однако эта информация не используется в рассматриваемых методах.

Тогда у нас имеется следующая интерпретация функции зазора $\delta_k(D)$. Обозначим

$$\begin{array}{rcl} l_k(x) & = & \frac{1}{S_k}\sum\limits_{i=0}^{k}\lambda_i[f(x_i)+\langle g_i, x-x_i\rangle], \quad (g_i \in \partial f(x_i)) \\ \hat{f}_k(D) & = & \min\limits_{x}\{l_k(x) :\ x \in \mathcal{F}_D\}, \\ f^*_D & = & \min\limits_{x}\{f(x) :\ x \in \mathcal{F}_D\}. \end{array}$$

Поскольку функция $f(\cdot)$ выпукла, в силу определения функции зазора (2.1.20) имеем

$$\tfrac{1}{S_k}\delta_k(D) \ = \ \tfrac{1}{S_k}\sum_{i=0}^{k}\lambda_i f(x_i) - \hat{f}_N(D) \ \geq \ f(\hat{x}_{k+1}) - f^*_D. \qquad (2.1.34)$$

Таким образом, мы можем оценивать скорость сходимости методов (2.1.31) и (2.1.32), применяемых к задаче (2.1.33). В получаемых оценках мы предполагаем, что

$$\|g\|_* \leq L \quad \forall g \in \partial f(x),\ \forall x \in Q.$$

1. Простые усреднения. В силу теоремы 2.1.2 и неравенств (2.1.30), (2.1.34) имеем

$$f(\hat{x}_{k+1}) - f^*_D \leq \tfrac{0.5+\sqrt{2k+1}}{k+1}\left(\gamma D + \tfrac{L^2}{2\sigma\gamma}\right). \qquad (2.1.35)$$

Заметим, что параметры D и L не используются в самом методе. Однако они необходимы для правильного выбора параметра γ. Оптимальный выбор этого параметра таков:

$$\gamma^* = \tfrac{L}{\sqrt{2\sigma D}}.$$

В методе простых двойственных усреднений (ПДУ) накапливаемая нижняя линейная оценка очень проста:

$$l_k(x) = \tfrac{1}{k+1}\sum_{i=0}^{k}[f(x_i)+\langle g_i, x-x_i\rangle].$$

Таким образом, соответствующая алгоритмическая схема выглядит так:

$$x_{k+1} \ = \ \arg\min_{x\in Q}\left\{ \tfrac{1}{k+1}\sum_{i=0}^{k}[f(x_i)+\langle g_i, x-x_i\rangle] + \mu_k d(x) \right\}, \quad k \geq 0, \qquad (2.1.36)$$

где $\{\mu_k\}_{k=0}^{\infty}$ – это последовательность шкалирующих параметров. В соответствии с правилами (2.1.31) можно выбрать $\mu_k = \frac{\beta_{k+1}}{k+1}$. Однако скорость сходимости метода (2.1.36) остается на том же уровне при любых $\mu_k = O\left(\frac{1}{\sqrt{k}}\right)$. Заметим, что в методе (2.1.36) нам не нужны значения целевой функции.

Для метода (2.1.36) не требуется ограниченность множества Q. Например, возможна ситуация $Q = E$. Тем не менее, если решение x^* задачи (2.1.33) существует, то по теореме 2.1.2 генерируемая последовательность $\{x_k\}_{k=0}^{\infty}$ будет ограниченной. Это свойство может показаться удивительным, так как $\mu_k \to 0$ и в методе (2.1.36) нет никаких специальных средств для обеспечения ограниченности генерируемой последовательности.

2. Взвешенные усреднения. В силу теоремы 2.1.3 и неравенств (2.1.30), (2.1.34) имеем

$$f(\hat{x}_{k+1}) - f^*_D \le \tfrac{0.5+\sqrt{2k+1}}{(k+1)\sqrt{\sigma}} L \left(\tfrac{1}{\rho} D + \tfrac{\rho}{2} \right). \tag{2.1.37}$$

Как и в методе (2.1.31), параметры D и L явно в методе (2.1.32) не используются. Но теперь для разумного выбора параметра ρ необходимо иметь оценку только для D. Оптимальный выбор ρ таков: $\rho^* = \sqrt{2D}$.

B. Прямо-двойственная задача

Для упрощения обозначений в этом разделе будем считать что прокс-центр множества Q находится в нуле:

$$x_0 = \arg\min_{x \in Q} \, d(x) \; = \; 0 \in E. \tag{2.1.38}$$

Предположим, что оптимальное решение x^* задачи (2.1.33) существует. Тогда, выбирая $D \ge d(x^*)$, мы можем переписать эту задачу в эквивалентной форме:

$$f^* \; = \; \min_x \{ f(x) : \; x \in \mathcal{F}_D \}. \tag{2.1.39}$$

Рассмотрим сопряженную функцию

$$f_*(s) = \sup_{x \in E} [\, \langle s, x \rangle - f(x) \,]. \tag{2.1.40}$$

Заметим, что при всех $x \in E$ выполнено включение

$$\partial f(x) \subseteq \operatorname{dom} f_*. \tag{2.1.41}$$

Поскольку $\operatorname{dom} f = E$, обратное представление в формуле (2.1.40) всегда возможно:

$$f(x) = \max_s [\, \langle s, x \rangle - f_*(s) : \; s \in \operatorname{dom} f_* \,], \quad x \in E.$$

Поэтому задачу (2.1.39) можно переписать в двойственной форме:

$$\begin{aligned} f^* \; &= \; \min_{x \in \mathcal{F}_D} \; \max_{s \in \operatorname{dom} f_*} [\, \langle s, x \rangle - f_*(s) \,] \; = \; \max_{s \in \operatorname{dom} f_*} \; \min_{x \in \mathcal{F}_D} [\, \langle s, x \rangle - f_*(s) \,] \\ &= \; \max_{s \in \operatorname{dom} f_*} [\, -\xi_D(-s) - f_*(s) \,]. \end{aligned}$$

Таким образом, получаем следующую двойственную задачу:

$$-f^* \; = \; \min_s [\, f_*(s) + \xi_D(-s) : \; s \in \operatorname{dom} f_* \,]. \tag{2.1.42}$$

Как обычно, задачи (2.1.33) и (2.1.42) можно объединить в одну *прямо-двойственную* задачу:

$$0 \; = \; \min_{x,s} [\, \psi_D(x,s) \stackrel{\mathrm{def}}{=} f(x) + f_*(s) + \xi_D(-s) : \; x \in Q, \; s \in \operatorname{dom} f_* \,]. \tag{2.1.43}$$

Покажем, что ДУ-методы сходятся к решению этой задачи.

Теорема 2.1.4 *Пусть пара* $(\hat{x}_{k+1}, \hat{s}_{k+1})$ *определена по правилам (2.1.19) с помощью последовательностей* X_k*,* G_k *и* Λ_k*? сгенерированных методом (2.1.24) при решении задачи (2.1.39). Тогда* $\hat{x}_{k+1} \in Q$*,* $\hat{s}_{k+1} \in \mathrm{dom}\, f_*$ *и*

$$\psi_D(\hat{x}_{k+1}, \hat{s}_{k+1}) \;\le\; \tfrac{1}{S_k}\delta_k(D). \tag{2.1.44}$$

Доказательство.

Действительно, точка $\hat{x}_{k+1}$ является допустимой как выпуклая комбинация допустимых точек. В силу включения (2.1.41) те же аргументы применимы и к $\hat{s}_{k+1}$. Наконец,

$$\begin{array}{rcl} \psi_D(\hat{x}_{k+1}, \hat{s}_{k+1}) & = & \xi_D(-\hat{s}_{k+1}) + f(\hat{x}_{k+1}) + f_*(\hat{s}_{k+1}) \\ & \overset{(2.1.19)}{\le} & \xi_D(-\hat{s}_{k+1}) + \frac{1}{S_k}\sum\limits_{i=0}^{k} \lambda_i[f(x_i) + f_*(g_i)] \\ & \overset{(2.1.11),(2.1.40)}{=} & \frac{1}{S_k}\xi_D(-s_{k+1}) + \frac{1}{S_k}\sum\limits_{i=0}^{k} \lambda_i\langle g_i, x_i\rangle \\ & \overset{(2.1.21),(2.1.38)}{=} & \frac{1}{S_k}\delta_k(D). \qquad\qquad \Box \end{array}$$

Таким образом, для соответствующих методов правые части неравенств (2.1.35), (2.1.37) устанавливают скорость стремления прямо-двойственной функции в (2.1.44) к нулю. В случае внутреннего решения x^* оптимальное решение двойственной задачи (2.1.42) достигается в $0 \in E^*$. В этом случае, скорость сходимости точек $\hat{s}_{k+1}$ к нулю вытекает из оценки (2.1.27).

В этом разделе мы показали возможности ДУ-методов (2.1.24) при решении общей прямо-двойственной задачи (2.1.43). Однако, как мы увидим в дальнейшем, эти методы могут генерировать и гораздо более детальную информацию. Для этого нужно просто использовать больше информации о структуре конкретной задачи.

C. Минимаксные задачи

Рассмотрим следующий вариант задачи (2.1.33):

$$\min_x \{ f(x) = \max_{1\le j\le p} f_j(x) : \; x \in Q \}, \tag{2.1.45}$$

где функции $f_j(\cdot)$, $j = 1, \ldots, p$, выпуклы на E. Эта задача допускает двойственное представление. Зафиксируем $D \ge d(x^*)$. Напомним что обозначение Δ_p используется для стандартного симплекса в $\mathbb{R}^p$:

$$\Delta_p \;=\; \left\{ y \in \mathbb{R}^p_+ : \; \sum_{j=1}^{p} y^{(j)} = 1 \right\}.$$

Тогда

$$\begin{array}{rcl} f^* & = & \min\limits_{x\in Q} \max\limits_{1\le j\le p} f_j(x) \;=\; \min\limits_{x\in \mathcal{F}_D} \max\limits_{y\in\Delta_p} \sum\limits_{j=1}^{p} y^{(j)} f_j(x) \\ & = & \max\limits_{y\in\Delta_p} \min\limits_{x\in\mathcal{F}_D} \sum\limits_{j=1}^{p} y^{(j)} f_j(x). \end{array}$$

Таким образом, определяя $\phi_D(y) = \min\limits_{x \in \mathcal{F}_D} \sum\limits_{j=1}^{p} y^{(j)} f_j(x)$, получаем двойственную задачу

$$f^* \;=\; \max_{y \in \Delta_p} \phi_D(y). \tag{2.1.46}$$

Покажем, что ДУ-методы могут генерировать и приближенное решение двойственной задачи. Для этого нам нужно воспользоваться структурой оракула $\mathcal{G}$, вычисляющего значение целевой функции задачи (2.1.45). В нашем случае

$$\partial f(x) \;=\; \mathrm{Conv}\,\{\partial f_j(x) : \; j \in I(x)\}, \quad I(x) \;=\; \{j : \; f_j(x) = f(x)\}.$$

Таким образом, для любого g_k в методе (2.1.24), применяемом для решения задачи (2.1.45), мы можем определить такой вектор $y_k \in \Delta_p$, что

$$\begin{aligned} y_k^{(j)} \;&=\; 0, \quad j \notin I(x_k), \\ g_k \;&=\; \sum_{j \in I(x_k)} y_k^{(j)} g_{k,j}, \end{aligned} \tag{2.1.47}$$

где $g_{k,j} \in \partial f_j(x_k)$ для $j \in I(x_k)$. Обозначим $\hat{y}_{k+1} = \frac{1}{S_k} \sum\limits_{i=0}^{k} y_i$.

Теорема 2.1.5 *Пусть пара $(\hat{x}_{k+1}, \hat{y}_{k+1})$ определена с помощью последовательностей X_k, G_k и Λ_k, сгенерированных методом (2.1.24) при решении задачи (2.1.45). Тогда эта пара является допустимым прямо-двойственным решением и*

$$0 \;\le\; f(\hat{x}_{k+1}) - \phi_D(\hat{y}_{k+1}) \;\le\; \tfrac{1}{S_k} \delta_k(D). \tag{2.1.48}$$

Доказательство.

Действительно, пара $(\hat{x}_{k+1}, \hat{y}_{k+1})$ допустима в силу выпуклости прямо-двойственного допустимого множества. Далее, обозначим $F(x) = (f_1(x), \ldots, f_p(x))^T \in \mathbb{R}^p$. Тогда в силу соотношений (2.1.47) для любого $k \ge 0$ имеем

$$\begin{aligned} \langle g_k, x_k - x \rangle \;&=\; \sum_{j \in I(x_k)} y_k^{(j)} \langle g_{k,j}, x_k - x \rangle \;\ge\; \sum_{j \in I(x_k)} y_k^{(j)} [f_j(x_k) - f_j(x)] \\ &=\; \langle y_k, F(x_k) - F(x) \rangle \;=\; f(x_k) - \langle y_k, F(x) \rangle. \end{aligned}$$

Таким образом,

$$\begin{aligned} \tfrac{1}{S_k} \delta_k(D) \;&=\; \tfrac{1}{S_k} \max_{x \in \mathcal{F}_D} \left\{ \sum_{i=0}^{k} \lambda_i \langle g_i, x_i - x \rangle \right\} \\ &\ge\; \tfrac{1}{S_k} \max_{x \in \mathcal{F}_D} \left\{ \sum_{i=0}^{k} \lambda_i [f(x_i) - \langle y_i, F(x) \rangle] \right\} \\ &=\; \tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(x_i) - \phi_D(\hat{y}_{k+1}) \;\ge\; f(\hat{x}_{k+1}) - \phi_D(\hat{y}_{k+1}). \qquad \Box \end{aligned}$$

Выпишем теперь явную форму ПДУ-метода, применяемого к задаче (2.1.45). Через e_j обозначим j-й координатный вектор в $\mathbb{R}^p$.

Инициализация Полагаем $l_0(x) \equiv 0$, $m_0 = 0 \in Z^p$.

Итерация $(k \geq 0)$

1. Выберем любое j_k^* : $f_{j_k^*}(x_k) = f(x_k)$.

2. Положим $l_{k+1}(x) = \frac{k}{k+1} l_k(x) + \frac{1}{k+1}[f(x_k) + \langle g_{k,j_k^*}, x - x_k \rangle]$. (2.1.49)

3. Вычислим $x_{k+1} = \arg\min\limits_{x \in Q} \left\{ l_{k+1}(x) + \frac{\gamma \hat{\beta}_{k+1}}{k+1} d(x) \right\}$.

4. Положим $m_{k+1} = m_k + e_{j_k^*}$.

Выход: $\hat{x}_{k+1} = \frac{1}{k+1} \sum\limits_{i=0}^{k} x_i, \quad \hat{y}_{k+1} = \frac{1}{k+1} m_{k+1}$.

В методе (2.1.49) каждая координата оптимального двойственного вектора приближается *частотой* с которой соответствующая функция оказывается активной в функции максимума.

Заметим, что ПДУ-метод может также применяться к *непрерывной* минимаксной задачи. В этом случае целевая функция имеет следующую форму:

$$f(x) = \max_y \{ \langle y, F(x) \rangle : \ y \in Q_d \},$$

где Q_d – замкнутое выпуклое ограниченное множество в $\mathbb{R}^p$. Для этой задачи приближение к оптимальному двойственному решению может быть получено с помощью усреднения векторов $y(x_k)$, определяемых как

$$y(x) \in \operatorname{Arg}\max_y \{ \langle y, F(x) \rangle : \ y \in Q_d \}.$$

(Заметим, что для вычисления $y(x)$ необходимо вычислить максимум линейной функции на выпуклом множестве.) Соответствующая модификация метода (2.1.49) тривиальна.

D. Задачи с простыми функциональными ограничениями

Предположим, что множество Q в задаче (2.1.33) имеет следующую структуру:

$$Q \ = \ \{ \ x \in \bar{Q} : \quad Ax = b, \quad F(x) \leq 0 \ \}, \tag{2.1.50}$$

где $\bar{Q}$ – выпуклое замкнутое множество в E, $\dim E = n$, $b \in \mathbb{R}^m$, A – $m \times n$-матрица и функция $F(x) : \bar{Q} \to \mathbb{R}^p$ покомпонентно выпукла. Обычно, в зависимости от важности функциональных компонент, мы можем решить какие из них должны быть включены в описание множества $\bar{Q}$, а какие нет. В любом случае представление множества Q в форме (2.1.50) не является единственным. Поэтому мы будем называть соответствующую двойственную задачу *сопряженной.*

Введем двойственные переменные $u \in \mathbb{R}^m$ и $v \in \mathbb{R}^p_+$. Предположим, что оптимальное решение x^* задачи (2.1.33), (2.1.50) существует и $D \geq d(x^*)$. Обозначим

$$\phi_D(u,v) \;=\; \min_{x \in \bar{Q}} \{ f(x) + \langle u, b - Ax \rangle + \langle v, F(x) \rangle : \; d(x) \leq D \}. \tag{2.1.51}$$

В этом определении $d(x)$ – прокс-функция множества Q с прокс-центром $x_0 \in Q$. Тогда задача сопряженная к (2.1.33), (2.1.50), записывается так:

$$\max_{u,v} \{ \phi_D(u,v) : \; u \in \mathbb{R}^m, \; v \in \mathbb{R}^p_+ \}. \tag{2.1.52}$$

Заметим, что сложность вычисления целевой функции в этой задаче может быть сравнима с решением исходной задачи (2.1.33). Тем не менее, мы покажем, что ДУ-методы (2.1.24) способны находить приближение к ее оптимальному решению. Рассмотрим две возможности.

1. Зафиксируем $D \geq d(x^*)$. Обозначим через $\hat{u}_k$, $\hat{v}_k$ оптимальные двойственные множители для существенных ограничений в следующих оптимизационных задачах:

$$\begin{aligned} \tfrac{1}{S_k}\delta_k \;&=\; \max_{x \in \bar{Q}} \left\{ \tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i \langle g_i, x_i - x \rangle : \; Ax = b, \; F(x) \leq 0, \; d(x) \leq D \right\} \\ &=\; \max_{x \in \bar{Q},\, d(x) \leq D} \quad \min_{u \in \mathbb{R}^m, v \in \mathbb{R}^p_+} \quad \hat{\mathcal{L}}(x,u,v), \end{aligned} \tag{2.1.53}$$

$$\hat{\mathcal{L}}(x,u,v) \;=\; \tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i \langle g_i, x_i - x \rangle + \langle u, Ax - b \rangle - \langle v, F(x) \rangle,$$

и пусть $\hat{x}_k$ – соответствующее оптимальное решение. Тогда для любого такого $x \in \bar{Q}$, что $d(x) \leq D$, имеем

$$\langle -\tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i g_i + A^T \hat{u}_k - \sum_{j=1}^{p} \hat{v}_k^{(j)} \psi_j, \; x - \hat{x}_k \; \rangle \;\leq\; 0,$$

для некоторого $\psi_j \in \partial F^{(j)}(\hat{x}_k)$, $j = 1, \ldots, p$, и

$$\sum_{j=1}^{p} \hat{v}_k^{(j)} F^{(j)}(\hat{x}_k) = 0.$$

Таким образом, поскольку функции f и F выпуклы, для любого такого x получаем

$$\begin{aligned} & f(x) + \langle \hat{u}_k, b - Ax \rangle + \langle \hat{v}_k, F(x) \rangle \\ \geq \ & \tfrac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i [f(x_i) + \langle g_i, x - x_i \rangle] + \langle \hat{u}_k, b - Ax \rangle \\ & + \sum\limits_{j=1}^{p} \hat{v}_k^{(j)} [F^{(j)}(\hat{x}_k) + \langle \psi_j, x - \hat{x}_k \rangle] \\ \geq \ & \tfrac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i [f(x_i) + \langle g_i, \hat{x}_k \ - x_i \rangle] \ = \ \tfrac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i f(x_i) \ - \ \tfrac{1}{S_k} \delta_k(D). \end{aligned}$$

Итак, мы доказали, что

$$0 \ \leq \ f(\hat{x}_{k+1}) - \phi_D(\hat{u}_k, \hat{v}_k) \ \leq \ \tfrac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i f(x_i) - \phi_D(\hat{u}_k, \hat{v}_k) \ \leq \ \tfrac{1}{S_k} \delta_k(D). \tag{2.1.54}$$

Поэтому качество этого прямо-двойственного объекта может быть получено из пункта 1 теоремы 2.1.1.

2. Приведенная выше стратегия формирования приближенного решения сопряженной задачи (2.1.52) с помощью двойственных множителей задачи (2.1.53) имеет два недостатка. Во-первых, нужно уметь выбирать параметр D достаточно большим. Во-вторых, для решения задачи (2.1.53) требуются дополнительные вычислительные ресурсы. Покажем, что приближенное решение задачи (2.1.52) может быть получено почти “задаром”, помощью объектов, необходимых для нахождения точки $\pi_{\beta_{k+1}}(-s_{k+1})$.

Обозначим через $\bar{u}_k$, $\bar{v}_k$ оптимальные множители для функциональных ограничений в следующей задаче:

$$\begin{aligned} & x_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1}) \\ = \ & \arg\max_{x \in \bar{Q}} \left\{ -\tfrac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i \langle g_i, x \rangle - \tfrac{\beta_{k+1}}{S_k} d(x) : \ Ax = b, \ F(x) \leq 0 \right\} \\ = \ & \arg\max_{x \in \bar{Q}} \quad \min_{u \in \mathbb{R}^m, v \in \mathbb{R}^p_+} \quad \bar{\mathcal{L}}(x, u, v), \end{aligned} \tag{2.1.55}$$

$$\bar{\mathcal{L}}(x, u, v) \ = \ -\tfrac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i \langle g_i, x \rangle - \tfrac{\beta_{k+1}}{S_k} d(x) + \langle u, Ax - b \rangle - \langle v, F(x) \rangle.$$

Тогда при любых $x \in \bar{Q}$ имеем

$$\langle -\tfrac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i g_i - \tfrac{\beta_{k+1}}{S_k} d' + A^T \bar{u}_k - \sum\limits_{j=1}^{p} \bar{v}_k^{(j)} \psi_j, \ x - x_{k+1} \ \rangle \ \leq \ 0,$$

для некоторых $d' \in \partial d(x_{k+1})$ и $\psi_j \in \partial F^{(j)}(x_{k+1})$, $j = 1, \dots, p$, и

$$\sum_{j=1}^{p} \bar{v}_k^{(j)} F^{(j)}(x_{k+1}) = 0.$$

Таким образом, для любых таких x получаем

$$\begin{array}{rl} & f(x) + \langle \bar{u}_k, b - Ax \rangle + \langle \bar{v}_k, F(x) \rangle \\ \\ \geq & \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i [f(x_i) + \langle g_i, x - x_i \rangle] + \langle \bar{u}_k, b - Ax \rangle \\ \\ & + \sum\limits_{j=1}^{p} \bar{v}_k^{(j)} [F^{(j)}(x_{k+1}) + \langle \psi_j, x - x_{k+1} \rangle] \\ \\ \geq & \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i [f(x_i) + \langle g_i, x_{k+1} - x_i \rangle] + \frac{\beta_{k+1}}{S_k} \langle d', \ x_{k+1} - x \ \rangle \\ \\ \geq & \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i f(x_i) - \frac{\beta_{k+1}}{S_k} d(x) - \frac{1}{S_k} \left[\sum\limits_{i=0}^{k} \lambda_i \langle g_i, x_i - x_{k+1} \rangle - \beta_{k+1} d(x_{k+1}) \right]. \end{array}$$

Поскольку в определении (2.1.51) необходимо, чтобы $d(x) \leq D$, в силу неравенства (2.1.22) эти оценки приводят к следующему

$$\phi_D(\bar{u}_k, \bar{v}_k) \ \geq \frac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(x_i) - \frac{1}{S_k} \Delta(\beta_{k+1}, D). \tag{2.1.56}$$

Поэтому оценка качества данного прямо-двойственного объекта может быть получена из пункта 1 теоремы 2.1.1.

2.1.4 Седловые задачи

Рассмотрим общую задачу нахождения седловой точки:

$$\min_{u \in U} \max_{v \in V} \ f(u, v), \tag{2.1.57}$$

где U – выпуклое замкнутое множество в пространстве E_u, V – выпуклое замкнутое множество в E_v, функция $f(\cdot, v)$ выпукла по первому аргументу на E_u при всяком $v \in V$ и функция $f(u, \cdot)$ вогнута по второму аргументу на E_v при всяком $u \in U$.

Для множества U имеется прокс-функция $d_u(\cdot)$ с центром u_0, которая строго выпукла на U относительно нормы $\| \cdot \|_u$ с параметром σ_u. Для множества V также имеется аналогичный объект.

Точка $x^* = (u^*, v^*) \in Q \stackrel{\mathrm{def}}{=} U \times V$ является решением задачи (2.1.57) тогда и только тогда, когда

$$f(u^*, v) \leq f(u^*, v^*) \leq f(u, v^*) \quad \forall u \in U, \ v \in V.$$

Поскольку при всех $x \overset{\mathrm{def}}{=} (u,v) \in Q$ выполняются неравенства

$$\begin{array}{rcl} f(u,v^*) & \le & f(u,v) + \langle g_v, v^* - v\rangle \quad \forall g_v \in \partial f_v(u,v), \\[3ex] f(u^*,v) & \ge & f(u,v) + \langle g_u, u^* - u\rangle \quad \forall g_u \in \partial f_u(u,v), \end{array}$$

заключаем, что

$$\langle g_u, u - u^*\rangle + \langle -g_v, v - v^*\rangle \ge 0 \tag{2.1.58}$$

при любом $g = (g_u, g_v) \in \partial f_u(u,v) \times \partial f_v(u,v)$. Таким образом, оракул

$$\mathcal{G}: \quad x = (u,v) \in Q \quad \Rightarrow \quad g(x) = (g_u(x), -g_v(x)) \tag{2.1.59}$$

удовлетворяет условию (2.1.18).

Далее, зафиксируем некоторое $\alpha \in (0,1)$. Тогда можно ввести следующую прокс-функцию множества Q:

$$d(x) = \alpha d_u(u) + (1-\alpha) d_v(v).$$

В этом случае прокс-центром множества Q будет $x_0 = (u_0, v_0)$. Задавая норму на $E = E_u \times E_v$ как

$$\|x\| = \left[\ \alpha\sigma_u \|u\|_u^2 + (1-\alpha)\sigma_v \|v\|_v^2\ \right]^{1/2}.$$

мы получаем, что функция $d(\cdot)$ имеет параметр выпуклости $\sigma = 1$. Норма в двойственном пространстве $E^* = E_u^* \times E_v^*$ задается теперь как

$$\|g\|_* = \left[\ \tfrac{1}{\alpha\sigma_u}\|g_u\|_{u,*}^2 + \tfrac{1}{(1-\alpha)\sigma_v}\|g_v\|_{v,*}^2\ \right]^{1/2}.$$

Предполагая теперь субдифференциалы функции f равномерно ограниченными,

$$\|g_u\|_{u,*} \le L_u, \quad \|g_v\|_{v,*} \le L_v, \quad \forall g \in \partial f_u(u,v) \times \partial f_v(u,v)\ \forall (u,v) \in Q,$$

мы получаем следующую границу на норму ответов оракула $\mathcal{G}$:

$$\|g\|_*^2 \le L^2 \overset{\mathrm{def}}{=} \tfrac{L_u^2}{\alpha\sigma_u} + \tfrac{L_v^2}{(1-\alpha)\sigma_v}. \tag{2.1.60}$$

Наконец, если $d_u(u^*) \le D_u$ и $d_v(v^*) \le D_v$, то $x^* = (u^*, v^*) \in \mathcal{F}_D$, где

$$D = \alpha D_u + (1-\alpha) D_v. \tag{2.1.61}$$

Применим теперь к задаче (2.1.57) ПДУ-метод (2.1.31). В соответствии с теоремой 2.1.2 получаем следующую оценку:

$$\tfrac{1}{S_k}\delta_k(D) \le \tfrac{\hat{\beta}_{k+1}}{k+1}\,\Omega, \quad \Omega \overset{\mathrm{def}}{=} \gamma D + \tfrac{L^2}{2\gamma},$$

где L и D определяются формулами (2.1.60) и (2.1.61) соответственно. При оптимальном выборе $\gamma = \frac{L}{\sqrt{2D}}$ получаем

$$\begin{array}{rcl} \Omega & = & [2L^2 D]^{1/2} = \left[2\left(\tfrac{L_u^2}{\alpha\sigma_u} + \tfrac{L_v^2}{(1-\alpha)\sigma_v}\right)\cdot(\alpha D_u + (1-\alpha)D_v)\right]^{1/2} \\[3ex] & = & \left[2\left(\tfrac{L_u^2 D_u}{\sigma_u} + \tfrac{L_v^2 D_v}{\sigma_v} + \tfrac{\alpha L_v^2 D_u}{(1-\alpha)\sigma_v} + \tfrac{(1-\alpha)L_u^2 D_v}{\alpha\sigma_u}\right)\right]^{1/2}. \end{array}$$

Минимизируя результат по α, имеем

$$\Omega = \sqrt{2}\left(L_u \sqrt{\tfrac{D_u}{\sigma_u}} + L_v \sqrt{\tfrac{D_v}{\sigma_v}} \right).$$

Таким образом, мы показали, что при правильном выборе параметров ПДУ-метод гарантирует следующую скорость убывания функции зазора:

$$\tfrac{1}{S_k}\delta_k(D) \le \tfrac{\hat{\beta}_{k+1}}{k+1}\sqrt{2}\left(L_u \sqrt{\tfrac{D_u}{\sigma_u}} + L_v \sqrt{\tfrac{D_v}{\sigma_v}} \right). \tag{2.1.62}$$

Осталось показать, что убывание зазора гарантирует сходимость к решению седловой задачи (2.1.57).

Рассмотрим две вспомогательные функции:

$$\begin{aligned} \phi_{D_v}(u) &= \max_{v \in V}\{ f(u,v) :\ d_v(v) \le D_v \}, \\ \psi_{D_u}(v) &= \min_{u \in U}\{ f(u,v) :\ d_u(u) \le D_u \}. \end{aligned}$$

В силу наших предположений функция $\phi_{D_v}(\cdot)$ выпукла на U, а функция $\psi_{D_u}(\cdot)$ вогнута на V. Более того, для любых $u \in U$ и $v \in V$ имеем

$$\psi_{D_u}(v) \le f^* \le \phi_{D_v}(u),$$

где $f^* = f(u^*, v^*)$.

Теорема 2.1.6 *Пусть последовательность $\hat{x}_{k+1} = (\hat{u}_{k+1}, \hat{v}_{k+1})$ определена по формулам (2.1.19) с помощью последовательностей X_k, G_k и Λ_k, построенных методом (2.1.24) при решении задачи (2.1.57), снабженной оракулом $\mathcal{G}$, см. соотношение (2.1.59). Тогда $\hat{x}_{k+1} \in Q$ и*

$$0 \le \phi_{D_v}(\hat{u}_{k+1}) - \psi_{D_u}(\hat{v}_{k+1}) \ \le\ \tfrac{1}{S_k}\delta_k(D). \tag{2.1.63}$$

Доказательство.
Действительно, поскольку функция $f(\cdot, v)$ выпукла, а функция $f(u, \cdot)$ вогнута, получаем

$$\begin{aligned} \tau_k &\overset{\text{def}}{=} \tfrac{1}{S_k} \max_{u \in U}\left\{ \sum_{i=0}^{k} \lambda_i \langle g_u(u_i, v_i), u_i - u \rangle :\ d_u(u) \le D_u \right\} \\ &\ge \tfrac{1}{S_k} \max_{u \in U}\left\{ \sum_{i=0}^{k} \lambda_i [f(u_i, v_i) - f(u, v_i)] :\ d_u(u) \le D_u \right\} \\ &\ge \tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(u_i, v_i) - \min_{u \in U}\{ f(u, \hat{v}_{k+1}) :\ d_u(u) \le D_u \} \\ &= \tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(u_i, v_i) - \psi_{D_u}(\hat{v}_{k+1}). \end{aligned}$$

Аналогично,

$$\begin{aligned}\sigma_k &\overset{\rm def}{=} \tfrac{1}{S_k} \max_{v\in V} \left\{ \sum_{i=0}^{k} \lambda_i \langle g_v(u_i,v_i), v-v_i\rangle : \ d_v(v) \le D_v \right\} \\ &\ge \tfrac{1}{S_k} \max_{v\in V} \left\{ \sum_{i=0}^{k} \lambda_i [f(u_i,v) - f(u_i,v_i)] : \ d_v(v) \le D_v \right\} \\ &\ge \max_{v\in V} \{ f(\hat u_{k+1}, v) : \ d_v(v) \le D_v \} - \tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(u_i,v_i) \\ &= \phi_{D_v}(\hat u_{k+1}) - \tfrac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(u_i,v_i).\end{aligned}$$

Таким образом,

$$\phi_{D_v}(\hat u_{k+1}) - \psi_{D_u}(\hat v_{k+1}) \ \le \ \tau_k + \sigma_k$$

$$\le \tfrac{1}{S_k} \max_{x\in Q} \left\{ \sum_{i=0}^{k} \lambda_i \langle g(x_i), x_i - x\rangle : \ d(x) \le \alpha D_u + (1-\alpha) D_v \right\} \ = \ \tfrac{1}{S_k}\delta_k(D). \qquad \Box$$

Заметим, что мы не предполагали ограниченность множества U и V. Параметр D из неравенства (2.1.63) не используется в работе ДУ-метода (2.1.24).

2.1.5 Вариационные неравенства

Пусть выпуклое замкнутое множество Q снабжено, как обычно, прокс-функцией $d(x)$. Рассмотрим непрерывный оператор $V(\cdot) : \ Q \to E^*$, который является *монотонным* на Q:

$$\langle V(x) - V(y), x-y\rangle \ge 0 \quad \forall x,y \in Q.$$

В этом разделе мы рассмотрим методы решения *вариационного неравенства* (ВН):

$$\text{Find } x^* \in Q\text{:} \quad \langle V(x), x - x^*\rangle \ge 0 \quad \forall x \in Q. \tag{2.1.64}$$

Иногда эта задача называется *слабым вариационным неравенством.* Поскольку оператор $V(\cdot)$ непрерывен и монотонен, задача (2.1.64) эквивалентно *сильному вариационному неравенству*

$$\text{найти } x^* \in Q\text{:} \quad \langle V(x^*), x - x^*\rangle \ge 0 \quad \forall x \in Q. \tag{2.1.65}$$

Мы предполагаем что решение этой задачи x^* существует.

Зафиксируем число $D > d(x^*)$. Стандартной мерой качества приближенного решения задачи (2.1.64) является следующая *функция близости*

$$f_D(x) = \max_{y\in Q} \{ \langle V(y), x-y\rangle : \ d(y) \le D\}. \tag{2.1.66}$$

Приведем два важных свойства этой функции.

Лемма 2.1.4 *Функция* $f_D(x)$ *корректно определена и выпукла на* E. *Для любого* $x \in Q$ *имеем* $f_D(x) \geq 0$ *и* $f_D(x^*) = 0$. *Обратно, если* $f_D(\hat{x}) = 0$ *при некотором* $\hat{x} \in Q$ *и* $d(\hat{x}) < D$, *то точка* $\hat{x}$ *является решением задачи (2.1.64).*

Доказательство.
Действительно, функция $f_D(x)$ корректно определена, поскольку множество $\mathcal{F}_D$ ограничено. Она выпукла по x как максимум параметрического семейства линейных функций. Если $x \in \mathcal{F}_D$, то, выбирая в (2.1.66) $y = x$, мы гарантируем, что $f_D(x) \geq 0$. Пусть $x \in Q$ и $d(x) > D$. Рассмотрим точку $y \in Q$, удовлетворяющую условию

$$d(y) = D, \quad y = \alpha x + (1 - \alpha)x^*$$

при некотором $\alpha \in (0, 1)$. Тогда $y \in \mathcal{F}_D$ и

$$\langle V(y), x - y \rangle = \tfrac{1-\alpha}{\alpha} \langle V(y), y - x^* \rangle \geq 0.$$

Таким образом, $f_D(x) \geq 0$ при всех $x \in Q$. С другой стороны, поскольку x^* является решением задачи (2.1.64), $\langle V(y), x^* - y \rangle \leq 0$ для всех $y \in \mathcal{F}_D$. Поэтому, так как $x^* \in \mathcal{F}_D$, получаем $f_D(x^*) = 0$.

Наконец, пусть $f_D(\hat{x}) = 0$ для некоторого $\hat{x} \in Q$ и $d(\hat{x}) < D$. Это означает, что точка $\hat{x}$ является решением слабого вариационного неравенства $\langle V(y), y - \hat{x} \rangle \geq 0$ при всех $y \in \mathcal{F}_D$. Поскольку оператор $V(y)$ непрерывен, мы заключаем, что $\hat{x}$ есть решение сильного вариационного неравенства

$$\langle V(\hat{x}), y - \hat{x} \rangle \geq 0 \quad \forall y \in \mathcal{F}_D.$$

Поэтому, минимум линейной функции $l(y) = \langle V(\hat{x}), y \rangle$, $y \in \mathcal{F}_D$, достигается в $y = \hat{x}$. Более того, в этой точке ограничение $d(y) \leq D$ не активно. Таким образом, это ограничение можно отбросить и мы заключаем, что точка $\hat{x}$ является решением задачи (2.1.65) и, следовательно задачи (2.1.64). □

Давайте воспользуемся для решения задачи (2.1.64) оракулом $\mathcal{G}:\ x \to V(x)$.

Лемма 2.1.5 *При любом* $k \geq 0$ *выполняется неравенство*

$$f_D(\hat{x}_{k+1}) \leq \tfrac{1}{S_k} \delta_k(D). \tag{2.1.67}$$

Доказательство.
Действительно, поскольку оператор $V(x)$ монотонен, имеем

$$\begin{aligned} f_D(\hat{x}_{k+1}) &= \max_x \{ \langle V(x), \hat{x}_{k+1} - x \rangle :\ x \in \mathcal{F}_D \} \\ &= \tfrac{1}{S_k} \max_x \left\{ \sum_{i=0}^{k} \lambda_i \langle V(x), x_i - x \rangle :\ x \in \mathcal{F}_D \right\} \\ &\leq \tfrac{1}{S_k} \max_x \left\{ \sum_{i=0}^{k} \lambda_i \langle V(x_i), x_i - x \rangle :\ x \in \mathcal{F}_D \right\} \equiv \tfrac{1}{S_k} \delta_k(D). \qquad \square \end{aligned}$$

Предположим, что $\|V(x)\|_* \leq L$ при всех $x \in Q$. Тогда, например, из теоремы 2.1.2 заключаем, что ПДУ-метод (2.1.31), применяемый к задаче (2.1.64), сходится со следующей скоростью:

$$f_D(\hat{x}_{k+1}) \leq \frac{\hat{\beta}_{k+1}}{k+1}\left(\gamma d(x^*) + \frac{L^2}{2\sigma\gamma}\right). \tag{2.1.68}$$

Заметим, что последовательность $\{x_k\}_{k=0}^{\infty}$ остается ограниченной даже в случае неограниченного множества Q:

$$\|x_k - x^*\|^2 \ \leq \ \frac{2}{\sigma}d(x^*) + \frac{L^2}{\sigma^2\gamma^2}. \tag{2.1.69}$$

2.1.6 Стохастическая оптимизация

Пусть вероятностное пространство Ξ снабжено вероятностной мерой, и пусть для выпуклого замкнутого множества Q имеется прокс-функция $d(x)$. Рассмотрим функцию цены

$$f: \ Q \times \Xi \mapsto R.$$

Это отображение определяет семейство случайных переменных $f(x, \xi)$ на Ξ. Мы предполагаем, что математическое ожидание по ξ, $\mathbf{E}_\xi(f(x,\xi))$, корректно определено при всех $x \in Q$. Задача стохастической оптимизации ставится следующим образом:

$$\min_x \left\{\phi(x) \equiv \mathbf{E}_\xi(f(x,\xi)) : \ x \in Q\right\}. \tag{2.1.70}$$

Предположим, что задача (2.1.70) разрешима. Обозначим через x^* одно из ее решений и пусть $\phi^* \overset{\text{def}}{=} \phi(x^*)$. В этом пункте мы предполагаем следующее.

Предположение 2.1.1 При любом $\xi \in \Xi$ функция $f(\cdot, \xi)$ является выпуклой и субдифференцируемой на Q.

Тогда функция $\phi(x)$ также является выпуклой на Q.

Заметим, что вычисление точного значения $\phi(x)$ может быть затруднительным, даже если известно распределение ξ. Действительно, если $\Xi \subset \mathbb{R}^m$, то вычисление такого значения $\hat{\phi}$, что $\mid \hat{\phi} - \phi(x) \mid \leq \hat{\epsilon}$, может потребовать до $O\left(\frac{1}{\hat{\epsilon}^m}\right)$ вычислений значений $f(x,\xi)$ при различных $\xi \in \Xi$. Поэтому стандартная детерминистическая аппроксимация решения задачи (2.1.70) является слишком сложной.

Однако возможно альтернативное определение приближенного решения задачи (2.1.70). Действительно, понятно, что никакая концепция решения не исключает сбоя при его реализации. При принятии решений в условиях неопределенности естественным образом отсутствует полный контроль над возможными последствиями. В такой ситуации, необязательно требовать, чтобы вычисляемое решение было детерминистическим. Рандомизированные алгоритмы также являются приемлемыми при условии, что их результаты удовлетворяют определенным критериям качества. А именно, пусть точка x является случайным результатом работы некоторого случайного алгоритма. Оказывается, достаточно предположить, что этот результат хорош только *в среднем*:

$$\mathbf{E}(\phi(x)) - \phi^* \leq \epsilon. \tag{2.1.71}$$

Тогда можно показать [100], что любой такой случайный алгоритм может быть превращен в алгоритм, генерирующий решения с определенным уровнем надежности.

В этом пункте мы показываем, что стохастическая версия ПДУ-метода (2.1.31), применяемая к задаче (2.1.70), также в состоянии генерировать случайную точку $\tilde{x} \in Q$, удовлетворяющую условию (2.1.71). Для этого нам потребуется одно дополнительное предположение.

Предположение 2.1.2 В любой точке $x \in Q$ при любой реализации случайного вектора $\xi \in \Xi$ ответы оракула $\mathcal{G}(x, \xi)$, $\nabla f(x, \xi) \in \partial_x f(x, \xi)$, равномерно ограничены:

$$\| \nabla f(x, \xi) \|_* \ \le \ L \quad \forall x \in Q, \ \forall \xi \in \Xi.$$

Выпишем стохастическую версию ПДУ-метода. Через $\tilde{\tau}$ будем обозначать реализацию случайной переменной τ. Таким образом, все $\tilde{\xi}_k \in \mathbb{R}^m$ в приводимой схеме обозначают реализации независимых одинаково распределенных случайных векторов ξ_k, соответствующих случайному вектору ξ из задачи (2.1.70).

Метод стохастических двойственных усреднений

Инициализация Полагаем $\tilde{s}_0 = 0 \in E^*$. Выбираем $\gamma > 0$.

Итерация ($k \ge 0$) (2.1.72)

1. Генерируем $\tilde{\xi}_k$ и вычисляем $\tilde{g}_k = \nabla f(\tilde{x}_k, \tilde{\xi}_k)$.

2. Полагаем $\tilde{s}_{k+1} = \tilde{s}_k + \tilde{g}_k, \ \tilde{x}_{k+1} = \pi_{\gamma \hat{\beta}_{k+1}}(-\tilde{s}_{k+1})$.

Обозначим через $\boldsymbol{\xi}_k$, $k \ge 0$, набор случайных переменных $(\xi_0, \dots, \xi_k)$. Метод (2.1.72) генерирует две последовательности случайных векторов

$$s_{k+1} = s_{k+1}(\boldsymbol{\xi}_k) \in E^*, \quad x_{k+1} = x_{k+1}(\boldsymbol{\xi}_k) \in Q, \quad k \ge 0,$$

с фиксированными начальными точками s_0 и x_0. Заметим, что их реализации $\tilde{s}_{k+1}$ и $\tilde{x}_{k+1}$ при разных запусках этого метода будут различными.

Таким образом, для каждого запуска метода можно определить функцию зазора

$$\tilde{\delta}_k(D) = \max_x \left\{ \sum_{i=0}^{k} \langle \nabla f(\tilde{x}_i, \tilde{\xi}_i), \tilde{x}_i - x \rangle \, : \, x \in \mathcal{F}_D, \right\}, \quad D \ge 0, \tag{2.1.73}$$

которая порождает случайную величину $\delta_k(D)$. Поскольку $\|\tilde{g}_k\|_* \le L$,теорема 2.1.2дает следующую равномерную оценку:

$$\tfrac{1}{S_k} \tilde{\delta}_k(D) \ \le \ \tfrac{\hat{\beta}_{k+1}}{k+1} \left(\gamma D + \tfrac{L^2}{2\sigma\gamma} \right). \tag{2.1.74}$$

Выберем параметр D достаточно большим: $D \geq d(x^*)$. Тогда

$$\tfrac{1}{S_k}\tilde{\delta}_k(D) \;\; \geq \;\; \tfrac{1}{k+1}\sum_{i=0}^{k}\langle \nabla f(\tilde{x}_i, \tilde{\xi}_i), \tilde{x}_i - x^* \rangle \; \geq \; \tfrac{1}{k+1}\sum_{i=0}^{k}[f(\tilde{x}_i, \tilde{\xi}_i) - f(x^*, \tilde{\xi}_i)] \,.$$

Поскольку все ξ_i независимы и одинаково распределены, в среднем мы получаем

$$\mathbf{E}_{\boldsymbol{\xi}_k}\left(\tfrac{1}{S_k}\delta_k(D)\right) \;\; \geq \;\; \tfrac{1}{k+1}\sum_{i=0}^{k}\mathbf{E}_{\boldsymbol{\xi}_k}\left(f(x_i, \xi_i)\right) - \phi^*. \tag{2.1.75}$$

Заметим, что случайный вектор x_i независим относительно всех ξ_j, $i \leq j \leq k$. Таким образом,

$$\mathbf{E}_{\boldsymbol{\xi}_k}\left(f(x_i, \xi_i)\right) = \mathbf{E}_{\boldsymbol{\xi}_k}\left(\phi(x_i)\right),$$

и мы заключаем, что

$$\tfrac{1}{k+1}\sum_{i=0}^{k}\mathbf{E}_{\boldsymbol{\xi}_k}\left(f(x_i, \xi_i)\right) \;\; = \;\; \tfrac{1}{k+1}\sum_{i=0}^{k}\mathbf{E}_{\boldsymbol{\xi}_k}\left(\phi(x_i)\right) \; \geq \; \mathbf{E}_{\boldsymbol{\xi}_k}\left(\phi\left(\tfrac{1}{k+1}\sum_{i=0}^{k}x_i\right)\right).$$

Итак, в силу неравенств (2.1.74), (2.1.75) мы доказали следующую теорему.

Теорема 2.1.7 *Пусть случайная последовательность точек* $\{x_k\}_{k=0}^{\infty}$ *генерируется методом (2.1.72). Тогда при любом* $k \geq 0$ *выполняются неравенства*

$$\mathbf{E}_{\boldsymbol{\xi}_k}\left(\phi\left(\tfrac{1}{k+1}\sum_{i=0}^{k}x_i\right)\right) \; - \; \phi^* \;\; \leq \;\; \tfrac{\hat{\beta}_{k+1}}{k+1}\left(\gamma d(x^*) + \tfrac{L^2}{2\sigma\gamma}\right).$$

2.1.7 Приложения в моделировании

Алгоритм сбалансированного развития

Оказывается, вариант ДУ-метода (2.1.24) имеет естественную интерпретацию рациональной динамической стратегии инвестиций. В этом пункте мы опишем соответствующую модель.

Предположим, что мы собираемся развивать нашу собственность с помощью заранее заданной последовательности инвестиций капитала $\Lambda = \{\lambda_k\}_{k=0}^{\infty}$. Эти деньги могут использоваться на покупку определенных продуктов с единичными ценами

$$p^{(j)} \geq 0, \quad j = 1, \ldots, N.$$

Все продукты неограниченно делимы и доступны на рынке в любых количествах. Таким образом, если для k-го вложения капитала мы покупаем количество $X_k^{(j)} \geq 0$ продукта j, $j = 1, \ldots, N$, то балансовое уравнение запишется как

$$X_k = \lambda_k x_k, \quad \langle p, x_k \rangle = 1, \quad x_k \geq 0, \tag{2.1.76}$$

где векторы $p = (p^{(1)}, \ldots, p^{(N)})^T$ и $x_k = (x_k^{(1)}, \ldots x^{(N)})^T$ лежат в $\mathbb{R}^N$.

Для отслеживания прогресса в развитии нашей собственности (или эффективности наших капиталовложений) мы введем m характеристик (свойств). Предположим, что после $k \geq 0$ капиталовложений мы можем *точно* измерить накопленный уровень $s_k^{(i)}$, $i = 1, \dots, m$, каждого свойства i. С другой стороны, у нас имеется набор стандартов $b^{(i)} > 0$, $i = 1, \dots, m$, которые понимаются как нижние границы на концентрацию каждого свойства в *идеально сбалансированном* хозяйстве стандартного размера.

Наконец, предположим, что качество продукта j описывается вектором

$$a_j = (a_j^{(1)}, \dots, a_j^{(m)})^T, \quad j = 1, \dots, N,$$

координаты которого дают концентрацию соответствующего свойства в единице товара. Вводя матрицу $A = (a_1, \dots, a_N) \in \mathbb{R}^{m \times N}$, мы можем описать динамику накопления свойств следующим образом:

$$s_{k+1} = s_k + \lambda_k g_k, \quad g_k = Ax_k, \quad k \geq 0. \qquad (2.1.77)$$

В соответствии с нашей системой стандартов стратегия *оптимально сбалансированного развития* может быть найдена как решение следующей задачи оптимизации:

$$\tau^* \overset{\text{def}}{=} \max_{x,\tau} \{\tau : \ Ax \geq \tau b, \ \langle p, x \rangle = 1, \ x \geq 0\}. \qquad (2.1.78)$$

Действительно, обозначим через x^* ее оптимальное решение. Тогда, выбирая в соотношениях (2.1.77) $x_k = x^*$, получаем

$$s_k = S_k \cdot Ax^* \ \geq \ S_k \cdot \tau^* b.$$

Таким образом, эффективность такого капиталовложения максимальна, и его качество не зависит даже от графика инвестиций.

Однако заметим, что для реализации этой стратегии необходимо заранее решить задачу линейного программирования (2.1.78), которая может быть сложной ввиду большого размера. Что делать, если это действительно так? Есть ли у нас какие-либо алгоритмы для реализации оптимальной стратегии в реальной жизни? Ответ на последний вопрос оказывается положительным.

Прежде всего перепишем задачу (2.1.78) в двойственной форме. С помощью функции Лагранжа

$$\mathcal{L}(x, \tau, y, \lambda) \ = \ \tau + \langle Ty, Ax - \tau b \rangle + \lambda \cdot (1 - \langle p, x \rangle)$$

где T – диагональная матрица, $T^{(i,i)} = \frac{1}{b^{(i)}}$, $i = 1, \dots, m$, и соответствующей седловой задачи

$$\max_{\tau;\, x \geq 0} \ \min_{\lambda;\, y \geq 0} \mathcal{L}(x, \tau, y, \lambda)$$

мы приходим к следующей двойственной постановке:

$$\tau^* \ = \ \min_{y,\lambda} \{\lambda : \ A^T T y \leq \lambda p, \ y \in \Delta_m\}.$$

Таким образом, вводя функцию $\phi(y) = \max\limits_{1 \le j \le N} \frac{1}{p^{(j)}} \langle a_j, Ty \rangle$, получаем

$$\tau^* \ = \ \min_y \{ \phi(y) : \ y \in \Delta_m \}. \tag{2.1.79}$$

Применим к задаче (2.1.79) одну из версий ДУ-метода (2.1.24). Для допустимого множества этой задачи Δ_m выберем энтропийную прокс-функцию

$$d(y) \ = \ \ln m + \sum_{i=1}^{m} y^{(i)} \ln y^{(i)}$$

с прокс-центром $y_0 = (\frac{1}{m}, \dots, \frac{1}{m})^T$. Напомним, что $d(y)$ сильно выпукла на Δ_m в l_1-норме (2.1.10) с параметром выпуклости $\sigma = 1$ и при всех $y \in \Delta_m$ выполняется неравенство

$$d(y) \ \le \ \ln m. \tag{2.1.80}$$

Более того, при такой прокс-функции вычисление проекции $\pi_\beta(s)$ очень простое:

$$\pi_\beta(s)^{(i)} \ \ = \ \ e^{s^{(i)}/\beta} \cdot \left[\sum_{j=1}^{m} e^{s^{(j)}/\beta} \right]^{-1}, \quad i = 1, \dots, m. \tag{2.1.81}$$

В приводимом алгоритме последовательность Λ такая же, как и в формулах (2.1.77).

Алгоритм сбалансированного развития

Инициализация Полагаем $s_0 = 0 \in E^*$. Выберем $\beta_0 > 0$.

Итерация $(k \ge 0)$

1. Формируем множество $J_k = \left\{ j : \ \frac{1}{p^{(j)}} \langle a_j, Ty_k \rangle = \phi(y_k) \right\}$.

2. Выбираем любое $x_k \ge 0$, $\langle p, x_k \rangle = 1$, $x_k^{(j)} = 0$ для $j \notin J_k$.

3. Пересчитываем $s_{k+1} = s_k + \lambda_k g_k$, $g_k = A x_k$.

4. Выбираем $\beta_{k+1} \ge \beta_k$ и для $i = 1, \dots, m$ определяем

$$y_{k+1}^{(i)} \ = \ e^{-s_{k+1}^{(i)}/(\beta_{k+1} b^{(i)})} \cdot \left[\sum_{j=1}^{m} e^{-s_{k+1}^{(j)}/(\beta_{k+1} b^{(j)})} \right]^{-1}.$$

(2.1.82)

Заметим, что в этой схеме $Tg_k \in \partial \phi(y_k)$ и

$$y_{k+1} = \pi_{\beta_{k+1}} \left(-T \sum_{j=0}^{k} \lambda_j g_j \right).$$

Таким образом, это действительно вариант метода (2.1.24), и в силу теоремы 2.1.1 и неравенства (2.1.80) имеем

$$\begin{array}{rcl} \frac{1}{S_k}\sum\limits_{i=0}^{k}\lambda_i\phi(y_i)-\tau^* & \leq & \frac{1}{S_k}\left(\beta_{k+1}\ln m+\frac{1}{2}L^2\sum\limits_{i=0}^{k}\frac{\lambda_i^2}{\beta_i}\right), \\ L \; = \; \max\limits_{1\leq j\leq N}\frac{1}{p^{(j)}}\|Ta_j\|_* & = & \max\limits_{\substack{1\leq j\leq N,\\ 1\leq i\leq m}}\frac{1}{p^{(j)}b^{(i)}}\,|a_j^{(i)}|. \end{array} \qquad (2.1.83)$$

Алгоритм (2.1.82) допускает следующую интерпретацию. Его первые три шага представляют собой естественную стратегию инвестиций. Действительно, алгоритм пересчитывает систему персональных цен для характеристик

$$u_k \overset{\text{def}}{=} Ty_k, \quad k\geq 0.$$

С помощью этих цен каждый продукт j получает персональную оценку его полезности $\langle a_j,u_k\rangle$. Таким образом, на Шаге 1 мы формируем множество продуктов с наилучшим соотношением *цена/качество*:

$$J_k=\left\{j:\;\frac{1}{p^{(j)}}\langle a_j,u_k\rangle=\max_{1\leq i\leq N}\frac{1}{p^{(i)}}\langle a_i,u_k\rangle\right\}.$$

На шаге 2 мы распределяем единицу бюджета произвольным образом среди наилучших продуктов. На шаге 3 мы покупаем эти продукты, используя бюджет λ_k, и пересчитываем вектор накопленных качеств. Нетривиальная часть интерпретации относится к шагу 4, где описывается эволюция персональных цен, возникающих к концу периода $(k+1)$, в зависимости от вектора накопленных характеристик s_{k+1}, наблюдаемых в течение этого периода. Для этой интерпретации надо объяснить смысл формул (2.1.81).

Соотношения (2.1.81) возникают в так называемой *логистической* модели в ситуации дискретного выбора (см., например, [26]). В этой модели необходимо выбрать один из m вариантов, имеющих полезность $s^{(i)}$, $i=1,\ldots,m$. Эти полезности наблюдаются со случайными аддитивными ошибками $\epsilon^{(i)}$, которые независимы и одинаково распределены в соответствии с двойной экспонентой:

$$F(\tau)=\mathbf{Prob}\left[\epsilon^{(i)}\leq\tau\right]=\exp\left\{-e^{-\gamma-\tau/\beta}\right\}, \quad i=1,\ldots,m,$$

где $\gamma\approx 0,5772$ – константа Эйлера и параметр $\beta>0$ отвечает за чувствительность к ошибкам. Тогда математическим ожиданием от функции полезности будет

$$U_\beta(s)=E\left(\max_{1\leq i\leq m}[s^{(i)}+\epsilon^{(i)}]\right),$$

Можно показать, что $U_\beta(s)=\beta d_*(s/\beta)+\text{const}$, где

$$d_*(s)=\ln\left(\sum_{i=0}^{m}e^{s^{(i)}}\right).$$

Таким образом, мы получаем следующие важные формулы для вероятностей выбора:

$$\mathbf{Prob}\left[s^{(i)}+\epsilon^{(i)}=\max_{1\le j\le m}[s^{(j)}+\epsilon^{(j)}]\right] \;=\; \nabla U_\beta(s)^{(i)}$$

$$=\; \nabla d_*(s/\beta)^{(i)} \;=\; e^{s^{(i)}/\beta}\cdot\left[\sum_{j=1}^{m} e^{s^{(j)}/\beta}\right]^{-1}, \quad i=1,\dots,m,$$

(ср. (2.1.81)). Чем меньше β, тем ближе эта схема к детерминистическому максимуму.

Используя логистическую модель, можно снабдить шаг 4 в алгоритме (2.1.82) следующей поведенческой интерпретацией.

- В течение $(k+1)$-го периода мы регулярно сравниваем уровень накопленных свойств со стандартом b. Каждая проверка выявляет наиболее дефицитную характеристику.
- Вышеупомянутые проверки проводятся с аддитивными ошибками[4], которые соответствуют логистической модели.
- В конце периода мы получаем частоты $y^{(i)}_{k+1}$, с которыми наши характеристики фиксировались как наиболее дефицитные. Они и дают нам новые цены:

$$u^{(i)}_{k+1} \quad=\quad \tfrac{1}{b^{(i)}}\, y^{(i)}_{k+1}, \quad i=1,\dots,m. \tag{2.1.84}$$

Заметим, что условия сходимости нашего метода могут быть получены из анализа правой части неравенства (2.1.83). Предположим, например, что у нас имеется последовательность одинаковых инвестиций объемом λ. Далее, в течение каждого периода мы сравниваем средние накопленные характеристики, используя при этом логистическую модель с параметром μ. Поэтому в методе (2.1.82) мы берем

$$\beta_{k+1}=\mu\cdot(k+1), \quad \lambda_k=\lambda, \quad k\ge 0.$$

Полагая $\beta_0=\mu$, можно оценить правую часть неравенства (2.1.83):

$$\tfrac{\mu}{\lambda}\ln m+\tfrac{\lambda L^2}{2\mu}\cdot\tfrac{2+\ln k}{k+1}, \quad k\ge 1.$$

Таким образом, качество нашей собственности стабилизируется на уровне $\tau^*+\frac{\mu}{\lambda}\ln m$. Для достижения сходимости к оптимальному балансу уровень неточности μ в проверках качества должен постепенно стремиться к нулю.

---

[4] Эти ошибки могут возникать естественным образом в силу неточной информации или ошибок измерения. Однако, даже если наши измерения точны, полезно вносить в них *искусственную рандомизацию.*

Предварительное планирование в сравнении с динамической настройкой

Рассмотрим мультистадийную задачу оптимизации в следующей постановке. Для k последовательных периодов, $k = 0, \dots, N$, наши затраты описываются разными выпуклыми целевыми функциями

$$f_0(x), \dots, f_N(x), \quad x \in Q,$$

где Q – выпуклое замкнутое множество. Таким образом, если мы собираемся применять в период k стратегию $x_k \in Q$, то общие затраты будут составлять

$$\Psi(X) = \sum_{k=0}^{N} f_k(x_k), \quad X = \{x_k\}_{k=0}^{N}.$$

В такой ситуации есть несколько возможностей для определения рациональности.

1. Если функции $\{f_k(x)\}_{k=0}^{N}$ известны заранее, то можно применить оптимальную *динамическую* стратегию X^*:

$$x_k \; = \; x_k^* \; \stackrel{\mathrm{def}}{=} \; \arg\min_x \left\{ f_k(x) : \; x \in Q \right\}, \quad k = 0, \dots, N. \tag{2.1.85}$$

Понятно что в этом случае наш выигрыш максимален. Однако у этой стратегии есть и недостатки. Во-первых, она требует больших вычислительных затрат, поскольку приходится решать $N + 1$ задачу оптимизации. Во-вторых, требуется каким-то образом узнать целевые функции из будущего. И наконец, у нас нет никаких гарантий малости расстояния между последовательными стратегиями x_k и x_{k+1}. Если это расстояние велико, то могут потребоваться дополнительные затраты на перестройку, которые не вошли в модель.

2. Менее претенциозная стратегия состоит в использовании следующего оптимального решения:

$$x^* \; \stackrel{\mathrm{def}}{=} \; \arg\min_x \left\{ \bar{f}_N(x) \stackrel{\mathrm{def}}{=} \tfrac{1}{N+1} \sum_{k=0}^{N} f_k(x) : \; x \in Q \right\}. \tag{2.1.86}$$

Здесь в каждый период мы применяем один и тот же вариант $x_k = x^*$, $k = 0, \dots, N$. Такая стратегия X_* называется оптимальной *статической* стратегией. Ясно, что ее вычисление и реализация гораздо проще, чем у стратегии X^*. Однако для нее все равно требуется знать целевые функции заранее.

3. Наконец, последняя возможность состоит в запуске процесса настройки, который будет каким-то образом генерировать последовательность X, принимая во внимание полученную в прошлом информацию. В этом случае никакая целевая функция заранее не известна.

Ясно, что в последней ситуации трудно ожидать многого. Тем не менее, давайте сравним ее возможную эффективность с эффективностью стратегии X_*. Предположим, что стратегия X генерируется следующим вариантом ПДУ-метода (2.1.31).

Инициализация Полагаем $s_0 = 0 \in E^*$. Выбираем $\gamma > 0$.

Итерации $(k = 0, \ldots, N)$ (2.1.87)

1. Вычисляем $g_k \in \partial f_k(x_k)$ и полагаем $s_{k+1} = s_k + g_k$.

2. Выбираем $\beta_{k+1} = \gamma \hat{\beta}_{k+1}$ и полагаем $x_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1})$.

Предположим, что при любом $k = 0, \ldots, N$ выполнено неравенство

$$\| g_k \|_* \le L \quad \forall g_k \in \partial f_k(x), \, \forall x \in Q.$$

Тогда в силу теоремы 2.1.2 получаем следующую оценку:

$$\tfrac{1}{N+1} \delta_N(D) \;\le\; \tfrac{\hat{\beta}_{N+1}}{N+1} \left(\gamma d(x^*) + \tfrac{L}{2\sigma\gamma} \right).$$

С другой стороны, для $D \ge d(x^*)$ имеем

$$\begin{aligned} \tfrac{\delta_N(D)}{N+1} \;&=\; \tfrac{1}{N+1} \max_x \left\{ \sum_{k=0}^{N} \langle g_k, x_k - x \rangle : \; x \in \mathcal{F}_D \right\} \\ &\ge\; \tfrac{1}{N+1} \max_x \left\{ \sum_{k=0}^{N} [f_k(x_k) - f_k(x)] \rangle : \; x \in \mathcal{F}_D \right\} \\ &=\; \tfrac{1}{N+1} \sum_{k=0}^{N} f_k(x_k) - \min_x \left\{ \bar{f}_N(x) : \; x \in Q \right\} \;=\; \tfrac{1}{N+1} \Psi(X) - \bar{f}_N(x^*). \end{aligned}$$

Таким образом,

$$\tfrac{1}{N+1} \Psi(X) \;\le\; \bar{f}_N(x^*) + \tfrac{\hat{\beta}_{N+1}}{N+1} \left(\gamma d(x^*) + \tfrac{L}{2\sigma\gamma} \right), \qquad (2.1.88)$$

и мы приходим к следующему выводу.

> *При больших N средняя эффективность динамической настройки по правилам (2.1.87) неограниченно приближается к эффективности оптимальной статической стратегии предварительного планирования в условиях* <u>*полной*</u> *информации о будущем.*

2.1.8 Обсуждение

В заключение сравним предложенные схемы двойственных усреднений с традиционными методами оптимизации.

1. Стратегия расходящегося ряда. Существует много способов определения последовательностей $\{\lambda_k\}_{k=0}^{\infty}$ и $\{\beta_k\}_{k=0}^{\infty}$, которые обеспечивают правильную скорость убывания дроби $\frac{1}{S_k}\delta_k(D)$. Выбор

$$\lambda_k = 1, \quad \beta_k = \gamma\hat{\beta}_k, \quad k \ge 0,$$

приводит к оценке $\frac{1}{S_k}\delta_k(D) = O(\frac{1}{\sqrt{k}})$. Это наилучшая скорость сходимости, возможная для черноящичных методов при решении задач негладкой минимизации неограниченного размера [79].

Посмотрим, что происходит при других способах. Возьмем, например,

$$\lambda_k = (k+1)^p, \quad \beta_k = \gamma(k+1)^q, \quad k \ge 0.$$

Заметим, что при $\lambda > -1$ справедлива оценка

$$\sum_{i=0}^{k}(i+1)^\lambda \; = \; O\left(\tfrac{1}{1+\lambda}(k+1)^{1+\lambda}\right).$$

В силу неравенства (2.1.25) нам нужно сравнить следующие величины:

$$\begin{aligned} \beta_{k+1} \;&\approx\; (k+1)^q, \\ \sum_{i=0}^{k}\lambda_i \;&\approx\; \tfrac{1}{1+p}(k+1)^{1+p}, \\ \sum_{i=0}^{k}\tfrac{\lambda_i^2}{\beta_i} \;&\approx\; \tfrac{1}{1+2p-q}(k+1)^{1+2p-q}. \end{aligned}$$

Таким образом, скорость сходимости ДУ-методов оказывается такой:

$$\max\left\{(1+p)(k+1)^{q-p-1}, \tfrac{1+p}{1+2p-q}(k+1)^{p-q}\right\}.$$

Поэтому оптимальное соотношение между степенями: $q - p = \frac{1}{2}$. Тогда функцию зазора можно оценить как

$$O\left(\max\left\{1, \tfrac{1}{p+0.5}\right\}\tfrac{p+1}{\sqrt{k+1}}\right). \qquad (2.1.89)$$

Таким образом, при любом $p > -\frac{1}{2}$ можно гарантировать оптимальный порядок скорости сходимости. Когда $p \downarrow -\frac{1}{2}$, константа в формуле (2.1.89) неограниченно растет. Тем не менее, для $p = -\frac{1}{2}$ можно гарантировать следующую скорость:

$$O\left(\tfrac{1+\ln(k+1)}{\sqrt{k+1}}\right).$$

В этом варианте мы получаем

$$\lambda_k = \tfrac{1}{\sqrt{k+1}}, \quad \beta_k = \gamma, \quad k \ge 0. \qquad (2.1.90)$$

Таким образом, метод (2.1.24) сводится к оптимальному варианту метода зеркального спуска с шагами, определяемыми расходящимся рядом. В ПДУ-методе (2.1.31) мы выбираем

$$\lambda_k = 1, \quad \beta_k \approx \gamma\sqrt{k+1}, \quad k \geq 0.$$

Заметим, что в технике сглаживания (см. гл. 5), которая основывается на быстрых градиентных методах, применяется гораздо более агрессивная стратегия усреднения:

$$\lambda_k \approx k+1, \quad k \geq 0.$$

Интересно, что в течение многих лет в стандартной теории выпуклой минимизации предлагалось пользоваться *наихудшим* способом выбора параметров (2.1.90) (см., например, п. 3.2.3 в [83]).

2. Методы с негладкой моделью (МНМ). Схема (2.1.36) имеет некоторые общие черты с методами, использующими негладкую модель. Однако отличий гораздо больше. Действительно,

- в МНМ прокс-центр сдвигается после каждой “успешной” итерации, а в (2.1.36) прокс-центр зафиксирован;
- в МНМ шкалирующий параметр μ_k увеличивается с целью нахождения лучшего значения целевой функции, а в методе (2.1.36) этот параметр уменьшается, что обеспечивает быстрое движение в прямом пространстве;
- наконец, оптимальная оценка сложности $O(\frac{1}{\epsilon^2})$ доказана только для одного МНМ - метода уровней, который сильно отличается от других МНМ. Для метода (2.1.36) эта оценка следует из неравенства (2.1.35).

В то же время на практике МНМ обычно сходятся очень быстро. Это объясняется использованием гораздо более точной кусочно-линейной модели. С точки зрения анализа по наихудшему случаю такая точность не нужна. Тем не менее, представляется интересным включить элементы негладкой модели в конструкцию (2.1.36).

2.2 Барьерный субградиентный метод

2.2.1 Введение

Мотивировка

В нелинейной оптимизации, работоспособность численных методов сильно зависит от наших возможностей решать вспомогательные оптимизационные задачи на допустимых множествах. Обычно при разработке методов предполагается, что мы в состоянии выполнить одну из следующих операций.

L. Максимизация линейной функции $\langle c, x \rangle$ на выпуклом множестве Q.

S. Максимизация разности $\langle c, x \rangle - d(x)$ по $x \in Q$, где d – сильно выпуклая прокс-функция множества Q.

B. Вычисление значения и двух первых производных некоторого самосогласованного барьера во внутренних точках множества Q.

Заметим, что в структурной оптимизации мы всегда рассматриваем оптимимизационные задачи в прямо-двойственной форме. Наиболее важным примером такого представления является билинейное седло

$$\min_{x \in Q_p} \max_{w \in Q_d} \langle Ax, w \rangle, \tag{2.2.1}$$

где множества Q_p и Q_d выпуклы и замкнуты в соответствующих пространствах и A – линейный оператор. Поскольку структура прямого и двойственного множеств может быть разной по сложности, мы получаем шесть возможных комбинаций реализуемых вспомогательных задач. Выпишем известные результаты о сложности соответствующих задач оптимизации.

- $\mathbf{L_p} \bigcap \mathbf{L_d}$. Насколько нам известно, такие задачи никто решать не умеет.
- $\mathbf{S_p} \bigcap \mathbf{S_d}$. Этот случай допускает применение техники сглаживания (см. гл. 5). Для него ϵ-решение задачи (2.2.1) может быть получено за

  $$O\left(\tfrac{1}{\epsilon} \cdot \|A\| \cdot [D_1 D_2]^{1/2} \right)$$

  градиентных итераций, где D_1 и D_2 – размеры прямого и двойственного множеств, а норма оператора $\|A\|$ вычислена относительно соответствующих норм прямого и двойственного пространств.
- $\mathbf{B_p} \bigcap \mathbf{B_d}$. В этом случае, методы внутреней точки позволяют вычислить ϵ-решение задачи (2.2.1) за

  $$O\left(\sqrt{\nu} \cdot \ln \tfrac{\nu}{\epsilon} \right)$$

  итераций метода Ньютона, где ν – параметр самосогласованного барьера для прямо-двойственного допустимого множества $Q_p \times Q_d$ (см. параграф 1.3).
- $\mathbf{S_p} \bigcap \mathbf{L_d}$. Эта комбинация описывает негладкую оптимизацию в модели Черного Ящика. Прямо-двойственные субградиентные методы вычисляют ϵ-решение задачи (2.2.1) за

  $$O\left(\tfrac{1}{\epsilon^2} \cdot \|A\|^2 \cdot D_1 \cdot D_2 \right)$$

  градиентных шагов (см. Параграф 2.1).
- $\mathbf{B_p} \bigcap \mathbf{S_d}$. Сложность этой комбинации до сих пор неизвестна.

- $\mathbf{B_p} \bigcap \mathbf{L_d}$. Этот последний оставшийся вариант изучается в данном параграфе. С точки зрения черноящичной оптимизации он соответствует задаче минимизации негладкой выпуклой функции на множестве, для которого известен самосогласованный барьер.

Содержание

В п. 2.2.2 мы изучим операцию сглаживания опорной функции выпуклого множества с помощью самосогласованного барьера. Соответствующий барьерный субградиентный метод (БСМ) описывается в п. 2.2.3. Для этого метода мы доказываем оценку сложности $O(\frac{1}{\epsilon^2})$ итераций, где ϵ – требуемая точность решения задачи. В п. 2.2.4 показывается что этот метод может естественным образом применяться для максимизации вогнутой неотрицательной функции. В этом случае удается гарантировать получение приближенного решения соответствующей задачи с *относительной точностью* δ за $O(\frac{1}{\delta^2})$ итераций. В п. 2.2.5 мы рассматриваем несколько приложений. Наконец, п. 2.2.6 обсуждается использование БСМ для решения задач оптимизации в *реальном времени*. Мы показываем, что этот подход может служить серьезной альтернативой для постановок стохастической оптимизации. В приложении мы обсуждаем сложность нахождения барьерной проекции на стандартный симплекс.

Обозначения

Для линейного оператора $A : E \to H^*$ через $A^* : H \to E^*$ обозначается сопряженный оператор:

$$\langle Ax, y \rangle \;=\; \langle A^* y, x \rangle, \quad x \in E,\; y \in H.$$

Для *вогнутой* функции $f(x)$ обозначение $\nabla f(x)$ используется для одного из ее суперградиентов в x:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle, \quad y, x \in \mathrm{dom}\, f.$$

Для функции от двух векторных переменных $\Psi(u,x)$ обозначение $\nabla_2 \Psi(u,x)$ относится к субградиенту по второму аргументу.

2.2.2 Сглаживание с помощью самосогласованного барьера

Пусть выпуклое замкнутое множество $Q \subset E$ не содержит прямых. Предположим, что Q снабжено ν-самосогласованным барьером F. Рассмотрим другое выпуклое множество $\hat{P} \subseteq E$. Нас будет интересовать их пересечение

$$P \;=\; \hat{P} \bigcap Q,$$

которое предполагается ограниченным. Обозначим через x_0 его *условный аналитический центр*:

$$x_0 \;=\; \arg\min_{x \in P_0} F(x) \;\in\; P_0 \overset{\mathrm{def}}{=} \hat{P} \bigcap \mathrm{int}\, Q \;\subseteq\; P. \tag{2.2.2}$$

Таким образом, $F(x) \geq F(x_0)$ при всех $x \in P$. Поскольку Q не содержит прямых, центр x_0 существует (см. теорему 1.3.3)).

Для множества P введем гладкую аппроксимацию его опорной функции:

$$U_\beta(s) \;=\; \max_{x \in \hat{P}} \{ \langle s, x - x_0 \rangle - \beta [F(x) - F(x_0)] \}, \quad s \in E^*, \tag{2.2.3}$$

где $\beta > 0$ – параметр сглаживания. Обозначим через $u^\star_\beta(s)$ единственное решение задачи максимизации (2.2.3). Тогда

$$\nabla U_\beta(s) \;=\; u^\star_\beta(s) - x_0, \quad s \in E^*. \tag{2.2.4}$$

Для произвольного $x \in \operatorname{int} Q$ рассмотрим следующую локальную норму:

$$\|h\|_x \;=\; \langle \nabla^2 F(x) h, h \rangle^{1/2}, \qquad h \in E,$$

$$\|s\|^*_x \;=\; \langle s, [\nabla^2 F(x)]^{-1} s \rangle^{1/2}, \quad s \in E^*.$$

Тогда для функции $U_\beta(\cdot)$ можно гарантировать следующий уровень гладкости.

Лемма 2.2.1 *Пусть* $\beta > 0$, $s \in E^*$ *и* $x = u^\star_\beta(s)$. *Тогда для любого такого* $g \in E^*$, *что* $\|g\|^*_x < \beta$, *выполняется неравенство*

$$U_\beta(s + g) \;\leq\; U_\beta(s) + \langle g, \nabla U_\beta(s) \rangle + \beta \omega_*(\tfrac{1}{\beta} \|g\|^*_x), \tag{2.2.5}$$

где $\omega_*(\tau) = -\tau - \ln(1 - \tau) \leq \frac{\tau^2}{2(1-\tau)}$.

Доказательство.
В силу определения (2.2.3) для любого $y \in P_0$ имеем

$$\langle s - \beta \nabla F(x), y - x \rangle \;\leq\; 0. \tag{2.2.6}$$

Более того, поскольку F – самосогласованная функция, для любой точки $y \in \operatorname{int} Q$ выполняется неравенство

$$F(y) \;\geq\; F(x) + \langle \nabla F(x), y - x \rangle + \omega(\|y - x\|_x), \tag{2.2.7}$$

где $\omega(t) = t - \ln(1 + t)$ (см. неравенство (1.3.6)). Поэтому

$$\begin{aligned}
& U_\beta(s + g) - U_\beta(s) - \langle g, \nabla U_\beta(s) \rangle \\
& \overset{(2.2.4)}{=} \max_{y \in P_0} \{ \langle s + g, y - x_0 \rangle - \beta [F(y) - F(x_0)] - \langle s + g, x - x_0 \rangle + \beta [F(x) - F(x_0)] \} \\
& = \max_{y \in P_0} \{ \langle s + g, y - x \rangle - \beta [F(y) - F(x)] \} \\
& \overset{(2.2.6)}{\leq} \max_{y \in P_0} \{ \langle g, y - x \rangle + \beta [\langle \nabla F(x), y - x \rangle - F(y) + F(x)] \} \\
& \overset{(2.2.7)}{\leq} \max_{y \in P_0} \{ \langle g, y - x \rangle - \beta \omega(\|y - x\|_x) \} \;\leq\; \sup_{\tau \geq 0} \{ \; \tau \|g\|^*_x - \beta \omega(\tau) \; \}.
\end{aligned}$$

Если $\|g\|_x^* < \beta$, то супремум в правой части равен $\beta\omega_*(\frac{1}{\beta}\|g\|_x^*)$. □

Рассмотрим теперь аффинную функцию $l(x)$, $x \in P$. Для $\beta \geq 0$ обозначим

$$l^\star(\beta) \;=\; \max_{x \in P_0}\{l(x) - \beta[F(x) - F(x_0)]\} \;\geq\; l(x_0) \;\overset{\mathrm{def}}{=}\; l_0. \tag{2.2.8}$$

Тогда $l^\star(0) = \max\limits_{x \in P} l(x) \;\overset{\mathrm{def}}{=}\; l^\star$.

Лемма 2.2.2 *Для любого* $\beta > 0$ *выполняется неравенство*

$$l^\star(\beta) \;\leq\; l^\star \;\leq\; l^\star(\beta) + \beta\nu\left(1 + \left[\ln\tfrac{l^\star - l_0}{\beta\nu}\right]_+\right), \tag{2.2.9}$$

где $[a]_+ = \max\{a, 0\}$. *Более того,*

$$l^\star - l_0 \;\leq\; \left[\sqrt{l^\star(\beta) - l_0} + \sqrt{\beta\nu}\right]^2. \tag{2.2.10}$$

Доказательство.

Первое из неравенств (2.2.9) следует из определений (2.2.2) и (2.2.8). Докажем второе неравенство. Рассмотрим произвольную точку $y^\star \in \operatorname{Arg}\max\limits_{x \in P} l(x)$. Обозначим

$$y(\alpha) = x_0 + \alpha(y^\star - x_0), \quad \alpha \in [0, 1].$$

В силу неравенства (1.3.29) имеем

$$F(y(\alpha)) \;\leq\; F(x_0) - \nu\ln(1 - \alpha), \quad \alpha \in [0, 1).$$

Поскольку функция $l(\cdot)$ линейна, это неравенство означает, что

$$\begin{aligned} l^\star(\beta) \;&\geq\; \max_{\alpha \in [0,1)}\{l(y(\alpha)) - \beta[F(y(\alpha)) - F(x_0)]\} \\ &\geq\; (1 - \alpha)l_0 + \alpha l^\star + \beta\nu\ln(1 - \alpha), \quad \alpha \in [0, 1). \end{aligned} \tag{2.2.11}$$

Максимум последнего выражения по α достигается в точке $\alpha^\star = \left[1 - \frac{\beta\nu}{l^\star - l_0}\right]_+$. Таким образом, если $\frac{l^\star - l_0}{\beta\nu} \leq 1$ (т. е. $\alpha^\star = 0$), то $l^\star \leq l_0 + \beta\nu$ и неравенство (2.2.9) следует из соотношения (2.2.8). Если $\alpha^\star > 0$, то мы получаем неравенство (2.2.9) с помощью простой подстановки.

С другой стороны, из оценки (2.2.11) получаем

$$l^\star - l_0 \;\leq\; \tfrac{1}{\alpha}\left[l^\star(\beta) - l_0 + \beta\nu\ln\left(1 + \tfrac{\alpha}{1-\alpha}\right)\right] \;\leq\; \tfrac{1}{\alpha}[l^\star(\beta) - l_0] + \tfrac{\beta\nu}{1-\alpha}.$$

Минимизируя это выражение по α, приходим к неравенству (2.2.10). □

Следствие 2.2.1 *Для любых* $\beta > 0$ *выполнено неравенство*

$$l^\star \;\leq\; l^\star(\beta) + \beta\nu\left[1 + 2\ln\left(1 + \sqrt{\tfrac{l^\star(\beta) - l_0}{\beta\nu}}\right)\right]. \tag{2.2.12}$$

2.2.3 Барьерный субградиентный метод

Рассмотрим задачу выпуклой оптимизации в следующей форме:

$$\text{найти } f_\star \overset{\text{def}}{=} \max_x \{ f(x) : \ x \in P \}, \tag{2.2.13}$$

где f – вогнутая функция и множество P удовлетворяет структурным предположениям, сформулированным в начале п. 2.2.2. Мы предполагаем, что функция f супердифференцируема на P_0 и что множество P является простым. Последнее означает, что вспомогательная задача оптимизации (2.2.3) может быть легко решена.

Приведем общую схему *барьерного субградиентного метода* (БСМ).

Инициализация Полагаем $s_0 = 0 \in E^*$.

Итерация $(k \geq 0)$ (2.2.14)

1. Выбираем $\beta_k > 0$ и вычисляем $x_k = u^\star_{\beta_k}(s_k)$.

2. Выбираем $\lambda_k > 0$ и полагаем $s_{k+1} = s_k + \lambda_k \nabla f(x_k)$.

Напомним, что обозначение $u^\star_\beta(s)$ используется для единственного решения задачи оптимизации (2.2.3). Таким образом, БСМ является *аффинно-инвариантным* методом.

Для анализа скорости сходимости метода (2.2.14) нам потребуются следующие *функции зазора*:

$$\begin{aligned} l_k(y) &= \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), y - x_i \rangle, \\ l^\star_k &\overset{\text{def}}{=} \max_{y \in P} l_k(y), \quad k \geq 0. \end{aligned}$$

Теорема 2.2.1 *Пусть параметры метода (2.2.14) удовлетворяют соотношению*

$$\lambda_k \| \nabla f(x_k) \|^*_{x_k} \ \leq \ \beta_k \ \leq \ \beta_{k+1}, \quad k \geq 0. \tag{2.2.15}$$

Обозначим $S_k = \sum\limits_{i=0}^{k} \lambda_i$, *и* $A_k = \sum\limits_{i=0}^{k} \beta_i \omega_* \left(\frac{\lambda_i}{\beta_i} \| \nabla f(x_i) \|^*_{x_i} \right)$. *Тогда для любого* $k \geq 0$ *выполняется неравенство:*

$$l^\star_k \ \leq \ A_k + \beta_{k+1} \nu \left[1 + 2 \ln \left(1 + \sqrt{\frac{A_k}{\beta_{k+1}\nu} + c(Q) \frac{S_k}{\beta_{k+1}} \| \nabla f(x_0) \|^*_{x_0}} \right) \right], \tag{2.2.16}$$

где $c(Q) = 1$, *если* Q *является конусом и барьер* $F(\cdot)$ *логарифмически однороден. В противном случае* $c(Q) = 3$.

Доказательство.

Для любого $k \geq 0$ имеем

$$U_{\beta_{k+1}}(s_{k+1}) \overset{(2.2.15)}{\leq} U_{\beta_k}(s_{k+1})$$

$$\overset{(2.2.5)}{\leq} U_{\beta_k}(s_k) + \lambda_k \langle \nabla f(x_k), u^\star_{\beta_k}(s_k) - x_0 \rangle + \beta_k \omega_* \left(\frac{\lambda_k}{\beta_k} \| \nabla f(x_k) \|^*_{x_k} \right).$$

Поскольку $U_{\beta_0}(0) = 0$, заключаем, что

$$\begin{aligned} \langle s_{k+1}, x_{k+1} - x_0 \rangle - \beta_{k+1}[F(x_{k+1}) - F(x_0)] &= U_{\beta_{k+1}}(s_{k+1}) \\ \leq \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), x_i - x_0 \rangle + \sum_{i=0}^{k} \beta_i \omega_* \left(\frac{\lambda_i}{\beta_i} \| \nabla f(x_i) \|^*_{x_i} \right). \end{aligned} \tag{2.2.17}$$

В силу условия оптимальности для (2.2.3) для всех $y \in P_0$ имеем

$$\langle s_{k+1}, y - x_{k+1} \rangle \leq \beta_{k+1} \langle \nabla F(x_{k+1}), y - x_{k+1} \rangle. \tag{2.2.18}$$

Заметим, что $s_{k+1} = \sum\limits_{i=0}^{k} \lambda_i \nabla f(x_i)$. Таким образом, для любого $y \in P_0$ получаем

$$\begin{aligned} \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), y - x_i \rangle &\overset{(2.2.17)}{\leq} \langle s_{k+1}, y - x_{k+1} \rangle + \beta_{k+1}[F(x_{k+1}) - F(x_0)] + A_k \\ &\overset{(2.2.18)}{\leq} \beta_{k+1}[F(x_{k+1}) + \langle \nabla F(x_{k+1}), y - x_{k+1} \rangle - F(x_0)] + A_k \\ &\leq \beta_{k+1}[F(y) - F(x_0)] + A_k. \end{aligned}$$

Поэтому $l^\star_k(\beta_{k+1}) \leq A_k$. С другой стороны, так как функция f выпукла, получаем

$$\begin{aligned} l_k(x_0) &= \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), x_0 - x_i \rangle \geq \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_0), x_0 - x_i \rangle \\ &\geq -\| \nabla f(x_0) \|^*_{x_0} \cdot \sum_{i=0}^{k} \lambda_i \| x_0 - x_i \|_{x_0}. \end{aligned}$$

В силу определения (2.2.2) имеем $\langle \nabla F(x_0), x_i - x_0 \rangle \geq 0$. Поэтому ввиду теоремы 1.3.21 получаем $\| x_i - x_0 \|_{x_0} \leq \nu + 2\sqrt{\nu} \leq 3\nu$ (напомним, что $\nu \geq 1$ в силу леммы 1.3.7). Если Q – конус, то $\| x_i - x_0 \|_{x_0} \leq \nu$ (см. лемму 1.3.6). Таким образом, в любом случае

$$\| x_i - x_0 \|_{x_0} \leq \nu \cdot c(Q), \quad i = 0, \dots, k,$$

и мы заключаем, что $l_k(x_0) \geq -\nu \cdot c(Q) S_k \| \nabla f(x_0) \|^*_{x_0}$. Используя эти наблюдения в неравенстве (2.2.12), приходим к оценке (2.2.16). □

Оценим теперь скорость сходимости метода (2.2.14) при решении некоторых специальных задач.

Определение 2.2.1 Мы говорим, что $f \in \mathcal{B}_M(P)$, если $\|\nabla f(x)\|_x^* \le M$ для любого $x \in P_0$.

Для функции $f \in \mathcal{B}_M(P)$ в методе (2.2.14) можно выбрать следующие параметры:

$$\lambda_k = 1,\ k \ge 0, \qquad \beta_0 = \beta_1, \quad \beta_k = M \cdot \left(1 + \sqrt{\tfrac{k}{\nu}}\right),\ k \ge 1. \tag{2.2.19}$$

Теорема 2.2.2 *Пусть задача (2.2.13) с $f \in \mathcal{B}_M(P)$ решается методом (2.2.14) с параметрами, заданными правилами (2.2.19). Тогда при любом $k \ge 0$ выполняется неравество*

$$\tfrac{1}{S_k}\, l_k^\star \;\le\; 2M \cdot \left(\sqrt{\tfrac{\nu}{k+1}} + \tfrac{\nu}{k+1}\right) \cdot \left(1 + \ln\left(2 + \tfrac{1}{2} c(Q)\sqrt{\nu(k+1)}\right)\right). \tag{2.2.20}$$

Доказательство.

Зададим $\tau_k = \frac{1}{M}\beta_k > 1$. В силу правил выбора параметров (2.2.19) и предположений теоремы имеем $S_k = k+1$ и

$$\begin{aligned} A_k \;&=\; \sum_{i=0}^{k} \beta_i \omega_*\left(\tfrac{\lambda_i}{\beta_i}\|\nabla f(x_i)\|_{x_i}^*\right) \;\le\; M\sum_{i=0}^{k} \tau_i \omega_*\left(\tfrac{1}{\tau_i}\right) \;\le\; \tfrac{1}{2} M \sum_{i=0}^{k} \tau_i \tfrac{\tau_i^{-2}}{1-\tau_i^{-1}} \\ &=\; \tfrac{1}{2} M \sum_{i=0}^{k} \tfrac{1}{\tau_i - 1} \;=\; \tfrac{\sqrt{\nu}}{2} M \left[1 + \sum_{i=1}^{k} \tfrac{1}{\sqrt{i}}\right] \;\le\; \sqrt{\nu} M \left[\tfrac{1}{2} + \sqrt{k}\right]. \end{aligned} \tag{2.2.21}$$

Более того, по тем же причинам

$$\tfrac{S_k}{\beta_{k+1}}\|\nabla f(x_0)\|_{x_0}^* \;\le\; \tfrac{k+1}{1+\sqrt{\frac{k+1}{\nu}}} \;\le\; \sqrt{\nu(k+1)},$$

$$\tfrac{A_k}{\beta_{k+1}\nu} \;\le\; \tfrac{\frac{1}{2}+\sqrt{k}}{\sqrt{\nu}+\sqrt{k+1}} \;\le\; 1.$$

Таким образом, подставляя полученные оценки в неравенство (2.2.16), получаем

$$\begin{aligned} \tfrac{l_k^\star}{S_k} \;&\le\; M\left[\tfrac{\sqrt{\nu}}{k+1}\left(\tfrac{1}{2} + \sqrt{k}\right) + \tfrac{\nu + \sqrt{\nu(k+1)}}{k+1}\left(1 + 2\ln\left(1 + \sqrt{1 + c(Q)\sqrt{\nu(k+1)}}\right)\right)\right] \\ &\le\; 2M \cdot \left(\sqrt{\tfrac{\nu}{k+1}} + \tfrac{\nu}{k+1}\right) \cdot \left(1 + \ln\left(2 + \tfrac{1}{2}c(Q)\sqrt{\nu(k+1)}\right)\right). \end{aligned}$$

В последнем неравенстве мы воспользовались оценкой $\frac{\sqrt{\nu}}{k+1}\left(\frac{1}{2} + \sqrt{k}\right) \le \sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1}$. □

С параметрами, заданными правилами (2.2.19), метод (2.2.14) принимает такую форму:

$$x_{k+1} = \arg\max_{x \in P_0}\left\{\tfrac{1}{k+1}\sum_{i=0}^{k}\langle \nabla f(x_i), x - x_i\rangle - M\tfrac{\sqrt{\nu}+\sqrt{k+1}}{\sqrt{\nu(k+1)}}\left[F(x) - F(x_0)\right]\right\}. \tag{2.2.22}$$

Поскольку функция f вогнута, имеем

$$\begin{aligned} \tfrac{1}{S_k}\, l_k^\star \;&=\; \tfrac{1}{S_k}\max_{y \in P}\sum_{i=0}^{k}\lambda_i\langle \nabla f(x_i), y - x_i\rangle \\ &\ge\; \tfrac{1}{S_k}\max_{y \in P}\sum_{i=0}^{k}\lambda_i[f(y) - f(x_i)] \;=\; f_\star - \tfrac{1}{S_k}\sum_{i=0}^{k}\lambda_i f(x_i). \end{aligned}$$

Таким образом, оценка (2.2.20) обосновывает следующую скорость сходимости по *прямым* переменным:

$$f_\star - \tfrac{1}{S_k}\sum_{i=0}^{k}\lambda_i f(x_i) \;\leq\; 2M\cdot\left(\sqrt{\tfrac{\nu}{k+1}}+\tfrac{\nu}{k+1}\right)\cdot\left(1+\ln\left(2+\tfrac{1}{2}c(Q)\sqrt{\nu(k+1)}\right)\right). \qquad (2.2.23)$$

Заметим, что значение $l^\star_k$ легко вычислить. Поэтому оно может использоваться в критерии остановки метода.

В завершение этого пункта покажем, что метод (2.2.22) может генерировать также и приближенные решения для двойственной задачи. Для этого нужно задействовать внутреннюю структуру нашей задачи. Предположим что она задана в *седловой форме*:

$$f(x) \;=\; \min_{w\in S}\;\Psi(w,x) \quad\to\quad \max_{x\in P}, \qquad (2.2.24)$$

где $S\subset E_1$ – выпуклое замкнутое множество и функция $\Psi(w,x)$ выпукла по $w\in S$ и вогнута и супердифференцируема по $x\in P$. Тогда двойственная задача запишется так:

$$\text{найти } f_\star \;=\; \min_{w\in S}\;\eta(w), \quad \eta(w)\;\stackrel{\mathrm{def}}{=}\;\max_{y\in P}\Psi(w,y). \qquad (2.2.25)$$

Поскольку множество P ограничено, решение этой задачи существует. Не ограничивая общности, всегда можно выбрать

$$\nabla f(x) \;=\; \nabla_2\Psi(w(x),x)$$

с некоторым $w(x)\in \mathrm{Arg}\min\limits_{w\in S}\Psi(w,x)\subseteq S$.

Предположим, что точка $w(x)$ вычислима для любого $x\in P$. Для полноты картины, приведем следующий результат с полным доказательством.

Лемма 2.2.3 *Положим* $\bar w_k = \frac{1}{S_k}\sum\limits_{i=0}^{k}\lambda_i w(x_i)$ *и* $\bar x_k = \frac{1}{S_k}\sum\limits_{i=0}^{k}\lambda_i x_i$. *Тогда*

$$\eta(\bar w_k) - f(\bar x_k) \;\leq\; \tfrac{1}{S_k}\,l^\star_k. \qquad (2.2.26)$$

Доказательство.

Поскольку функция Ψ вогнута по второму аргументу, для любого $y\in P$ имеем

$$\begin{aligned}\langle\nabla f(x_i),y-x_i\rangle \;&=\; \langle\nabla_2\Psi(w(x_i),x_i),y-x_i\rangle\\ &\geq\; \Psi(w(x_i),y)-\Psi(w(x_i),x_i)=\Psi(w(x_i),y)-f(x_i).\end{aligned}$$

Поэтому

$$\begin{aligned}\tfrac{1}{S_k}\,l^\star_k \;&=\; \tfrac{1}{S_k}\max_{y\in P}\sum_{i=0}^{k}\lambda_i\langle\nabla f(x_i),y-x_i\rangle \;\geq\; \tfrac{1}{S_k}\max_{y\in P}\sum_{i=0}^{k}\lambda_i[\Psi(w(x_i),y)-f(x_i)]\\ &\geq\; \max_{y\in P}\Psi(\bar w_k,y)-\tfrac{1}{S_k}\sum_{i=0}^{k}\lambda_i f(x_i) \;=\; \eta(\bar w_k)-\tfrac{1}{S_k}\sum_{i=0}^{k}\lambda_i f(x_i)\\ &\geq\; \eta(\bar w_k)-f(\bar x_k). \qquad \Box\end{aligned}$$

Таким образом, метод (2.2.22) действительно может генерировать приближенные двойственные решения:

$$\eta(\bar{w}_k) - f(\bar{x}_k) \;\leq\; 2M \cdot \left(\sqrt{\tfrac{\nu}{k+1}} + \tfrac{\nu}{k+1} \right) \cdot \left(1 + \ln \left(2 + \tfrac{1}{2} c(Q) \sqrt{\nu(k+1)} \right) \right). \qquad (2.2.27)$$

2.2.4 Максимизация положительной вогнутой функции

Рассмотрим теперь выпуклую задачу оптимизации

$$\text{Find } \psi_\star \overset{\text{def}}{=} \max_x \{ \psi(x) : \; x \in P \}, \qquad (2.2.28)$$

где множество $P = \hat{P} \bigcap Q$ удовлетворяет предположениям задачи (2.2.13). Однако теперь мы предполагаем, что функция ψ вогнута и *положительна* на $\operatorname{int} Q$:

$$\psi(x) \;>\; 0 \quad \forall x \in \operatorname{int} Q. \qquad (2.2.29)$$

Лемма 2.2.4 *Пусть функция ψ вогнута и положительна на $\operatorname{int} Q$. Тогда при любом $x \in \operatorname{int} Q$ выполняется неравенство*

$$\| \nabla \psi(x) \|^*_x \;\leq\; \psi(x). \qquad (2.2.30)$$

Доказательство.
Выберем произвольную точку $x \in \operatorname{int} Q$ и $r \in [0,1)$. Положим

$$y \;=\; x - \tfrac{r}{\| \nabla \psi(x) \|^*_x} [\nabla^2 F(x)]^{-1} \nabla \psi(x).$$

В силу теоремы 1.3.5 $y \in \operatorname{int} Q$. Таким образом,

$$0 \leq \psi(y) \leq \psi(x) + \langle \nabla \psi(x), y - x \rangle = \psi(x) - r \| \nabla \psi(x) \|^*_x.$$

Поскольку число $r \in [0,1)$ произвольно, получаем оценку (2.2.30). □

Этот результат имеет важные следствия. Возьмем от целевой функции задачи (2.2.28) натуральный логарифм:

$$f(x) \;\overset{\text{def}}{=}\; \ln \psi(x). \qquad (2.2.31)$$

Лемма 2.2.5 *Пусть функция ψ вогнута и положительна в смысле (2.2.29). Тогда функция $f \in \mathcal{B}_1(Q)$ является вогнутой на Q.*

Доказательство.
Действительно, хорошо известно что логарифм вогнутой функции тоже является вогнутым. Осталось заметить, что $\nabla f(x) = \frac{1}{\psi(x)} \nabla \psi(x)$, и применить неравенство (2.2.30). □

Таким образом, для решения задачи (2.2.28) мы можем применить метод (2.2.14) к задаче (2.2.13) с целевой функцией, заданной формулой (2.2.31). Получившаяся схема выглядит следующим образом:

$$x_{k+1} = \arg\max_{x \in P_0} \left\{ \frac{1}{k+1} \sum_{i=0}^{k} \langle \frac{\nabla \psi(x_i)}{\psi(x_i)}, x - x_i \rangle - \frac{\sqrt{\nu}+\sqrt{k+1}}{\sqrt{\nu(k+1)}} \left[F(x) - F(x_0) \right] \right\}. \tag{2.2.32}$$

Для схемы (2.2.32) можно гарантировать определенную скорость сходимости с *относительной точностью.*

Теорема 2.2.3 *Пусть последовательность* $\{x_k\}_{k=0}^{\infty}$ *генерируется методом (2.2.32) для задачи (2.2.28). Тогда для любого* $k \geq 0$ *выполняются неравенства*

$$\begin{array}{rcl} \left[\prod\limits_{i=0}^{k} \psi(x_i) \right]^{\frac{1}{k+1}} & \geq & \psi_\star \cdot \exp\left\{ -2 \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \left(1 + \ln \left(2 + \frac{c(Q)}{2} \sqrt{\nu(k+1)} \right) \right) \right\} \\ & \geq & \psi_\star \cdot \left[1 - 2 \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \left(1 + \ln \left(2 + \frac{c(Q)}{2} \sqrt{\nu(k+1)} \right) \right) \right]. \end{array} \tag{2.2.33}$$

Доказательство.

Действительно, мы просто применяем метод (2.2.22) к функции f, заданной в (2.2.31). Поскольку $f \in \mathcal{B}_1(Q) \subseteq \mathcal{B}_1(P)$, с помощью неравенства (2.2.20) получаем

$$f_\star - \frac{1}{k+1} \sum_{i=0}^{k} f(x_i) \ \leq \ \delta_k \overset{\mathrm{def}}{=} 2 \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \left(1 + \ln \left(2 + \frac{c(Q)}{2} \sqrt{\nu(k+1)} \right) \right).$$

Поэтому $\left[\prod\limits_{i=0}^{k} \psi(x_i) \right]^{\frac{1}{k+1}} \geq \psi_\star \cdot e^{-\delta_k} \ \geq \ \psi_\star \cdot (1 - \delta_k)$. Это в точности оценка (2.2.33). □

Покажем, как можно решать задачу, двойственную к (2.2.28). Для простоты предположим что

$$\psi(x) \ = \ \min_{u \in \Omega} \Psi_0(u, x), \tag{2.2.34}$$

где $\Omega \subset E_1$ – выпуклое замкнутое множество. В этом случае условие (2.2.29) можно записать как

$$\Psi_0(u, x) \ \geq 0, \quad u \in \Omega, \ x \in P. \tag{2.2.35}$$

Заметим, что

$$\begin{array}{rcl} \max\limits_{x \in P} \ln \psi(x) & = & \max\limits_{x \in P} \ \min\limits_{\tau > 0} \ \min\limits_{u \in \Omega} \ \left[\tau \Psi_0(u, x) - \ln \tau - 1 \right] \\ & = & \max\limits_{x \in P} \ \min\limits_{\substack{v \in \tau\Omega, \\ \tau > 0}} \ \left[\tau \Psi_0 \left(\frac{1}{\tau} v, x \right) - \ln \tau - 1 \right] \\ & = & \min\limits_{\substack{v \in \tau\Omega, \\ \tau > 0}} \ \left\{ \eta(w) \equiv \eta(v, \tau) \overset{\mathrm{def}}{=} -1 - \ln \tau + \tau \psi^\star \left(\frac{1}{\tau} v \right) \right\}, \end{array}$$

где $\psi^\star(u) = \max\limits_{x \in P} \Psi_0(u,x)$.

Обозначим через $u(x)$ решение задачи минимизации (2.2.34). Тогда точка $w(x)$ определяется следующим образом:

$$w(x) = (v(x), \tau(x)), \quad v(x) = \tau(x) u(x), \quad \tau(x) = \tfrac{1}{\psi(x)}.$$

В соответствии с леммой 2.2.3 мы можем сформировать $\bar{w}_k = (\bar{v}_k, \bar{\tau}_k)$, где

$$\bar{v}_k = \tfrac{1}{k+1} \sum_{i=0}^{k} \tfrac{u(x_i)}{\psi(x_i)}, \quad \bar{\tau}_k = \tfrac{1}{k+1} \sum_{i=0}^{k} \tfrac{1}{\psi(x_i)}.$$

Обозначим $\bar{x}_k = \frac{1}{k+1} \sum\limits_{i=0}^{k} x_i$, и $\bar{u}_k = \frac{\bar{v}_k}{\bar{\tau}_k} = \sum\limits_{i=0}^{k} \frac{u(x_i)}{\psi(x_i)} / \left[\sum\limits_{i=0}^{k} \frac{1}{\psi(x_i)} \right] \in \Omega$. Тогда в силу неравенства (2.2.26) получаем

$$\begin{aligned} \tfrac{1}{S_k}\, l_k^\star \;&\geq\; \eta(\bar{w}_k) - \ln \psi(\bar{x}_k) \;=\; -1 - \ln \bar{\tau}_k + \bar{\tau}_k \psi^\star \left(\tfrac{1}{\bar{\tau}_k} \bar{v}_k \right) - \ln \psi(\bar{x}_k) \\ &=\; -1 - \ln \bar{\tau}_k + \bar{\tau}_k \psi^\star \left(\bar{u}_k \right) - \ln \psi(\bar{x}_k) \;\geq\; \ln \tfrac{\psi^\star(\bar{u}_k)}{\psi(\bar{x}_k)}. \end{aligned}$$

Поэтому

$$\psi(\bar{x}_k) \;\geq\; \psi^\star(\bar{u}_k) \cdot \exp\left\{ -\tfrac{1}{S_k}\, l_k^\star \right\}. \tag{2.2.36}$$

2.2.5 Приложения

В этом параграфе мы рассмотрим несколько примеров приложений метода (2.2.32). Будем говорить, что значение $\bar{\phi}$ является δ-приближением оптимального значения $\phi_\star > 0$ с *относительной точностью*, если

$$\phi_\star \geq \bar{\phi} \geq \phi_\star \cdot e^{-\delta}, \quad \delta > 0.$$

В оценках трудоемкости компактное обозначение $O^*(\cdot)$ используется для оценок, в которых опущены несущественные логарифмические множители. Поскольку скорость сходимости (2.2.33) не зависит от данных конкретной задачи, мы можем сказать, что соответствующий метод является *чисто полиномиальной аппроксимирующей схемой.*

Вычисление дробных покрытий

Рассмотрим следующую задачу по вычислению *дробных покрытий*:

$$\text{найти } \phi_\star \overset{\text{def}}{=} \min_y \{ \langle b, y \rangle : \; A^T y \geq c, \; y \geq 0 \in \mathbb{R}^m \}, \tag{2.2.37}$$

где $A = (a_1, \ldots, a_n)$ – $(m \times n)$-матрица с неотрицательными коэффициентами, а векторы $b \in \mathbb{R}^m$ и $c \in \mathbb{R}^n$ имеют положительные коэффициенты. Обозначим

$$\psi(y) \;=\; \min_{1 \leq i \leq n} \tfrac{1}{c(i)} \langle a_i, y \rangle.$$

Функция ψ является вогнутой и однородной степени единица. Таким образом,

$$\begin{aligned}\phi_\star &= \min_y \left\{ \tfrac{\langle b,y\rangle}{\psi(y)} : \; y \geq 0 \in \mathbb{R}^m \right\} \\ &= \left[\max_y \left\{ \tfrac{\psi(y)}{\langle b,y\rangle} : \; y \geq 0 \in \mathbb{R}^m \right\} \right]^{-1} \\ &= \left[\max_y \left\{ \psi(y) : \; \langle b,y\rangle = 1, \; y \geq 0 \in \mathbb{R}^m \right\} \right]^{-1}.\end{aligned}$$

Следовательно, задача (2.2.37) может быть записана в форме (2.2.28) с $K = \mathbb{R}^m_+$,

$$F(y) \; = \; -\sum_{j=1}^m \ln y^{(j)}, \quad \nu = m,$$

и $\hat{P} = \{y : \; \langle b,y\rangle = 1\}$. Поэтому в соответствии с оценкой (2.2.33), δ-аппроксимация значения $\phi_\star = \psi_\star^{-1}$ в относительной шкале может быть найдена за $O^*(\frac{m}{\delta^2})$ итераций метода (2.2.32). Каждая итерация метода требует $O(mn)$ операций для вычисления $\psi(y)$ и ее суперградиента и порядка $O(m \ln m)$ операций для решения вспомогательной задачи максимизации (2.2.32) (см. приложение). Конечно же, эта вычислительная стратегия хороша, только когда $m << n$. В противном случае лучше решать задачу двойственную к (2.2.37) с помощью техники сглаживания (см. гл. 5).

Максимальные многопродуктовые потоки

Рассмотрим сеть, состоящую из множеств вершин $\mathcal{N}$, $|\mathcal{N}| = n$, и направленных дуг

$$\mathcal{A} \; = \; \{\alpha = (i,j), \; i,j \in \mathcal{N}\}, \quad |\mathcal{A}| = m.$$

Предположим, что все дуги имеют ограниченные пропускные способности. Формально это означает что вектор потоков $f \in \mathbb{R}^m_+$ должен удовлетворять неравенство

$$f \leq \bar{f}.$$

Введем множество пар источник-сток $\mathcal{OD} \; = \; \{(i,j), \; i,j \in \mathcal{N}\}$. Каждая пара $(i,j) \in \mathcal{OD}$ порождает между вершинами i и j направленный поток $f_{i,j} \in \mathbb{R}^m_+$ мощностью $d_{i,j}$. Это означает, что векторы $f_{i,j}$ удовлетворяют систему линейных уравнений

$$Bf_{i,j} \; = \; d_{i,j}(e_i - e_j), \quad (i,j) \in \mathcal{OD},$$

где B – балансовая матрица сети и $e_{(\cdot)}$ - соответствующий базисный вектор в $\mathbb{R}^n$.

Задача нахождения *максимального многопродуктового потока* записывается следующим образом:

$$\begin{aligned}\text{найти } \lambda_\star \overset{\text{def}}{=} \max_{\lambda, f_{i,j}} \{\lambda : \; & Bf_{i,j} \; = \; \lambda \cdot d_{i,j}(e_i - e_j), \; f_{i,j} \; \geq \; 0, \; (i,j) \in \mathcal{OD}, \\ & \sum_{(i,j)\in\mathcal{OD}} f_{i,j} \leq \bar{f} \; \}.\end{aligned} \tag{2.2.38}$$

Хорошо известно [116], что эта задача может быть выписана в двойственной форме:

$$\begin{array}{rcl} \psi_\star \overset{\text{def}}{=} \lambda_\star^{-1} & = & \max\limits_t \{\psi(t) : \langle \bar{f}, t\rangle = 1,\ t \geq 0 \in \mathbb{R}^m\}, \\ \psi(t) & = & \sum\limits_{(i,j)\in\mathcal{OD}} d_{i,j} \cdot SP_{i,j}(t), \end{array} \qquad (2.2.39)$$

где функции $SP_{i,j}(t)$ являются *кратчайшими расстояниями* между вершинами i и j относительно неотрицательных времен проезда по дугам, записанных в векторе $t \in \mathbb{R}^m$.

Ясно, что функция ψ из формулы (2.2.39) удовлетворяет всем предположениям задачи (2.2.28). Таким образом, задача (2.2.39) может быть решена методом (2.2.32). В соответствии с оценкой (2.2.33) δ-аппроксимация числа $\psi_\star$ в относительной шкале может быть найдена за $O^*(\frac{m}{\delta^2})$ итераций. Каждая итерация требует вычисления кратчайших расстояний для всех пар источник–сток. Сложность решения вспомогательной задачи максимизации для метода (2.2.32) составляет $O(m \ln m)$ операций (см. приложение). Заметим, что мы также в состоянии вычислить приближенные двойственные решения (потоки между источником и стоком), используя прием описанный в конце п. 2.2.4.

Минимаксные задачи с неотрицательными компонентами

Рассмотрим следующую минимаксную задачу:

$$\text{найти } \psi_\star \overset{\text{def}}{=} \min_x \left\{ \max_{1\leq i\leq m} f_i(x) :\ x \in S \right\}, \qquad (2.2.40)$$

где множество S выпукло и замкнуто и все функции $f_i(x)$ выпуклы и неотрицательны на S. Предположим, что функция

$$\psi(y) = \min_{x\in S} \left\{ \sum_{i=1}^m y^{(i)} f_i(x) \right\}$$

определена при всяком $y \geq 0 \in \mathbb{R}^m$. Более того, предположим, что значения этой функции и ее суперградиенты легко вычислимы.

Тогда задачу (2.2.40) можно переписать в двойственной форме

$$\psi_\star = \max_y \{\psi(y) :\ \langle e, y\rangle = 1,\ y \geq 0 \in \mathbb{R}^m\}, \qquad (2.2.41)$$

где $e \in \mathbb{R}^m$ – вектор из всех единиц.

Заметим, что задача (2.2.41) удовлетворяет всем предположениям задачи (2.2.28). Таким образом, в соответствии с оценкой (2.2.33) δ-аппроксимация значения $\psi_\star$ в относительной шкале находится методом (2.2.32) за $O^*(\frac{m}{\delta^2})$ итераций. Каждая итерация этой схемы требует решения задачи минимизации взвешенной суммы функций f_i и барьера F.

Полуопределенная релаксация булевского квадратичного программирования

Рассмотрим следующую задачу максимизации:

$$\text{найти } f_\star \overset{\text{def}}{=} \max_x \{\langle Ax, x\rangle : \ x^{(i)} = \pm 1, \ i = 1, \dots, n\}, \tag{2.2.42}$$

где $n \times n$-матрица A симметрична и положительно полуопределена. Ясно, что это НП-сложная задача. Однако хорошо известно, что ее оптимальное значение может быть приближенно вычислено за полиномиальное время с некоторой относительной точностью, не зависящей от размерности пространства (см. [81]). А именно, положим

$$\psi_\star \ = \ \min_y \{\langle e, y\rangle : \ D(y) \succeq A\}, \tag{2.2.43}$$

где $D(y)$ – диагональная $n \times n$-матрица с вектором y на диагонали. Тогда

$$\tfrac{2}{\pi}\psi_\star \leq f_\star \leq \psi_\star.$$

Обычно задача (2.2.43) решается с помощью методов внутренней точки. Однако очень часто бессмысленно вычислять приближение $\psi_\star$ с высокой относительной точностью. Поэтому давайте попробуем применить к ее решению дешевый градиентный метод.

Получим сначала удобное представление для $\psi_\star$.

Лемма 2.2.6 *Пусть $A = L^T L$. Тогда*

$$\psi_\star \ = \ \max_X \left\{ \psi(X) \overset{\text{def}}{=} \left[\sum_{i=1}^n \langle Xq_i, q_i\rangle^{1/2} \right]^2 : \ \langle I, X\rangle = 1, \ X \succeq 0 \right\}, \tag{2.2.44}$$

где q_i – столбцы матрицы L, I – единичная матрица и скалярное произведение в пространстве симметрических матриц определено естественным образом.

Доказательство.

Действительно, поскольку $A \succ 0$, получаем, что

$$\begin{aligned} \psi_\star \ &= \ \min_u \left\{ \sum_{i=1}^n \tfrac{1}{u^{(i)}} : \ A^{-1} \succeq D(u) \right\} \\ &= \ \min_u \ \max_{Y \succeq 0} \ \left\{ \sum_{i=1}^n \tfrac{1}{u^{(i)}} + \langle Y, D(u) - A^{-1}\rangle \right\} \\ &= \ \max_{Y \succeq 0} \ \min_u \ \left\{ \sum_{i=1}^n \left(\tfrac{1}{u^{(i)}} + Y^{(i,i)} \, u^{(i)} \right) - \langle Y, A^{-1}\rangle \right\} \end{aligned}$$

Таким образом, $\psi_\star = \max\limits_{Y \succeq 0} \left\{ 2 \sum\limits_{i=1}^n \left[Y^{(i,i)} \right]^{1/2} - \langle Y, A^{-1}\rangle \right\}$. Максимизируя целевую функцию этой задачи вдоль фиксированного направления $Y \succeq 0$, получаем

$$\psi_\star \ = \max_{Y \succeq 0} \left\{ \tfrac{1}{\langle Y, A^{-1}\rangle} \left[\sum_{i=1}^n \left[Y^{(i,i)} \right]^{1/2} \right]^2 \right\}.$$

Выбирая в этой задаче новую переменную $X = L^{-T} Y L^{-1}$, приходим к представлению (2.2.44). □

Заметим, что функция ψ в формуле (2.2.44) вогнута. Более того, она дифференцируема и положительна при любом $X \succ 0$. В нашем случае Q – конус неотрицательно определенных матриц с барьером

$$F(X) \; = \; - \ln \det X, \quad \nu = n.$$

Поэтому задача (2.2.44) удовлетворяет условиям задачи (2.2.28). Следовательно, значение $\psi_\star$ может быть приближено с помощью метода (2.2.32) за $O^*(\frac{n}{\delta^2})$ итераций, где δ – требуемая относительная точность. В нашем случае каждая итерация метода (2.2.32) требует представления $(n \times n)$-матрицы в форме UTU^T, где U – ортогональная, а T – тридиагональная матрица. После этого можно применить эффективную процедуру поиска описанную в приложении.

2.2.6 Оптимизация в реальном времени как альтернатива стохастическому программированию

Принятие решений в условиях неопределенности

Рассмотрим повторяющийся процесс принятия решения с неопределенностью в размере дохода. Предположим, что у нас есть $N + 1$ периодов времени, в каждом из которых осуществляется полный производственный цикл (т. е. в начале периода происходит вложение капитала, а в конце все средства производства и товар опять превращаются в деньги). В начале k-го периода мы выбираем производственную стратегию

$$x_k \in P, \quad k = 0, \dots, N,$$

где структура множества P удовлетворяет условиям п. 2.2.2. Результаты различной экономической деятельности за этот период описываются *производственной функцией* $\psi_k(x) \geq 0$, $x \in P$. Значение $\psi_k(x)$ равно приросту капитала, инвестированного в начале периода k в соответствии с производственной стратегией $x \in P$. Однако функция $\psi_k(\cdot)$ становится известной только к концу k-го периода. Таким образом, ее можно использовать только при выборе стратегии для следующих периодов.

Предположим сначала, что мы заранее знаем все производственные функции

$$\psi_k(x), \quad k = 0, \dots, N.$$

Однако в силу некоторых обстоятельств мы вынуждены применять во всех периодах фиксированную стратегию $x \in P$. В этом случае, конечно же, ее стоит выбрать как

$$x_N^\star \; \stackrel{\mathrm{def}}{=} \; \arg\max_{x \in P} \prod_{k=0}^{N} \psi_k(x).$$

Тогда средняя эффективность нашего выбора будет составлять

$$\psi_N^\star \;=\; \left[\prod_{k=0}^{N} \psi_k(x^\star)\right]^{\frac{1}{N+1}}.$$

Однако обычно будущее нам неизвестно. В то же время, часто мы можем применять для каждого k-го периода свою производственную стратегию $x_k \in P$. Давайте посмотрим на ее возможную эффективность.

Предположим, что мы знаем ν-самосогласованный барьер $F(x)$ для множества Q. Тогда мы можем применить следующий вариант метода (2.2.32):

$$x_{k+1} = \arg\max_{x \in P_0} \left\{ \tfrac{1}{k+1} \sum_{i=0}^{k} \langle \tfrac{\nabla \psi_i(x_i)}{\psi_i(x_i)}, x - x_i \rangle - \tfrac{\sqrt{\nu}+\sqrt{k+1}}{\sqrt{\nu}(k+1)} \left[F(x) - F(x_0)\right] \right\}. \tag{2.2.45}$$

В этом случае после $N+1$ периодов средняя скорость роста есть

$$\Psi_N \;\overset{\text{def}}{=}\; \left[\prod_{k=0}^{N} \psi_k(x_k)\right]^{\frac{1}{N+1}}.$$

Теорема 2.2.4 *При любом* $N \geq 0$ *выполняется неравенство* $\Psi_N \geq \psi_N^\star \cdot e^{-\delta_N}$*, где*

$$\delta_N \;=\; 2\left(\sqrt{\tfrac{\nu}{N+1}} + \tfrac{\nu}{N+1}\right) \cdot \left(1 + \ln\left(2 + \tfrac{c(Q)}{2}\sqrt{\nu(N+1)}\right)\right) \;\to\; 0$$

при $N \to \infty$.

Доказательство.
Наши рассуждения очень похожи на доказательства теорем 2.2.1 и 2.2.2. Обозначим

$$f_k(x) = \ln \psi_k(x), \quad f(x) = \tfrac{1}{N+1} \sum_{k=0}^{N} f_k(x), \quad s_k = \sum_{i=0}^{k} \nabla f_i(x_i) = \sum_{i=0}^{k} \tfrac{\nabla \psi_i(x_i)}{\psi_i(x_i)}.$$

Заметим, что метод (2.2.45) может рассматриваться как применение схемы (2.2.14), (2.2.19) к изменяющейся целевой функции.

При любом $k \geq 0$ имеем

$$\begin{aligned} U_{\beta_{k+1}}(s_{k+1}) \;&\leq\; U_{\beta_k}(s_{k+1}) \\ &\overset{(2.2.5)}{\leq}\; U_{\beta_k}(s_k) + \langle \nabla f_k(x_k), u_{\beta_k}^\star(s_k) - x_0 \rangle + \beta_k \omega_* \left(\tfrac{1}{\beta_k} \|\nabla f_k(x_k)\|_{x_k}^*\right) \\ &\overset{(2.2.30)}{\leq}\; U_{\beta_k}(s_k) + \langle \nabla f_k(x_k), u_{\beta_k}^\star(s_k) - x_0 \rangle + \beta_k \omega_* \left(\tfrac{1}{\beta_k}\right). \end{aligned}$$

Поскольку $U_{\beta_0}(0) = 0$, мы заключаем, что

$$\begin{aligned} &\langle s_{N+1}, x_{N+1} - x_0 \rangle - \beta_{N+1}[F(x_{N+1}) - F(x_0)] \\ &= \; U_{\beta_{N+1}}(s_{N+1}) \;\leq\; \sum_{i=0}^{N} \langle \nabla f_i(x_i), x_i - x_0 \rangle + \sum_{i=0}^{N} \beta_i \omega_* \left(\tfrac{1}{\beta_i}\right) \\ &\overset{(2.2.21)}{\leq}\; \sum_{i=0}^{N} \langle \nabla f_i(x_i), x_i - x_0 \rangle + \sqrt{\nu}\left[\tfrac{1}{2} + \sqrt{N}\right]. \end{aligned} \tag{2.2.46}$$

В силу условий оптимальности для задачи (2.2.3) при всех $y \in P_0$ имеем

$$\langle s_{N+1}, y - x_{N+1} \rangle \;\le\; \beta_{N+1} \langle \nabla F(x_{N+1}), y - x_{N+1} \rangle. \tag{2.2.47}$$

Таким образом, пользуясь вогнутостью всех функций f_i, при любом $y \in P$ получаем

$$\begin{array}{rcl} l_N(y) & \stackrel{\mathrm{def}}{=} & \sum\limits_{i=0}^{N} \langle \nabla f_i(x_i), y - x_i \rangle \\[2ex] & \stackrel{(2.2.46)}{\le} & \langle s_{N+1}, y - x_{N+1} \rangle + \beta_{N+1}[F(x_{N+1}) - F(x_0)] + \sqrt{\nu}\left[\frac{1}{2} + \sqrt{N}\right] \\[2ex] & \stackrel{(2.2.47)}{\le} & \beta_{N+1}[F(x_{N+1}) + \langle \nabla F(x_{N+1}), y - x_{N+1} \rangle - F(x_0)] + \sqrt{\nu}\left[\frac{1}{2} + \sqrt{N}\right] \\[2ex] & \le & \beta_{N+1}[F(y) - F(x_0)] + \sqrt{\nu}\left[\frac{1}{2} + \sqrt{N}\right]. \end{array}$$

Поэтому $l^{\star}_N(\beta_{N+1}) \le \sqrt{\nu}\left[\frac{1}{2} + \sqrt{N}\right]$. С другой стороны, применяя те же рассуждения, что и в конце доказательства теоремы 2.2.1, мы получим

$$\begin{array}{rcl} l_N(x_0) & = & \sum\limits_{i=0}^{N} \langle \nabla f_i(x_i), x_0 - x_i \rangle \;\ge\; \sum\limits_{i=0}^{N} \langle \nabla f_i(x_0), x_0 - x_i \rangle \\[2ex] & \ge & -\nu \cdot c(Q) \cdot (N+1). \end{array}$$

Таким образом, $l^{\star}_N(\beta_{N+1}) - l_N(x_0) \le \sqrt{\nu}\left(\frac{1}{2} + \sqrt{N}\right) + \nu \cdot c(Q) \cdot (N+1)$. Поскольку $\beta_{N+1} = 1 + \sqrt{\frac{N+1}{\nu}}$, с помощью оценки (2.2.12) получаем

$$\begin{array}{rcl} \frac{l^{\star}_N}{N+1} & \le & \frac{\sqrt{\nu}}{N+1}\left(\frac{1}{2} + \sqrt{N}\right) \\[2ex] & & + \frac{\nu + \sqrt{\nu(N+1)}}{N+1}\left[1 + 2\ln\left(1 + \sqrt{\frac{\sqrt{\nu}\left(\frac{1}{2} + \sqrt{N}\right) + \nu \cdot c(Q) \cdot (N+1)}{\nu + \sqrt{\nu(N+1)}}}\right)\right] \\[2ex] & \le & \frac{\sqrt{\nu}}{N+1}\left(\frac{1}{2} + \sqrt{N}\right) \\[2ex] & & + \frac{\nu + \sqrt{\nu(N+1)}}{N+1}\left[1 + 2\ln\left(1 + \sqrt{1 + c(Q)\sqrt{\nu(N+1)}}\right)\right] \;\le\; \delta_N \end{array}$$

(см. рассуждения в конце доказательства теоремы 2.2.2). С другой стороны,

$$\begin{array}{rcl} \frac{1}{N+1}\, l^{\star}_N & = & \frac{1}{N+1} \max\limits_{y \in P} \left\{ \sum\limits_{i=0}^{N} \langle \nabla f_i(x_i), y - x_i \rangle \right\} \\[2ex] & \ge & \frac{1}{N+1} \max\limits_{y \in P} \left\{ \sum\limits_{i=0}^{N} [f_i(y) - f_i(x_i)] \right\} \;=\; \ln \psi^{\star}_N - \ln \Psi_N. \end{array}$$

□

Приведем теперь несколько приложений этой теоремы.

Управление портфелем ценных бумаг

Пусть вектор $x \in \Delta_n$ описывает структуру нашего портфеля. Обозначим через $c_k^{(i)} \geq 0$, $i = 1, \dots, n$, коэффициенты роста цены акции i за период $k \geq 0$. Тогда оптимальный портфель с *постоянными пропорциями* определяется следующим образом:

$$x_N^\star \;=\; \arg\max_{x \in P} \prod_{k=0}^{N} \langle c_k, x \rangle, \quad \psi_N^\star \;=\; \left[\prod_{k=0}^{N} \langle c_k, x_N^\star \rangle \right]^{1/(N+1)}.$$

Для множества $Q = \mathbb{R}^n_+$, запишем стандартный n-самосогласованный барьер

$$F(x) \;=\; -\sum_{i=1}^{n} \ln x^{(i)}.$$

Тогда мы можем воспользоваться следующим вариантом метода (2.2.45):

$$x_{k+1} = \arg\max_{x \in P_0} \left\{ \frac{1}{k+1} \sum_{i=0}^{k} \frac{\langle c_i, x - x_i \rangle}{\langle c_i, x_i \rangle} - \frac{\sqrt{\nu} + \sqrt{k+1}}{\sqrt{\nu}(k+1)} \left[F(x) - F(x_0) \right] \right\}, \quad k \geq 0. \tag{2.2.48}$$

В этом случае после $N+1$ периода мы получим следующую среднюю скорость роста стоимости нашего портфеля:

$$\Psi_N \;\stackrel{\mathrm{def}}{=}\; \left[\prod_{k=0}^{N} \langle c_k, x_k \rangle \right]^{\frac{1}{N+1}}.$$

В силу теоремы 2.2.4 имеем $\Psi_N \geq \psi_N^\star \cdot e^{-\delta_N(n)}$. Заметим, что каждый шаг метода (2.2.48) реализуем за $O(n \ln n)$ арифметических операций (см. Приложение).

Процессы с полным производственным циклом

Предположим, что в экономике имеется n эластичных производственных процессов. В начале k-го периода мы знаем цену $a_k^{(i)} > 0$ производства единицы i-го товара, $i = 1, \dots, n$. Эта цена вычисляется исходя из цен на сырье, труд, оборудование и т. д. Однако цена продажи $b_k^{(i)} \geq 0$ единицы товара i становится известной только в конце периода k, в тот момент, когда товар продается. Эта цена может зависеть от действий конкурентов, меняющихся предпочтений покупателей, и т. д. Обозначим через $x^{(i)}$ долю нашего капитала вкладываемую в i-й процесс. Тогда мы получим следующую производственную функцию:

$$\begin{aligned} \psi_k(x) \;&=\; \sum_{i=1}^{n} \frac{b_k^{(i)}}{a_k^{(i)}} \cdot x^{(i)}, \\ x \;&=\; (x^{(1)}, \dots, x^{(n)})^T \in Q \;\stackrel{\mathrm{def}}{=}\; \mathbb{R}^n_+, \\ \hat{P} \;&=\; \Delta_n. \end{aligned} \tag{2.2.49}$$

К этой модели можно применить метод (2.2.45) с барьером

$$F(x) \;=\; -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu = n.$$

Трудоемкость одной итерации этого метода, на которой приходится решать вспомогательную задачу (2.2.45), будет опять составлять $O(n \ln n)$ арифметических операций (см. приложение).

Обсуждение

Теорема 2.2.4, применяемая в условиях неопределенности, дает нам *абсолютную* и свободную от какого-либо *риска* гарантию для определенного уровня эффективности стратегии оптимизации (2.2.45), реализуемой в реальном времени. Для достижения такого результата нам не нужно вводить стандартные инструменты, связанные со случайными событиями, мерами риска, постановками стохастической или робастной минимизации. В теореме 2.2.4 мы сравниваем эффективность процесса *динамической настройки* с эффективностью *статической* стратегии. Это сравнение может показаться не вполне корректным. Однако давайте посмотрим, что происходит в стандартной постановке задачи стохастической оптимизации

$$x_\star = \arg\max_{x \in P} \mathcal{E}_\zeta [f(x, \zeta)], \tag{2.2.50}$$

где $\mathcal{E}_\zeta[\cdot]$ обозначает усреднение относительно случайного вектора ζ. Вычисленная стратегия $x_\star$ *обязана* быть статической в силу своего происхождения (в противном случае в максимизации *среднего значения* будет мало смысла). В то же время качество и точность модели $f(x, \xi)$, как правило, построенной с помощью анализа *прошлого*, вряд ли сравнимо с качеством модели, основанной на *точном знании будущего*. Таким образом, по транзитивности мы можем надеяться, что правильная динамическая стратегия настройки даст нам гораздо лучшие результаты, чем стандартное стохастическое программирование. Конечно же, мы имеем ввиду только те ситуации где динамическая настройка возможна.

Основным недостатком стратегии динамической настройки (2.2.45) является ее медленная скорость сходимости. Поэтому она эффективна только в процессах, в которых ожидаемая прибыль по сравнению с числом периодов и параметром барьерной функции достаточно велика. Поэтому перспективная область ее применения скорее относится к долгосрочным процессам планирования производства, а не к игре на фондовом рынке.

2.3 Градиентные методы минимизации составных функций

2.3.1 Введение

Мотивировка

В этом параграфе мы предлагаем новые оптимизационные методы для приближенного нахождения глобальных и локальных минимумов *составной* целевой функции $\phi(x)$. Мы предполагаем, что

$$\phi(x) \quad = \quad f(x) + \Psi(x), \tag{2.3.1}$$

где $f(x)$ – дифференцируемая функция, заданная черноящичным оракулом, а $\Psi(x)$ – выпуклая замкнутая функция общего вида. Однако мы предполагаем, что функция $\Psi(x)$ является *простой*. Это означает, что мы в состоянии находить в явном виде минимум суммы этой функции с простыми вспомогательными функциями. Приведем несколько примеров.

1. Условная минимизация. Пусть множество Q является выпуклым и замкнутым. Определим Ψ как индикаторную функцию множества Q:

$$\Psi(x) \quad = \quad \begin{cases} \quad 0, & \text{если } x \in Q, \\ +\infty, & \text{иначе.} \end{cases}$$

Тогда безусловная минимизация составной функции (2.3.1) эквивалентна минимизации функции f на множестве Q. Позже мы увидим, что наше предположение о простоте функции Ψ в данном случае сводится к условию нахождения евклидовой проекции на множество Q в явном виде.

2. Барьерное представление допустимого множества. Пусть целевая функция выпуклой задачи условной минимизации

$$\text{найти } f^* \quad = \quad \min_{x \in Q} \; f(x)$$

задается череноящичным оракулом, а допустимое множество Q – ν-самосогласованным барьером $F(x)$ (см. параграф 1.3). Положим $\Psi(x) = \frac{\epsilon}{\nu} F(x)$, $\phi(x) = f(x) + \Psi(x)$, и $x^* = \arg\min\limits_{x \in Q} f(x)$. Тогда для любого $\hat{x} \in \operatorname{int} Q$, имеем

$$\begin{aligned} f(\hat{x}) \quad &\leq \quad f(x^*) + \langle \nabla \phi(\hat{x}), \hat{x} - x^* \rangle + \tfrac{\epsilon}{\nu} \langle \nabla F(\hat{x}), x^* - \hat{x} \rangle \\ &\overset{(1.3.25)}{\leq} \quad f^* + \| \nabla \phi(\hat{x}) \|^* \cdot \| \hat{x} - x^* \| + \epsilon. \end{aligned}$$

Таким образом, любая точка $\hat{x}$ с маленькой нормой градиента функции ϕ хорошо аппроксимирует решение задачи условной минимизации. Заметим, что функция ϕ не принадлежит никакому из стандартных классов функций, выделяемых условиями на ограниченность некоторых производных.

3. Разреженные решения в наименьших квадратах. Во многих приложениях необходимо минимизировать следующую целевую функцию:

$$\phi(x) \quad = \quad \tfrac{1}{2}\|Ax - b\|_2^2 + \|x\|_1 \overset{\mathrm{def}}{=} f(x) + \Psi(x), \tag{2.3.2}$$

где A – матрица соответствующего размера и $\|\cdot\|_k$ обозначает стандартную l_k-норму. Хорошо известно, что аддитивная l_1-добавка увеличивает разреженность оптимального решения.

С формальной точки зрения целевая функция $\phi(x)$ в формуле (2.3.2) является негладкой. Поэтому стандартным черноящичным градиентным методам потребуется $O(\frac{1}{\epsilon^2})$ итераций для вычисления ϵ-решения. Методы структурной оптимизации, основанные на технике сглаживания, потребуют $O(\frac{1}{\epsilon})$ итераций (см. гл. 5). В то же время, как мы увидим, та же задача может быть решена за $O(\frac{1}{\epsilon^{1/2}})$ итераций специальных градиентных методов.

Содержание

В п. 2.3.2 мы рассматриваем задачу нахождения локального минимума негладкой невыпуклой функции. Прежде всего мы показываем, что для общих негладких невыпуклых функций даже нахождение направления спуска из точки негладкости является НП-сложной задачей. Однако для специальной формы целевой функции (2.3.1) можно определить *составное градиентное отображение*, которое делает вышеупомянутую проблему разрешимой. Целевая функция этой вспомогательной задачи формируется как сумма целевой функции обычного градиентного отображения (см. определение (1.2.2)) и добавочной функции Ψ. Мы приводим некоторые свойства этого объекта, которые будут полезны в дальнейшем.

В п. 2.3.3 мы рассматриваем простейший градиентный метод, основанный на составном градиентном отображении. Мы показываем, что как в выпуклом, так и в невыпуклом случаях удается воспроизвести в точности те же результаты, что и в обычном гладком случае (т. е. когда $\Psi \equiv 0$). Например, для негладкой функции f максимальный отрицательный уклон функции ϕ для релаксационной минимизирующей последовательности возрастает как $O\left(-\frac{1}{k^{1/2}}\right)$, где k - это счетчик итераций. Таким образом, во всех предельных точках такой последовательности будут отсутствовать направления спуска (см. теорему 2.3.3). Если функция f выпукла и имеет липщицев градиент, то градиентный метод сходится как $O(\frac{1}{k})$. Эта версия градиентного метода снабжена автоматической настройкой длины шага, в которой требуется всего одно дополнительное вычисление значения целевой функции на итерацию.

В п. 2.3.4, мы вводим аппарат модифицированных оценочных функций (ср. с п. 1.2.2), и применяем его сначала для обоснования двойственного варианта градиентного метода. Далее мы описываем быстрый градиентный метод, который сходится как $O(\frac{1}{k^2})$. По сравнению со стандартным вариантом быстрого метода (1.2.18) новая версия может эффективно настраивать исходные оценки для неизвестной константы Липшица для градиента.

В п. 2.3.5 приводятся примеры применения быстрого метода. Мы показываем, как минимизировать сильно выпуклую функцию с известным параметром выпуклости (п. 2.3.5), как находить точку с маленькой невязкой в уравнениях условия оптимальности (п. 2.3.5), и как оценить неизвестный параметр сильной выпуклости (п. 2.3.5). В п. 2.3.6 приводятся результаты численных экспериментов.

Обозначения

В этом параграфе мы используем евклидовы нормы, заданные положительно определенным самосопряженным оператором $B : E \to E^*$:

$$\begin{aligned} \|h\| &= \langle Bh, h\rangle^{1/2}, \quad h \in E, \\ \|s\|_* &= \langle s, B^{-1}s\rangle^{1/2}, \quad s \in E^*. \end{aligned} \tag{2.3.3}$$

2.3.2 Составное градиентное отображение

Задача нахождения направления спуска негладкой невыпуклой функции (или доказательство того, что такого направления не существует) является одной из наиболее сложных задач численного анализа. Чтобы в этом убедиться, рассмотрим следующую ситуацию. Зафиксируем произвольный целочисленный вектор $c \in Z^n_+$ и положим $\gamma = \sum\limits_{i=1}^{n} c^{(i)} \geq 1$. Рассмотрим функцию

$$\phi(x) \;=\; \left(1 - \tfrac{1}{\gamma}\right) \max_{1\leq i\leq n} |x^{(i)}| - \min_{1\leq i\leq n} |x^{(i)}| + |\langle c, x\rangle|. \tag{2.3.4}$$

Ясно что ϕ – кусочно-линейная функция, $\phi(0) = 0$. Тем не менее, имеется следующий пессимистический результат.

Лемма 2.3.1 *Решение вопроса о существовании направления $x \in \mathbb{R}^n$, $\phi(x) < 0$, является НП-сложным.*

Доказательство.
Действительно, пусть вектор $\sigma \in \mathbb{R}^n$ с координатами ± 1 удовлетворяет уравнению $\langle c, \sigma\rangle = 0$. Тогда $\phi(\sigma) = -\frac{1}{\gamma} < 0$.

Предположим теперь, что $\phi(x) < 0$ для некоторого $x \in \mathbb{R}^n$. Всегда можно отшкалировать x так, чтобы выполнилось условие $\max\limits_{1\leq i\leq n} |x^{(i)}| = 1$. Обозначим $\delta = |\langle c, x\rangle|$. Ввиду нашего предположения имеем

$$|x^{(i)}| \overset{(2.3.4)}{>} 1 - \tfrac{1}{\gamma} + \delta, \quad i = 1, \ldots, n.$$

Обозначив $\sigma^{(i)} = \operatorname{sign} x^{(i)}$, $i = 1, \ldots, n$, можно переписать это неравенство в виде $\sigma^{(i)} x^{(i)} > 1 - \frac{1}{\gamma} + \delta$. Таким образом, $|\sigma^{(i)} - x^{(i)}| = 1 - \sigma^{(i)} x^{(i)} < \frac{1}{\gamma} - \delta$, и мы заключаем, что

$$|\langle c, \sigma\rangle| \;\leq\; |\langle c, x\rangle| + |\langle c, \sigma - x\rangle| \;\leq\; \delta + \gamma \max_{1\leq i\leq n} |\sigma^{(i)} - x^{(i)}| \;<\; (1-\gamma)\delta + 1 \;\leq\; 1.$$

Поскольку вектор c целочисленный, это возможно тогда и только тогда, когда $\langle c, \sigma \rangle = 0$. Осталось заметить, что проверка разрешимости целочисленного линейного уравнения в булевских переменных является НП-сложной (это стандартная *задача о камнях*). $\Box$

Таким образом, мы показали, что нахождение направления спуска негладкой невыпуклой функции может быть сложным. Рассматривая теперь функцию $\max\{-1, \phi(x)\}$, мы распространяем этот результат на задачу нахождения локального минимума унимодальной негладкой невыпуклой функции.

Мы видим, что, в отличие от гладкой минимизации, для негладких невыпуклых функций трудно даже обеспечить монотонность минимизирующей последовательности. Поэтому в настоящем параграфе мы ограничиваемся только целевыми функциями с очень специальной структурой. А именно, мы рассматриваем минимизацию функций вида

$$\phi(x) \overset{\text{def}}{=} f(x) + \Psi(x) \tag{2.3.5}$$

на выпуклом множестве Q, где функция f дифференцируема, а функция Ψ выпукла и замкнута на Q. Для характеризации решения нашей задачи введем *конус допустимых направлений* и соответствующий двойственный конус, обычно называемый *нормальным*:

$$\begin{aligned} \mathcal{F}(y) &= \{u = \tau \cdot (x - y),\ x \in Q,\ \tau \geq 0\} \subseteq E, \\ \\ \mathcal{N}(y) &= \{s :\ \langle s, x - y \rangle \geq 0,\ x \in Q\} \subseteq E^*, \quad y \in Q. \end{aligned}$$

Тогда условия оптимальности первого порядка для локального минимума x^* записываются следующим образом:

$$\phi'_* \overset{\text{def}}{=} \nabla f(x^*) + \xi^* \in \mathcal{N}(x^*), \tag{2.3.6}$$

где $\xi^* \in \partial\Psi(x^*)$. Другими словами,

$$\langle \phi'_*, u \rangle \geq 0 \quad \forall u \in \mathcal{F}(x^*). \tag{2.3.7}$$

Поскольку функция Ψ выпукла, последнее условие эквивалентно следующему:

$$D\phi(x^*)[u] \geq 0 \quad \forall u \in \mathcal{F}(x^*). \tag{2.3.8}$$

Заметим, что в случае выпуклой функции f любое из условий (2.3.6) – (2.3.8) является достаточным для того, чтобы точка x^* была точкой глобального минимума функции ϕ на Q.

Последний вариант условия оптимальности хорошо подходит для определения приближенного решения нашей задачи.

Определение 2.3.1 Точка $\bar{x} \in Q$ удовлетворяет условиям оптимальности первого порядка для локального минимума функция ϕ на множестве Q с точностью $\epsilon \geq 0$, если

$$D\phi(\bar{x})[u] \geq -\epsilon \quad \forall u \in \mathcal{F}(\bar{x}),\ \|u\| = 1. \tag{2.3.9}$$

Заметим, что в случае $\mathcal{F}(\bar{x}) = E$, $0 \notin \nabla f(\bar{x}) + \partial \Psi(\bar{x})$, это условие сводится к следующему неравенству:

$$\begin{aligned} -\epsilon &\le \min_{\|u\|=1} D\phi(\bar{x})[u] \; = \; \min_{\|u\|=1} \max_{\xi \in \partial \Psi(\bar{x})} \langle \nabla f(\bar{x}) + \xi, u \rangle \\ &= \min_{\|u\| \le 1} \max_{\xi \in \partial \Psi(\bar{x})} \langle \nabla f(\bar{x}) + \xi, u \rangle \; = \; \max_{\xi \in \partial \Psi(\bar{x})} \min_{\|u\| \le 1} \langle \nabla f(\bar{x}) + \xi, u \rangle \\ &= -\min_{\xi \in \partial \Psi(\bar{x})} \|\nabla f(\bar{x}) + \xi\|_*. \end{aligned}$$

Для нахождения точки $\bar{x}$, удовлетворяющей условию (2.3.9), мы будем пользоваться *составным градиентным отображением.* А именно, для любого $y \in Q$ определим

$$\begin{aligned} m_L(y; x) &= f(y) + \langle \nabla f(y), x - y \rangle + \tfrac{L}{2}\|x - y\|^2 + \Psi(x), \\ T_L(y) &= \arg\min_{x \in Q} \, m_L(y; x), \end{aligned} \tag{2.3.10}$$

где L – положительная константа. Напомним, что в обычном градиентном отображении, предложенном в работе [79], $\Psi(\cdot) \equiv 0$. Теперь мы определим аналог градиентного направления для гладкой функции:

$$g_L(y) \;=\; L \cdot B(y - T_L(y)) \;\in\; E^*, \tag{2.3.11}$$

где оператор $B \succ 0$ определяет норму (2.3.3). (Когда конкретная целевая функция не ясна из контекста, мы пользуемся обозначением $g_L(y)[\phi]$.) Легко заметить, что для $Q \equiv E$ и $\Psi \equiv 0$ мы получаем $g_L(y) = \nabla \phi(y) \equiv \nabla f(x)$ при любых $L > 0$. Наше предположение о простоте функции Ψ означает в точности реализуемость операции (2.3.10).

Перечислим основные свойства составного градиентного отображения. Почти все они вытекают из следующих условий оптимальности для задачи (2.3.10):

$$\langle \nabla f(y) + LB(T_L(y) - y) + \xi_L(y), x - T_L(y) \rangle \;\ge\; 0, \quad \forall x \in Q, \tag{2.3.12}$$

где $\xi_L(y) \in \partial \Psi(T_L(y))$. Обозначим

$$\phi'(T_L(y)) \;=\; \nabla f(T_L(y)) + \xi_L(y) \;\in\; \partial \phi(T_L(y)). \tag{2.3.13}$$

Покажем, что этот объект наследует много важных свойств градиента гладкой функции.

Всюду далее мы предполагаем, что первая составляющая целевой функции (2.3.5) имеет липшицев градиент:

$$\|\nabla f(x) - \nabla f(y)\|_* \;\le\; L_f \|x - y\|, \quad x, y \in Q, \tag{2.3.14}$$

Тогда

$$|f(x) - f(y) - \langle \nabla f(y), x - y \rangle| \;\overset{(1.1.15)}{\le}\; \tfrac{L_f}{2}\|x - y\|^2, \quad x, y \in Q. \tag{2.3.15}$$

Прежде всего оценим локальное изменение функции ϕ. Обозначим

$$S_L(y) \;=\; \tfrac{\|\nabla f(T_L(y)) - \nabla f(y)\|_*}{\|T_L(y) - y\|} \;\le\; L_f.$$

Теорема 2.3.1 *Для любой точки* $y \in Q$ *выполняются неравенства*

$$\phi(y) - \phi(T_L(y)) \;\geq\; \tfrac{2L-L_f}{2L^2}\|g_L(y)\|_*^2, \tag{2.3.16}$$

$$\langle \phi'(T_L(y)), y - T_L(y)\rangle \;\geq\; \tfrac{L-L_f}{L^2}\|g_L(y)\|_*^2. \tag{2.3.17}$$

Более того, для любого $x \in Q$ *верно следующее:*

$$\begin{aligned}\langle \phi'(T_L(y)), x - T_L(y)\rangle \;&\geq\; -\left(1 + \tfrac{1}{L}S_L(y)\right) \cdot \|g_L(y)\|_* \cdot \|T_L(y) - x\| \\ &\geq\; -\left(1 + \tfrac{L_f}{L}\right) \cdot \|g_L(y)\|_* \cdot \|T_L(y) - x\|.\end{aligned} \tag{2.3.18}$$

Доказательство.
Для упрощения обозначений положим $T = T_L(y)$ и $\xi = \xi_L(y)$. Тогда

$$\begin{aligned}\phi(T) \;&\overset{(2.3.14)}{\leq}\; f(y) + \langle \nabla f(y), T - y\rangle + \tfrac{L_f}{2}\|T-y\|^2 + \Psi(T)\\ &\overset{(2.3.12),\, x=y}{\leq}\; f(y) + \langle LB(T-y) + \xi, y - T\rangle + \tfrac{L_f}{2}\|T-y\|^2 + \Psi(T)\\ &=\; f(y) + \tfrac{L_f - 2L}{2}\|T-y\|^2 + \Psi(T) + \langle \xi, y - T\rangle\\ &\leq\; \phi(y) - \tfrac{2L-L_f}{2}\|T-y\|^2.\end{aligned}$$

Принимая во внимание определение (2.3.11), получаем оценку (2.3.16). Далее,

$$\begin{aligned}\langle \nabla f(T) + \xi, y - T\rangle \;&=\; \langle \nabla f(y) + \xi, y - T\rangle - \langle \nabla f(T) - \nabla f(y), T - y\rangle\\ &\overset{(2.3.12),\, x=y}{\geq}\; \langle LB(y-T), y - T\rangle - \langle \nabla f(T) - \nabla f(y), T - y\rangle\\ &\overset{(2.3.14)}{\geq}\; (L - L_f)\|T-y\|^2 \overset{(2.3.11)}{=} \tfrac{L-L_f}{L^2}\|g_L(y)\|_*^2.\end{aligned}$$

Таким образом, мы получили оценку (2.3.17). Наконец,

$$\begin{aligned}\langle \nabla f(T) + \xi, T - x\rangle \;&\overset{(2.3.12)}{\leq}\; \langle \nabla f(T), T - x\rangle + \langle \nabla f(y) + LB(T-y), x - T\rangle\\ &=\; \langle \nabla f(T) - \nabla f(y), T - x\rangle - \langle g_L(y), x - T\rangle\\ &\overset{(2.3.11)}{\leq}\; \left(1 + \tfrac{1}{L}S_L(y)\right) \cdot \|g_L(y)\|_* \cdot \|T - x\|,\end{aligned}$$

и мы приходим к неравенствам (2.3.18). □

Следствие 2.3.1 *Для любых* $y \in Q$ *и* $u \in \mathcal{F}(T_L(y))$, $\|u\| = 1$ *выполняются неравенства*

$$D\phi(T_L(y))[u] \;\; \geq \;\; -\left(1 + \tfrac{L_f}{L}\right) \cdot \|g_L(y)\|_*. \tag{2.3.19}$$

Отметим следующую зависимость величины $\|g_L(y)\|_*$ от L.

Лемма 2.3.2 *Норма градиентного направления* $\|g_L(y)\|_*$ *возрастает по* L*, а норма шага* $\|T_L(y) - y\|$ *убывает по* L*.*

Доказательство.

Действительно, рассмотрим функцию

$$\omega(\tau) \;\; = \;\; \min_{x \in Q} \left[f(y) + \langle \nabla f(y), x - y \rangle + \tfrac{1}{2\tau}\|x - y\|^2 + \Psi(x) \right].$$

Целевая функция в этой задаче минимизации совместно выпукла по x и τ. Таким образом, функция $\omega(\tau)$ является выпуклой по τ. Поскольку минимум в этой задаче достигается в единственной точке, функция $\omega(\tau)$ дифференцируема и

$$\omega'(\tau) \;\; = \;\; -\tfrac{1}{2}\|\tfrac{1}{\tau}[T_{1/\tau}(y) - y]\|^2 \;\; = \;\; -\tfrac{1}{2}\|g_{1/\tau}(y)\|_*^2.$$

Так как функция $\omega(\cdot)$ выпукла, $\omega'(\tau)$ возрастает по τ. Поэтому $\|g_{1/\tau}(y)\|_*$ убывает по τ.

Второе утверждение вытекает из вогнутости функции

$$\hat{\omega}(L) \;\; = \;\; \min_{x \in Q} \left[f(y) + \langle \nabla f(y), x - y \rangle + \tfrac{L}{2}\|x - y\|^2 + \Psi(x) \right]. \qquad \Box$$

Посмотрим теперь на глобальные свойства составного градиентного отображения.

Теорема 2.3.2 *При любых* $y \in Q$ *выполняются неравенства*

$$m_L(y; T_L(y)) \;\; \leq \;\; \phi(y) - \tfrac{1}{2L}\|g_L(y)\|_*^2, \tag{2.3.20}$$

$$m_L(y; T_L(y)) \;\; \leq \;\; \min_{x \in Q} \left[\phi(x) + \tfrac{L + L_f}{2}\|x - y\|^2 \right]. \tag{2.3.21}$$

Если функция f *выпукла, то*

$$m_L(y; T_L(y)) \;\; \leq \;\; \min_{x \in Q} \left[\phi(x) + \tfrac{L}{2}\|x - y\|^2 \right]. \tag{2.3.22}$$

Доказательство.

Заметим, что функция $m_L(y; x)$ сильно выпукла по x с параметром выпуклости L. Поэтому

$$\phi(y) - m_L(y; T_L(y)) \;\; = \;\; m_L(y; y) - m_L(y; T_L(y)) \;\; \geq \;\; \tfrac{L}{2}\|y - T_L(y)\|^2 \;\; = \;\; \tfrac{1}{2L}\|g_L(y)\|_*^2.$$

Далее, если функция f выпукла, то

$$\begin{aligned} m_L(y; T_L(y)) \;\; &= \;\; \min_{x \in Q} \left[f(y) + \langle \nabla f(y), x - y \rangle + \tfrac{L}{2}\|x - y\|^2 + \Psi(x) \right] \\ &\leq \;\; \min_{x \in Q} \left[f(x) + \Psi(x) + \tfrac{L}{2}\|x - y\|^2 \right] \\ &= \;\; \min_{x \in Q} \left[\phi(x) + \tfrac{L}{2}\|x - y\|^2 \right]. \end{aligned}$$

Для невыпуклой функции f, мы можем воспользоваться непосредственным следствием неравенства (2.3.15):

$$f(y) + \langle \nabla f(y), x - y \rangle \;\; \leq \;\; f(x) + \tfrac{L_f}{2}\|x - y\|^2. \qquad \Box$$

Замечание 2 В силу оценки (2.3.14) для $L \geq L_f$ имеем

$$\phi(T_L(y)) \;\; \leq \;\; m_L(y; T_L(y)). \tag{2.3.23}$$

Поэтому в данном случае неравенство (2.3.22) гарантирует, что

$$\phi(T_L(y)) \;\; \leq \;\; \min_{x \in Q} \left[\phi(x) + \tfrac{L}{2}\|x - y\|^2 \right]. \tag{2.3.24}$$

Наконец, докажем следующее полезное неравенство для сильно выпуклой функции ϕ.

Лемма 2.3.3 *Пусть функция ϕ сильно выпукла с параметром выпуклости $\mu_\phi > 0$. Тогда для любых $y \in Q$ выполняются неравенства*

$$\|T_L(y) - x^*\| \;\; \leq \;\; \tfrac{1}{\mu_\phi} \cdot \left(1 + \tfrac{1}{L} S_L(y)\right) \cdot \|g_L(y)\|_* \; \leq \; \tfrac{1}{\mu_\phi} \cdot \left(1 + \tfrac{L_f}{L}\right) \cdot \|g_L(y)\|_*, \tag{2.3.25}$$

где x^* – единственный минимум функции ϕ на Q.

Доказательство.
Действительно, в силу неравенства (2.3.18) имеем

$$\begin{aligned} \left(1 + \tfrac{L_f}{L}\right) \cdot \|g_L(y)\|_* \cdot \|T_L(y) - x^*\| \;\; &\geq \;\; \left(1 + \tfrac{1}{L} S_L(y)\right) \cdot \|g_L(y)\|_* \cdot \|T_L(y) - x^*\| \\ &\geq \;\; \langle \phi'(T_L(y)), T_L(y) - x^* \rangle \; \geq \; \mu_\phi \|T_L(y) - x^*\|^2, \end{aligned}$$

и мы получаем (2.3.25). $\Box$

Теперь мы можем анализировать разные методы минимизации, использующие составное градиентное отображение. Начинаем с самого простого метода, который будет представлен в следующем пункте.

2.3.3 Составной градиентный метод

Рассмотрим сначала градиентный шаг с простейшей стратегией удваивания шагового параметра с восстановлением. Условие прерывания этого шага называется *полной релакса-*

цией).

Градиентная итерация $\mathcal{G}(x, M)$

Полагаем: $L := M.$

Повторяем: $T := T_L(x),$

если $\phi(T) > m_L(x; T)$**, то** $L := L \cdot \gamma_u,$ (2.3.26)

Пока не выполнится: $\phi(T) \leq m_L(x; T).$

Выход: $\mathcal{G}(x, M).T \;=\; T, \quad \mathcal{G}(x, M).L \;=\; L,$

$\mathcal{G}(x, M).S \;=\; S_L(x).$

Если используемая целевая функция ясна из контекста, мы пользуемся сокращенным обозначением $\mathcal{G}_\phi(x, M)$.

Для запуска градиентного метода нам потребуются начальная оптимистическая оценка L_0 для константы Липшица L_f:

$$0 \;<\; L_0 \;\;\leq\;\; L_f, \tag{2.3.27}$$

и два параметра для процедуры подбора длины шага $\gamma_u > 1$ и $\gamma_d \geq 1$. Пусть нам известна стартовая точка $y_0 \in Q$. При $k \geq 0$ рассмотрим следующий итеративный процесс.

Прямой градиентный метод $\mathcal{GM}(y_0, L_0)$

$$\begin{aligned} y_{k+1} &= \mathcal{G}(y_k, L_k).T, \\ M_k &= \mathcal{G}(y_k, L_k).L, \\ L_{k+1} &= \max\{L_0, M_k/\gamma_d\}. \end{aligned} \tag{2.3.28}$$

Таким образом, $y_{k+1} = T_{M_k}(y_k)$. Поскольку функция f удовлетворяет неравенству (2.3.15), в цикле (2.3.26) значение L увеличивается, только если $L \leq L_f$. Принимая во внимание условие (2.3.27), мы приходим к следующим границам:

$$L_0 \leq L_k \;\leq\; M_k \;\leq\; \gamma_u L_f. \tag{2.3.29}$$

Более того, если $\gamma_d \geq \gamma_u$, то

$$L_k \;\leq\; L_f, \quad k \geq 0. \tag{2.3.30}$$

Заметим, что в методе (2.3.26) отсутствуют какие-либо ограничения на число повторений цикла. Однако легко видеть, что общее число вызовов оракула N_k после k итераций метода (2.3.28) не может быть большим.

Лемма 2.3.4 *В методе (2.3.28) при любом* $k \geq 0$ *выполняется неравенство*

$$N_k \;\leq\; \left[1 + \frac{\ln \gamma_d}{\ln \gamma_u}\right] \cdot (k+1) + \frac{1}{\ln \gamma_u} \cdot \left(\ln \frac{\gamma_u L_f}{\gamma_d L_0}\right)_+. \tag{2.3.31}$$

Доказательство.
Обозначим через $n_i \geq 1$ число вызовов оракула на итерации $i \geq 0$. Тогда

$$L_{i+1} \;\geq\; \frac{1}{\gamma_d} \cdot L_i \cdot \gamma_u^{n_i - 1}.$$

Таким образом,

$$n_i \;\leq\; 1 + \frac{\ln \gamma_d}{\ln \gamma_u} + \frac{1}{\ln \gamma_u} \cdot \ln \frac{L_{i+1}}{L_i}.$$

Поэтому мы можем оценить

$$N_k \;=\; \sum_{i=0}^{k} n_i \;=\; \left[1 + \frac{\ln \gamma_d}{\ln \gamma_u}\right] \cdot (k+1) + \frac{1}{\ln \gamma_u} \cdot \ln \frac{L_{k+1}}{L_0}.$$

Осталось заметить, что $L_{k+1} \overset{(2.3.29)}{\leq} \max\left\{L_0, \frac{\gamma_u}{\gamma_d} L_f\right\}$. □

Обычно параметры настройки выбирают следующим образом:

$$\gamma_u \;=\; \gamma_d \;=\; 2 \;\overset{(2.3.31)}{\Rightarrow}\; N_k \;\leq\; 2(k+1) + \log_2 \frac{L_f}{L_0}, \quad L_k \;\overset{(2.3.30)}{\leq}\; L_f. \tag{2.3.32}$$

Итак, вычислительные затраты градиентного метода (2.3.28) хорошо описываются оценками на число итераций. В оставшейся части этого пункта мы получим эти оценки для различных ситуаций.

Начнем с общего невыпуклого случая. Обозначим

$$\delta_k \;=\; \min_{0 \leq i \leq k} \frac{1}{2M_i} \|g_{M_i}(y_i)\|_*^2,$$

$$i_k \;=\; 1 + \arg\min_{0 \leq i \leq k} \frac{1}{2M_i} \|g_{M_i}(y_i)\|_*^2.$$

Теорема 2.3.3 *Пусть функция ϕ ограничена снизу на Q некоторым числом ϕ_*. Тогда*

$$\delta_k \quad \leq \quad \tfrac{\phi(y_0)-\phi^*}{k+1}. \tag{2.3.33}$$

Более того, для любого $u \in \mathcal{F}(y_{i_k})$, $\|u\| = 1$, выполняется неравенство

$$D\phi(y_{i_k})[u] \quad \geq \quad -\tfrac{(1+\gamma_u)L_f}{L_0^{1/2}} \cdot \sqrt{\tfrac{2(\phi(y_0)-\phi_*)}{k+1}}. \tag{2.3.34}$$

Доказательство.
Действительно, в силу критерия прерывания в методе (2.3.26) имеем

$$\phi(y_i) - \phi(y_{i+1}) \quad \geq \quad \phi(y_i) - m_{M_i}(y_i; T_{M_i}(y_i)) \quad \overset{(2.3.20)}{\geq} \quad \tfrac{1}{2M_i}\|g_{M_i}(y_i)\|_*^2.$$

Просуммировав эти неравенства по $i = 0, \ldots, k$, получим оценку (2.3.33).

Обозначим $j_k = i_k - 1$. Поскольку $y_{i_k} = T_{M_{j_k}}(y_{j_k})$, при любом $u \in \mathcal{F}(y_{i_k})$, $\|u\| = 1$, имеем

$$\begin{aligned} D\phi(y_{i_k})[u] \quad & \overset{(2.3.19)}{\geq} \quad -\left(1 + \tfrac{L_f}{M_{j_k}}\right) \cdot \|g_{M_{j_k}}(y_{j_k})\|_* \; = \; -\left(1 + \tfrac{L_f}{M_{j_k}}\right) \cdot \sqrt{2M_{j_k}\delta_k} \\ & \overset{(2.3.33)}{\geq} \quad -\tfrac{M_{j_k}+L_f}{M_{j_k}^{1/2}} \cdot \sqrt{\tfrac{2(\phi(y_0)-\phi_*)}{k+1}} \quad \overset{(2.3.29)}{\geq} \quad -\tfrac{(1+\gamma_u)L_f}{L_0^{1/2}} \cdot \sqrt{\tfrac{2(\phi(y_0)-\phi_*)}{k+1}}. \qquad \square \end{aligned}$$

Опишем теперь поведение градиентного метода (2.3.28) в выпуклом случае.

Теорема 2.3.4 *Пусть функция f выпукла на Q. Предположим, что она достигает своего минимума на Q в точке x^* и что ее множества уровня ограничены:*

$$\|y - x^*\| \quad \leq \quad R \quad \forall y \in Q: \; \phi(y) \leq \phi(y_0). \tag{2.3.35}$$

Если $\phi(y_0) - \phi(x^*) \geq \gamma_u L_f R^2$, то $\phi(y_1) - \phi(x^*) \leq \frac{\gamma_u L_f R^2}{2}$. В противном случае при всех $k \geq 0$ выполняется неравенство

$$\phi(y_k) - \phi(x^*) \quad \leq \quad \tfrac{2\gamma_u L_f R^2}{k+2}. \tag{2.3.36}$$

Более того, для любого $u \in \mathcal{F}(y_{i_k})$, $\|u\| = 1$, выполняется неравенство

$$D\phi(y_{i_k})[u] \quad \geq \quad -\tfrac{4(1+\gamma_u)L_f R}{[(k+2)(k+4]^{1/2}} \cdot \sqrt{\gamma_u \tfrac{L_f}{L_0}}. \tag{2.3.37}$$

Доказательство.
Поскольку $\phi(y_{k+1}) \leq \phi(y_k)$ при всех $k \geq 0$, оценка $\|y_k - x^*\| \leq R$ верна для всех точек, полученных методом (2.3.28). Рассмотрим

$$y_k(\alpha) \quad = \quad \alpha x^* + (1-\alpha)y_k \; \in \; Q \quad \alpha \in [0,1].$$

Тогда

$$\begin{aligned} \phi(y_{k+1}) \quad & \leq \quad m_{M_k}(y_k; T_{M_k}(y_k)) \overset{(2.3.22)}{\leq} \min_{y \in Q} \left[\phi(y) + \tfrac{M_k}{2} \|y - y_k\|^2 \right] \\ (y = y_k(\alpha)) \quad & \leq \quad \min_{0 \leq \alpha \leq 1} \left[\phi(\alpha x^* + (1-\alpha) y_k) + \tfrac{M_k \alpha^2}{2} \|y_k - x^*\|^2 \right] \\ & \overset{(2.3.29)}{\leq} \quad \min_{0 \leq \alpha \leq 1} \left[\phi(y_k) - \alpha(\phi(y_k) - \phi(x^*)) + \tfrac{\gamma_u L_f R^2}{2} \cdot \alpha^2 \right]. \end{aligned}$$

Если $\phi(y_0) - \phi(x^*) \geq \gamma_u L_f R^2$, то оптимальным значением последней оптимизационной задачи является $\alpha = 1$ и мы получаем

$$\phi(y_1) - \phi(x^*) \quad \leq \quad \tfrac{\gamma_u L_f R^2}{2}.$$

В противном случае оптимальное значение есть

$$\alpha \quad = \quad \tfrac{\phi(y_k) - \phi(x^*)}{\gamma_u L_f R^2} \quad \leq \quad \tfrac{\phi(y_0) - \phi(x^*)}{\gamma_u L_f R^2} \leq 1$$

и мы получаем

$$\phi(y_{k+1}) \quad \leq \quad \phi(y_k) - \tfrac{[\phi(y_k) - \phi(x^*)]^2}{2\gamma_u L_f R^2}. \tag{2.3.38}$$

Обозначив $\lambda_k = \frac{1}{\phi(y_k) - \phi(x^*)}$, из этого неравенства выводим, что

$$\lambda_{k+1} \quad \geq \quad \lambda_k + \tfrac{\lambda_{k+1}}{2\lambda_k \gamma_u L_f R^2} \quad \geq \quad \lambda_k + \tfrac{1}{2\gamma_u L_f R^2}.$$

Поэтому для $k \geq 0$ имеем

$$\lambda_k \quad \geq \quad \tfrac{1}{\phi(y_0) - \phi(x^*)} + \tfrac{k}{2\gamma_u L_f R^2} \quad \geq \quad \tfrac{k+2}{2\gamma_u L_f R^2}.$$

Далее, зафиксируем целое m, $0 < m < k$. Поскольку

$$\phi(y_i) - \phi(y_{i+1}) \quad \geq \quad \tfrac{1}{2M_i} \|g_{M_i}(y_i)\|_*^2, \quad i = 0, \ldots, k,$$

имеем

$$\begin{aligned} (k - m + 1)\delta_k \quad & \leq \quad \sum_{i=m}^{k} \tfrac{1}{2M_i} \|g_{M_i}(y_i)\|_*^2 \\ & \leq \quad \phi(y_m) - \phi(y_{k+1}) \\ & \leq \quad \phi(y_m) - \phi(x^*) \\ & \overset{(2.3.36)}{\leq} \quad \tfrac{2\gamma_u L_f R^2}{m+2}. \end{aligned}$$

Обозначим $j_k = i_k - 1$. Тогда для любого $u \in \mathcal{F}(y_{i_k})$, $\|u\| = 1$, получаем

$$\begin{aligned} D\phi(y_{i_k})[u] &\overset{(2.3.19)}{\geq} -\left(1 + \tfrac{L_f}{M_{j_k}}\right) \cdot \|g_{M_{j_k}}(y_{j_k})\|_* \\ &= -\left(1 + \tfrac{L_f}{M_{j_k}}\right) \cdot \sqrt{2M_{j_k}\delta_k} \\ &\overset{(2.3.36)}{\geq} -2\tfrac{M_{j_k}+L_f}{M_{j_k}^{1/2}} \cdot \sqrt{\tfrac{\gamma_u L_f R^2}{(m+2)(k+1-m)}} \\ &\overset{(2.3.29)}{\geq} -2(1+\gamma_u)L_f R \cdot \sqrt{\tfrac{\gamma_u L_f}{L_0(m+2)(k+1-m)}}. \end{aligned}$$

Выбирая $m = \lfloor \frac{k}{2} \rfloor$, получаем $(m+2)(k+1-m) \geq \frac{(k+2)(k+4)}{4}$. □

Теорема 2.3.5 *Пусть функция ϕ сильно выпукла на Q с параметром выпуклости μ_ϕ. Если $\frac{\mu_\phi}{L_f} \geq 2\gamma_u$, то при любых $k \geq 0$ выполняются неравенства*

$$\phi(y_k) - \phi(x^*) \;\leq\; \left(\tfrac{\gamma_u L_f}{\mu_\phi}\right)^k (\phi(y_0) - \phi(x^*)) \;\leq\; \tfrac{1}{2^k}(\phi(y_0) - \phi(x^*)). \tag{2.3.39}$$

В противном случае

$$\phi(y_k) - \phi(x^*) \;\leq \left(1 - \tfrac{\mu_\phi}{4\gamma_u L_f}\right)^k \cdot (\phi(y_0) - \phi(x^*)). \tag{2.3.40}$$

Доказательство.

Поскольку функция ϕ строго выпукла, при всех $k \geq 0$ имеем

$$\phi(y_k) - \phi(x^*) \;\geq\; \tfrac{\mu_\phi}{2}\|y_k - x^*\|^2. \tag{2.3.41}$$

Обозначим $y_k(\alpha) = \alpha x^* + (1-\alpha)y_k \in Q$, $\alpha \in [0,1]$. Тогда

$$\begin{aligned} \phi(y_{k+1}) &\overset{(2.3.22)}{\leq} \min_{0\leq\alpha\leq1} \left[\phi(\alpha x^* + (1-\alpha)y_k) + \tfrac{M_k\alpha^2}{2}\|y_k - x^*\|^2\right] \\ &\overset{(2.3.29)}{\leq} \min_{0\leq\alpha\leq1} \left[\phi(y_k) - \alpha(\phi(y_k) - \phi(x^*)) + \tfrac{\gamma_u L_f}{2} \cdot \alpha^2\|y_k - x^*\|^2\right] \\ &\overset{(2.3.41)}{\leq} \min_{0\leq\alpha\leq1} \left[\phi(y_k) - \alpha\left(1 - \alpha \cdot \tfrac{\gamma_u L_f}{\mu_\phi}\right)(\phi(y_k) - \phi(x^*))\right]. \end{aligned}$$

Минимум последнего выражения достигается при $\alpha^* = \min\left\{1, \frac{\mu_\phi}{2\gamma_u L_f}\right\}$. Поэтому если $\frac{\mu_\phi}{2\gamma_u L_f} \geq 1$, то $\alpha^* = 1$ и мы получаем

$$\phi(y_{k+1}) - \phi(x^*) \;\leq\; \tfrac{\gamma_u L_f}{\mu_\phi}(\phi(y_k) - \phi(y^*)) \;\leq\; \tfrac{1}{2}(\phi(y_k) - \phi(y^*)).$$

Если $\frac{\mu_\phi}{2\gamma_u L_f} \leq 1$, то $\alpha^* = \frac{\mu_\phi}{2\gamma_u L_f}$ и $\phi(y_{k+1}) - \phi(x^*) \leq \left(1 - \frac{\mu_\phi}{4\gamma_u L_f}\right) \cdot (\phi(y_k) - \phi(y^*))$. □

Замечание 3 1) В теореме 2.3.5 число обусловленности $\frac{L_f}{\mu_\phi}$ может быть меньше единицы.

2) Для сильно выпуклой функции ϕ оценки для производных по направлению могут быть получены с помощью неравенств (2.3.39), (2.3.40) с учетом того факта, что

$$\phi(y_k) - \phi(x^*) \quad \overset{(2.3.16):L=L_f}{\geq} \quad \tfrac{1}{2L_f}\|g_{L_f}(y_k)\|_*^2$$

и неравенства (2.3.19). Таким образом, неравенство (2.3.39) приводит к оценке

$$D\phi(y_{k+1})[u] \quad \geq \quad -2\left(\tfrac{\gamma_u L_f}{\mu_\phi}\right)^{k/2} \cdot \sqrt{2L_f(\phi(y_0) - \phi_*)}, \tag{2.3.42}$$

а неравенство (2.3.40) дает

$$D\phi(y_{k+1})[u] \quad \geq \quad -2\left(1 - \tfrac{\mu_\phi}{4\gamma_u L_f}\right)^{k/2} \cdot \sqrt{2L_f(\phi(y_0) - \phi_*)}, \tag{2.3.43}$$

что верно для всех $u \in \mathcal{F}(y_{k+1})$ с $\|u\| = 1$.

3) Оценка (2.3.31) может создать впечатление, что большое значение параметра γ_u может уменьшить общее число обращений к оракулу. Это не так, поскольку γ_u также входит в оценку скорости сходимости соответствующих методов (см. например, оценку (2.3.36)). Таким образом, разумные значения этого параметра лежат на отрезке $[2, 3]$.

2.3.4 Быстрый градиентный метод

В предыдущем пункте мы показали, что для выпуклой функции f градиентный метод (2.3.28) сходится со скоростью $O(\frac{1}{k})$. Однако хорошо известно, что на выпуклых задачах этот метод не самый быстрый. Покажем, что аналогичное ускорение может быть достигнуто и для составных функций.

Рассмотрим задачу

$$\min_{x \in E} \quad [\ \phi(x) = f(x) + \Psi(x)\ ], \tag{2.3.44}$$

где функция f выпукла и удовлетворяет условию (2.3.14), а функция Ψ замкнута и сильно выпукла на E с параметром выпуклости $\mu_\Psi \geq 0$. Будем считать, что этот параметр известен. Случай $\mu_\Psi = 0$ соответствует выпуклой функции Ψ. Обозначим через x^* оптимальное решение задачи (2.3.44).

В задаче (2.3.44) допускается случай $\mathrm{dom}\,\Psi \neq E$, т. е. формулировка (2.3.44) включает в себя также задачи с ограничениями. Для задачи (2.3.44) условия оптимальности (2.3.12), определяющие составное градиентное отображение, записываются в более простой форме:

$$\begin{array}{rcl} T_L(y) & \in & \mathrm{dom}\,\Psi, \\ \\ \nabla f(y) + \xi_L(y) & = & LB(y - T_L(y)) \ \equiv \ g_L(y), \end{array} \tag{2.3.45}$$

где $\xi_L(y) \in \partial\Psi(T_L(y))$.

Для доказательства оценок скорости сходимости разных методов, применяемых для решения задачи (2.3.44), мы будем пользоваться аппаратом оценочных функций (ср. с п. 1.2.2). Принимая во внимание специальную форму целевой функции в задаче (2.3.44), мы будем рекурсивно пересчитывать следующие последовательности:

- минимизирующую последовательность $\{x_k\}_{k=0}^{\infty}$;
- последовательность возрастающих масштабных коэффициентов $\{A_k\}_{k=0}^{\infty}$:

$$A_0 = 0, \quad A_k \stackrel{\rm def}{=} A_{k-1} + a_k, \quad k \geq 1;$$

- последовательность оценочных функций

$$\psi_k(x) \;=\; l_k(x) + A_k \Psi(x) + \tfrac{1}{2}\|x - x_0\|^2 \quad k \geq 0, \tag{2.3.46}$$

  где $x_0 \in \operatorname{dom} \Psi$ – стартовая точка, а функции $l_k(x)$ линейны по $x \in E$.

Однако, по сравнению с п. 1.2.2, мы добавим возможность настройки оценок для константы Липшица L_f, использующую начальный оптимистический прогноз L_0, удовлетворяющий (2.3.27), и два параметра настройки $\gamma_u > 1$ и $\gamma_d \geq 1$.

Для этих объектов мы будем рекурсивно поддерживать выполнение соотношений

$$\left.\begin{array}{rrcl} \mathcal{R}_k^1: & A_k \phi(x_k) & \leq & \psi_k^* \equiv \min\limits_x \psi_k(x), \\ \\ \mathcal{R}_k^2: & \psi_k(x) & \leq & A_k \phi(x) + \frac{1}{2}\|x - x_0\|^2 \; \forall x \in E \end{array}\right\}, \quad k \geq 0. \tag{2.3.47}$$

Из них очевидным образом вытекает следующая оценка скорости сходимости:

$$\phi(x_k) - \phi(x^*) \;\leq\; \tfrac{\|x^* - x_0\|^2}{2A_k}, \quad k \geq 1. \tag{2.3.48}$$

Обозначим $v_k = \arg\min\limits_{x \in E} \; \psi_k(x)$. Поскольку $\mu_{\psi_k} \geq 1$, для любого $x \in E$ имеем

$$A_k \phi(x_k) + \tfrac{1}{2}\|x - v_k\|^2 \;\overset{\mathcal{R}_k^1}{\leq}\; \psi_k^* + \tfrac{1}{2}\|x - v_k\|^2 \;\leq\; \psi_k(x) \;\overset{\mathcal{R}_k^2}{\leq}\; A_k \phi(x) + \tfrac{1}{2}\|x - x_0\|^2.$$

Поэтому, выбирая $x = x^*$, получаем два полезных следствия соотношений (2.3.47):

$$\|x^* - v_k\| \leq \|x^* - x_0\|, \qquad \|v_k - x_0\| \leq 2\|x^* - x_0\|, \quad k \geq 1. \tag{2.3.49}$$

Заметим, что соотношения (2.3.47) могут использоваться для обоснования скорости сходимости двойственного варианта градиентного метода (2.3.28). Для $v_0 \in \operatorname{dom} \Psi$ положим $\psi_0(x) = \frac{1}{2}\|x - v_0\|^2$ и выберем оценку L_0 удовлетворяющую условию (2.3.27).

Двойственный градиентный метод $\mathcal{DG}(v_0, L_0)$, $k \geq 0$.

$$\begin{array}{rcl} y_k & = & \mathcal{G}(v_k, L_k).T, \quad M_k \;=\; \mathcal{G}(v_k, L_k).L, \\ L_{k+1} & = & \max\{L_0, M_k/\gamma_d\}, \quad a_{k+1} \;=\; \frac{1}{M_k}, \\ \psi_{k+1}(x) & = & \psi_k(x) + \frac{1}{M_k}[f(v_k) + \langle \nabla f(v_k), x - v_k\rangle + \Psi(x)]. \end{array} \tag{2.3.50}$$

В этой схеме операция $\mathcal{G}$ определена в схеме (2.3.26). Поскольку функция Ψ простая, точки v_k легко вычислимы.

Заметим, что соотношения $\mathcal{R}_0^1$ и $\mathcal{R}_k^2$, $k \geq 0$, выполнены по тривиальным причинам. Соотношения $\mathcal{R}_k^1$ могут быть обоснованы по индукции. Положим $x_0 = y_0$, $\phi_k = \min\limits_{0 \leq i \leq k-1} \phi(y_i)$, и $x_k : \phi(x_k) = \phi_k$ для $k \geq 1$. Тогда

$$\begin{array}{rcl} \psi_{k+1}^* & = & \min\limits_x \left\{ \psi_k(x) + \frac{1}{M_k}[f(v_k) + \langle \nabla f(v_k), x - v_k\rangle + \Psi(x)] \right\} \\ & \overset{\mathcal{R}_k^1}{\geq} & A_k\phi_k + \min\limits_x \left\{ \frac{1}{2}\|x - v_k\|^2 + \frac{1}{M_k}[f(v_k) + \langle \nabla f(v_k), x - v_k\rangle + \Psi(x)] \right\} \\ & \overset{(2.3.10)}{=} & A_k\phi_k + a_{k+1} m_{M_k}(v_k; y_k) \\ & \overset{(2.3.26)}{\geq} & A_k\phi_k + a_{k+1}\phi(y_k) \;\geq\; A_{k+1}\phi_{k+1}. \end{array}$$

Таким образом, соотношения $\mathcal{R}_k^1$ выполнены при всех $k \geq 0$. Поскольку значения M_k удовлетворяют неравенствам (2.3.29), для метода (2.3.50) получается такая оценка скорости сходимости:

$$\phi(x_k) - \phi(x^*) \;\leq\; \frac{\gamma_u L_f}{2k}\|x^* - v_0\|^2, \quad k \geq 1. \tag{2.3.51}$$

Заметим, что константа в правой части этого неравенства в четыре раза меньше, чем константа в (2.3.36). Однако каждая итерация двойственного метода в два раза сложнее чем итерация прямого метода (2.3.28).

Заметим, что метод (2.3.50) не использует все возможности аппарата оценочных функций. Рассмотрим ускоренную версию метода (2.3.50). Эта версия требует задания стартовой точки $x_0 \in \operatorname{dom}\Psi$, нижней оценки $L_0 > 0$ для константы Липшица L_f и нижней оценки $\mu \in [0, \mu_\Psi]$ для параметра выпуклости функции Ψ.

Быстрый градиентный метод $\mathcal{A}(x_0, L_0, \mu)$

```
Инициализация   $\psi_0(x) = \frac{1}{2}\|x - x_0\|^2,\ A_0 = 0.$

Итерация $(k \geq 0)$

ПОЛАГАЕМ:  $L := L_k$.
ПОВТОР:    найти $a$ из квадратного уравнения $\frac{a^2}{A_k + a} = 2\frac{1 + \mu A_k}{L}$,   $(*)$
           полагаем $y = \frac{A_k x_k + a v_k}{A_k + a}$, и вычисляем $T_L(y)$.
ЕСЛИ:      $\langle \phi'(T_L(y)), y - T_L(y) \rangle \geq \frac{1}{L}\|\phi'(T_L(y))\|_*^2$,   $(**)$
           ТО СТОП ПОВТОР ИНАЧЕ $L := L \cdot \gamma_u$.
ПОЛАГАЕМ:  $y_k := y,\ M_k := L,\ a_{k+1} := a$,
           $L_{k+1} := M_k / \gamma_d,\ x_{k+1} := T_{M_k}(y_k)$,
           $\psi_{k+1}(x) := \psi_k(x) + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle + \Psi(x)]$.
```

(2.3.52)

По сравнению с градиентной итерацией (2.3.26) для остановки внутреннего цикла здесь используется *усиленное условие релаксации* $(**)$.

Лемма 2.3.5 *Условие (**) в методе (2.3.52) выполнено при* $L \geq L_f$.

Доказательство.
Обозначим $T = T_L(y)$. Умножая представление

$$\phi'(T) \;=\; \nabla f(T) + \xi_L(y) \overset{(2.3.45)}{=} LB(y - T) + \nabla f(T) - \nabla f(y) \tag{2.3.53}$$

на вектор $y - T$, получаем

$$\begin{aligned}\langle \phi'(T), y - T\rangle &= L\|y - T\|^2 - \langle \nabla f(y) - \nabla f(T), y - T\rangle \\ &\overset{(2.3.53)}{=} \tfrac{1}{L}\left[\|\phi'(T)\|_*^2 + 2L\langle \nabla f(y) - \nabla f(T), y - T\rangle - \|\nabla f(y) - \nabla f(T)\|_*^2\right] \\ &\quad -\langle \nabla f(y) - \nabla f(T), y - T\rangle \\ &= \tfrac{1}{L}\|\phi'(T)\|_*^2 + \langle \nabla f(y) - \nabla f(T), y - T\rangle - \tfrac{1}{L}\|\nabla f(y) - \nabla f(T)\|_*^2.\end{aligned}$$

Поэтому для $L \geq L_f$ условие (**) выполнено. □

Таким образом, мы всегда можем гарантировать выполнение неравенств

$$L_k \;\leq\; M_k \;\leq\; \gamma_u L_f. \tag{2.3.54}$$

Если $\gamma_d \geq \gamma_u$, то верхняя граница в формуле (2.3.30) остается справедливой.

Установим связь между общим числом вызовов оракула N_k после k итераций и количеством итераций.

Лемма 2.3.6 *В методе (2.3.52) при любом $k \geq 0$ выполняется соотношение*

$$N_k \;\leq\; 2\left[1 + \tfrac{\ln \gamma_d}{\ln \gamma_u}\right] \cdot (k+1) + \tfrac{2}{\ln \gamma_u} \cdot \ln \tfrac{\gamma_u L_f}{\gamma_d L_0}. \tag{2.3.55}$$

Доказательство.

Обозначим через $n_i \geq 1$ число вызовов оракула на итерации $i \geq 0$. При каждом проходе внутреннего цикла оракул вызывается дважды, для вычисления $\nabla f(y)$ и для вычисления $\nabla f(T_L(y))$. Таким образом,

$$L_{i+1} \;=\; \tfrac{1}{\gamma_d} \cdot L_i \cdot \gamma_u^{0.5 n_i - 1}.$$

Следовательно,

$$n_i \;=\; 2\left[1 + \tfrac{\ln \gamma_d}{\ln \gamma_u} + \tfrac{1}{\ln \gamma_u} \cdot \ln \tfrac{L_{i+1}}{L_i}\right].$$

Поэтому

$$N_k \;=\; \sum_{i=0}^{k} n_i \;=\; 2\left[1 + \tfrac{\ln \gamma_d}{\ln \gamma_u}\right] \cdot (k+1) + \tfrac{2}{\ln \gamma_u} \cdot \ln \tfrac{L_{k+1}}{L_0}.$$

Осталось заметить, что $L_{k+1} \overset{(2.3.54)}{\leq} \frac{\gamma_u}{\gamma_d} L_f$. □

Таким образом, каждая итерация метода (2.3.52) требует примерно в два раза больше вызовов оракула чем одна итерация градиентного метода:

$$\gamma_u \;=\; \gamma_d \;= 2 \quad \Rightarrow \quad N_k \;\leq\; 4(k+1) + 2\log_2 \tfrac{L_f}{L_0}, \quad L_k \;\leq L_f. \tag{2.3.56}$$

Однако скоро мы увидим, что скорость сходимости метода (2.3.52) гораздо выше.

Начнем с двух вспомогательных утверждений.

Лемма 2.3.7 *Пусть* $\mu_\Psi \geq \mu$. *Тогда последовательности* $\{x_k\}$, $\{A_k\}$ *и* $\{\psi_k\}$, *сгенерированные методом* $\mathcal{A}(x_0, L_0, \mu)$, *удовлетворяют соотношениям (2.3.47) при всех* $k \geq 0$.

Доказательство.

Действительно, в силу выбора начальных параметров в методе (2.3.52), $A_0 = 0$ и $\psi_0^* = 0$. Поэтому при $k = 0$, оба соотношения (2.3.47) тривиальны.

Предположим, что соотношения $\mathcal{R}_k^1$, $\mathcal{R}_k^2$ выполнены при некотором $k \geq 0$. В силу $\mathcal{R}_k^2$ при любом $x \in E$ имеем

$$\begin{aligned} \psi_{k+1}(x) &\leq A_k\phi(x) + \tfrac{1}{2}\|x - x_0\|^2 + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle + \Psi(x)] \\ &\leq (A_k + a_{k+1})\phi(x) + \tfrac{1}{2}\|x - x_0\|^2, \end{aligned}$$

и это есть $\mathcal{R}_{k+1}^2$. Покажем, что соотношение $\mathcal{R}_{k+1}^1$ также выполнено.

Действительно, в силу соотношения (2.3.46) функция $\psi_k(x)$ является сильно выпуклой с параметром $1 + \mu A_k$. Поэтому силу $\mathcal{R}_k^1$ при любом $x \in E$ имеем

$$\begin{aligned} \psi_k(x) &\geq \psi_k^* + \tfrac{1+\mu A_k}{2}\|x - v_k\|^2 \\ &\geq A_k\phi(x_k) + \tfrac{1+\mu A_k}{2}\|x - v_k\|^2. \end{aligned} \tag{2.3.57}$$

Таким образом,

$$\begin{aligned} \psi_{k+1}^* = \min_{x \in E} \quad & \{\psi_k(x) + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle + \Psi(x)]\} \\ \overset{(2.3.57)}{\geq} \min_{x \in E} \quad & \left\{A_k\phi(x_k) + \tfrac{1+\mu A_k}{2}\|x - v_k\|^2 + a_{k+1}[\phi(x_{k+1}) + \langle \phi'(x_{k+1}), x - x_{k+1}\rangle]\right\} \\ \geq \min_{x \in E} \quad & \{(A_k + a_{k+1})\phi(x_{k+1}) + A_k\langle \phi'(x_{k+1}), x_k - x_{k+1}\rangle \\ & + a_{k+1}\langle \phi'(x_{k+1}), x - x_{k+1}\rangle + \tfrac{1+\mu A_k}{2}\|x - v_k\|^2\} \\ \overset{(2.3.52)}{=} \min_{x \in E} \quad & \{A_{k+1}\phi(x_{k+1}) + \langle \phi'(x_{k+1}), A_{k+1}y_k - a_{k+1}v_k - A_kx_{k+1}\rangle \\ & + a_{k+1}\langle \phi'(x_{k+1}), x - x_{k+1}\rangle + \tfrac{1+\mu A_k}{2}\|x - v_k\|^2\} \\ = \min_{x \in E} \quad & \{A_{k+1}\phi(x_{k+1}) + A_{k+1}\langle \phi'(x_{k+1}), y_k - x_{k+1}\rangle \\ & + a_{k+1}\langle \phi'(x_{k+1}), x - v_k\rangle + \tfrac{1+\mu A_k}{2}\|x - v_k\|^2\}. \end{aligned}$$

Минимум последней задачи достигается при $x = v_k - \frac{a_{k+1}}{1+\mu A_k}B^{-1}\phi'(x_{k+1})$. Таким образом, мы доказали неравенство

$$\psi_{k+1}^* \quad \geq \quad A_{k+1}\phi(x_{k+1}) + A_{k+1}\langle \phi'(x_{k+1}), y_k - x_{k+1}\rangle - \tfrac{a_{k+1}^2}{2(1+\mu A_k)}\|\phi'(x_{k+1})\|_*^2.$$

С другой стороны, в силу критерия остановки внутреннего цикла в методе (2.3.52), имеем

$$\langle \phi'(x_{k+1}), y_k - x_{k+1} \rangle \;\; \geq \;\; \tfrac{1}{M_k} \| \phi'(x_{k+1}) \|_*^2 .$$

Осталось заметить, что в методе 2.3.52) число a_{k+1} удовлетворяет квадратному уравнению

$$A_{k+1} \;\; \equiv \;\; A_k + a_{k+1} \;\; = \;\; \tfrac{M_k a_{k+1}^2}{2(1+\mu A_k)} .$$

Таким образом, соотношение $\mathcal{R}_{k+1}^1$ выполнено. □

Итак, для получения из неравенства (2.3.48) оценки скорости сходимости метода $\mathcal{A}(x_0, L_0, \mu)$ нам осталось оценить скорость роста масштабных коэффициентов $\{A_k\}_{k=0}^{\infty}$.

Лемма 2.3.8 *При любом* $\mu \geq 0$ *масштабные коэффициенты растут следующим образом:*

$$A_k \;\; \geq \;\; \tfrac{k^2}{2\gamma_u L_f}, \quad k \geq 0. \qquad (2.3.58)$$

Если $\mu > 0$*, то скорость роста линейна:*

$$A_k \;\; \geq \;\; \tfrac{2}{\gamma_u L_f} \cdot \left[1 + \sqrt{\tfrac{\mu}{2\gamma_u L_f}} \right]^{2(k-1)}, \quad k \geq 1. \qquad (2.3.59)$$

Доказательство.

Действительно, в силу уравнения $(*)$ в методе (2.3.52) имеем

$$\begin{array}{rcl}
A_{k+1} & \leq & A_{k+1}(1 + \mu A_k) \\[2ex]
 & = & \frac{M_k}{2}(A_{k+1} - A_k)^2 \\[2ex]
 & = & \frac{M_k}{2} \left[A_{k+1}^{1/2} - A_k^{1/2} \right]^2 \left[A_{k+1}^{1/2} + A_k^{1/2} \right]^2 \\[2ex]
 & \leq & 2A_{k+1} M_k \left[A_{k+1}^{1/2} - A_k^{1/2} \right]^2 \\[2ex]
 & \overset{(2.3.54)}{\leq} & 2A_{k+1} \gamma_u L_f \left[A_{k+1}^{1/2} - A_k^{1/2} \right]^2 .
\end{array}$$

Таким образом, для любого $k \geq 0$ получаем $A_k^{1/2} \geq \frac{k}{\sqrt{2\gamma_u L_f}}$. Если $\mu > 0$, то по тем же причинам, что и раньше, получаем

$$\mu A_k A_{k+1} \;\; < \;\; A_{k+1}(1 + \mu A_k) \;\; \leq \;\; 2A_{k+1}\gamma_u L_f \left[A_{k+1}^{1/2} - A_k^{1/2} \right]^2 .$$

Поэтому $A_{k+1}^{1/2} \geq A_k^{1/2} \left[1 + \sqrt{\frac{\mu}{2\gamma_u L_f}} \right]$. Поскольку $A_1 = \frac{2}{M_0} \overset{(2.3.54)}{\geq} \frac{2}{\gamma_u L_f}$, мы получаем оценку (2.3.59). □

Теперь мы можем подвести итог.

Теорема 2.3.6 *Предположим, что градиент функции f удовлетворяет условию Липшица с константой L_f и пусть параметр L_0 удовлетворяет условию (2.3.27). Тогда скорость сходимости метода $\mathcal{A}(x_0, L_0, 0)$, применяемого для решения задачи (2.3.44), оценивается следующим образом:*

$$\phi(x_k) - \phi(x^*) \ \leq \tfrac{\gamma_u L_f \|x^* - x_0\|^2}{k^2}, \quad k \geq 1. \tag{2.3.60}$$

Если к тому же функция Ψ сильно выпукла, то последовательность $\{x_k\}_{k=1}^{\infty}$, генерируемая методом $\mathcal{A}(x_0, L_0, \mu_\Psi)$, сходится в соответствии как с неравенством (2.3.60), так и с неравенством

$$\phi(x_k) - \phi(x^*) \ \leq \tfrac{\gamma_u L_f}{4} \|x^* - x_0\|^2 \cdot \left[1 + \sqrt{\tfrac{\mu_\Psi}{2\gamma_u L_f}}\right]^{-2(k-1)}, \quad k \geq 1. \tag{2.3.61}$$

В следующем пункте мы покажем, как можно использовать полученные результаты при решении различных задач оптимизации.

2.3.5 Примеры применения

Строго выпуклая целевая функция с известным параметром

Рассмотрим следующую выпуклую задачу условной минимизации:

$$\min_{x \in Q} \ \hat{f}(x), \tag{2.3.62}$$

где множество Q выпукло и замкнуто, а сильно выпуклая функция $\hat{f}$ имеет липшицев градиент. Пусть параметр выпуклости $\mu_{\hat{f}}$ нам известен. Обозначим через $\sigma_Q(x)$ индикаторную функцию множества Q:

$$\sigma_Q(x) \ = \ \begin{cases} \ \ \ 0, & x \in Q, \\ +\infty, & \text{иначе}. \end{cases}$$

Задачу (2.3.62) можно решать двумя разными способами.

1. Изменение парметра выпуклости целевой функции. Для $\mu \in (0, \mu_{\hat{f}}]$ положим

$$f(x) \ = \ \hat{f}(x) - \tfrac{\mu}{2}\|x - x_0\|^2, \quad \Psi(x) \ = \ \sigma_Q(x) + \tfrac{\mu}{2}\|x - x_0\|^2. \tag{2.3.63}$$

Заметим, что функция f в формуле (2.3.63) выпукла и ее градиент удовлетворяет условию Липшица с константой $L_f = L_{\hat{f}} - \mu$. Более того, функция $\Psi(x)$ является сильно выпуклой с параметром μ. С другой стороны,

$$\phi(x) \ = \ f(x) + \Psi(x) \ = \ \hat{f}(x) + \sigma_Q(x).$$

Таким образом, соответствующая задача безусловной минимизации (2.3.44) совпадает с задачей условной минимизации (2.3.62). Поскольку все условия теоремы 2.3.6 выполнены, метод $\mathcal{A}(x_0, L_0, \mu)$ имеет следующую скорость сходимости:

$$\hat{f}(x_k) - \hat{f}(x^*) \;\; \leq \frac{\gamma_u(L_{\hat{f}}-\mu)\|x^*-x_0\|^2}{2\left[1+\sqrt{\frac{\mu}{2\gamma_u(L_{\hat{f}}-\mu)}}\right]^{2(k-1)}}, \quad k \geq 1. \tag{2.3.64}$$

Это означает, что ϵ-решение задачи (2.3.62) может быть получено за

$$O\left(\sqrt{\frac{L_{\hat{f}}}{\mu}} \cdot \ln \frac{1}{\epsilon}\right) \tag{2.3.65}$$

итераций. Заметим, что та же задача может быть решена и градиентным методом (2.3.28). Однако в соответствии с неравенством (2.3.40) его производительность будет гораздо хуже. Потребуется

$$O\left(\frac{L_{\hat{f}}}{\mu} \cdot \ln \frac{1}{\epsilon}\right)$$

итераций.

2. Рестарт. Для задачи (2.3.62) зададим следующее представление в виде составной функции (см. задачу (2.3.44)):

$$f(x) \;=\; \hat{f}(x), \quad \Psi(x) \;=\; \sigma_Q(x). \tag{2.3.66}$$

Зафиксируем некоторое число $N \geq 1$ в качестве верхней границы на число итераций в методе $\mathcal{A}$. Рассмотрим следующий двухуровневый процесс.

$$\begin{array}{l} \text{Выбираем } u_0 \in Q. \\ \\ \text{Вычисляем } u_{k+1} \text{ как результат } N \text{ итераций метода } \mathcal{A}(u_k, L_0, 0),\ k \geq 0. \end{array} \tag{2.3.67}$$

В силу определения (2.3.66) имеем

$$\hat{f}(u_{k+1}) - \hat{f}(x^*) \;\overset{(2.3.60)}{\leq}\; \frac{\gamma_u L_{\hat{f}}\|x^*-u_k\|^2}{N^2} \;\leq\; \frac{2\gamma_u L_{\hat{f}}[\hat{f}(u_k)-\hat{f}(x^*)]}{\mu_{\hat{f}} \cdot N^2}.$$

Таким образом, выбирая $N = 2\sqrt{\frac{\gamma_u L_{\hat{f}}}{\mu_{\hat{f}}}}$, получаем

$$\hat{f}(u_{k+1}) - \hat{f}(x^*) \;\leq\; \tfrac{1}{2}[\hat{f}(u_k) - \hat{f}(x^*)].$$

Поэтому оценки сложности для этого подхода такие же, как и в формуле (2.3.65).

Приближенное условие оптимальности первого порядка

В некоторых приложениях нам необходимо найти точку с маленькой невязкой в системе уравнений, отвечающих за условия оптимальности. Поскольку

$$\begin{array}{rcl} D\phi(T_L(x))[u] & \overset{(2.3.19)}{\geq} & -\left(1 + \frac{L_f}{L}\right) \cdot \|g_L(x)\|_* \\ \\ & \overset{(2.3.16)}{\geq} & -(L + L_f) \cdot \sqrt{\frac{\phi(x)-\phi(x^*)}{2L-L_f}} \\ \\ & & \forall u \in \mathcal{F}(T_L(x)),\ \|u\| = 1, \end{array} \tag{2.3.68}$$

верхняя оценка на эту невязку может быть получена из оценок скорости сходимости метода (2.3.52) в форме (2.3.60) или (2.3.61). Однако в этом случае первое неравенство не дает хорошего результата. Действительно, оно может гарантировать, что правая часть неравенства (2.3.68) стремится к нулю со скоростью $O(\frac{1}{k})$. Эта скорость типична для градиентного метода (см. неравенство (2.3.37)), но от быстрого градиентного метода (2.3.52) мы можем ожидать гораздо большего. Покажем, как можно получить лучший результат.

Рассмотрим следующую задачу условной минимизации:

$$\min_{x\in Q} \; f(x), \tag{2.3.69}$$

где множество Q выпукло и замкнуто, а выпуклая функция f имеет липшицев градиент. Зафиксируем параметр точности $\delta > 0$ и стартовую точку $x_0 \in Q$. Положим

$$\Psi(x) \;\; = \;\; \sigma_Q(x) + \tfrac{\delta}{2}\|x - x_0\|^2.$$

Рассмотрим теперь задачу безусловной минимизации (2.3.44) с составной целевой функцией $\phi(x) = f(x) + \Psi(x)$. Заметим, что функция Ψ являтся сильно выпуклой с параметром $\mu_\Psi = \delta$. Поэтому в силу теоремы 2.3.6 метод $\mathcal{A}(x_0, L_0, \delta)$ сходится следующим образом:

$$\phi(x_k) - \phi(x^*) \;\; \le \;\; \tfrac{\gamma_u L_f}{4}\|x^* - x_0\|^2 \cdot \left[1 + \sqrt{\tfrac{\delta}{2\gamma_u L_f}}\right]^{-2(k-1)}. \tag{2.3.70}$$

Для простоты выберем $\gamma_u = \gamma_d$, чтобы выполнялось условие $L_k \le L_f$ при всех $k \ge 0$.

Вычислим теперь $T_k = \mathcal{G}(x_k, L_k).T$ и $M_k = \mathcal{G}(x_k, L_k).L$. Тогда

$$\phi(x_k) - \phi(x^*) \;\ge\; \phi(x_k) - \phi(T_k) \;\; \overset{(2.3.20)}{\ge} \;\; \tfrac{1}{2M_k}\|g_{M_k}(x_k)\|_*^2, \quad L_0 \;\le\; M_k \;\le\; \gamma_u L_f,$$

и мы получаем следующую оценку:

$$\|g_{M_k}(x_k)\|_* \;\; \overset{(2.3.70)}{\le} \;\; \tfrac{\gamma_u L_f}{2^{1/2}}\|x^* - x_0\| \cdot \left[1 + \sqrt{\tfrac{\delta}{2\gamma_u L_f}}\right]^{1-k}. \tag{2.3.71}$$

В нашем случае условия оптимальности первого порядка (2.3.45) для вычисления $T_{M_k}(x_k)$ записываются следующим образом:

$$\nabla f(x_k) + \delta B(T_k - x_0) + \xi_k \;\; = \;\; g_{M_k}(x_k), \tag{2.3.72}$$

где $\xi_k \in \partial\sigma_Q(T_k)$. Заметим, что при любом $y \in Q$ выполняются соотношения

$$0 \;\; = \;\; \sigma_Q(y) \; \ge \sigma_Q(T_k) + \langle \xi_k, y - T_k\rangle \;\; = \;\; \langle \xi_k, y - T_k\rangle. \tag{2.3.73}$$

Поэтому для любого направления $u \in \mathcal{F}(T_k)$, $\|u\| = 1$, получаем

$$\begin{aligned}
\langle \nabla f(T_k), u\rangle \;\; & \overset{(2.3.14)}{\ge} \;\; \langle \nabla f(x_k), u\rangle - \tfrac{L_f}{M_k}\|g_{M_k}(x_k)\|_* \\
& \overset{(2.3.72)}{=} \;\; \langle g_{M_k}(x_k) - \delta B(T_k - x_0) - \xi_k, u\rangle - \tfrac{L_f}{M_k}\|g_{M_k}(x_k)\|_* \\
& \overset{(2.3.73)}{=} \;\; -\delta \cdot \|T_k - x_0\| - \left(1 + \tfrac{L_f}{M_k}\right) \cdot \|g_{M_k}(x_k)\|_*.
\end{aligned}$$

Предположим теперь, что размер множества Q не превосходит R и $\delta = \epsilon \cdot L_0$. Определим число итераций k из неравенства

$$\left[1 + \sqrt{\tfrac{\epsilon L_0}{2\gamma_u L_f}}\right]^{1-k} \;\leq\; \epsilon.$$

Тогда невязка в системе условий оптимальности первого порядка удовлетворяет неравенству

$$\langle \nabla f(T_k), u \rangle \;\geq\; -\epsilon \cdot R \cdot \left[L_0 + \tfrac{\gamma_u L_f}{2^{1/2}} \cdot \left(1 + \tfrac{L_f}{L_0}\right)\right], \quad u \in \mathcal{F}(T_k),\ \|u\| = 1. \tag{2.3.74}$$

Таким образом, требуемое число итераций k будет порядка $O\left(\frac{1}{\sqrt{\epsilon}} \ln \frac{1}{\epsilon}\right)$.

Неизвестный параметр сильной выпуклости

В п. 2.3.5 мы обсудили две эффективные стратегии минимизации сильно выпуклой функции с известным параметром выпуклости $\mu_{\hat{f}}$. Однако очень часто эта информация недоступна. Мы можем легко получить только верхнюю оценку для этой величины, например, с помощью неравенства

$$\mu_{\hat{f}} \;\leq\; S_L(x) \;\leq\; L_{\hat{f}}, \quad x \in Q.$$

Покажем, что эта информация может быть использована для эффективной минимизации сильно выпуклой функции.

Предположим, что для задачи (2.3.62) у нас имеется некоторая оценка μ для параметра $\mu_{\hat{f}}$. Зафиксируем стартовую точку $u_0 \in Q$. Обозначим $\phi_0(x) = \hat{f}(x) + \sigma_Q(x)$. Выберем

$$x_0 = \mathcal{G}_{\phi_0}(u_0, L_0).T, \quad M_0 = \mathcal{G}_{\phi_0}(u_0, L_0).L, \quad S_0 = \mathcal{G}_{\phi_0}(u_0, L_0).S$$

и минимизируем составную функцию (2.3.63) с помощью метода $\mathcal{A}(x_0, M_0, \mu)$, снабженного следующим критерием остановки.

$$\boxed{\begin{array}{lll} \text{ВЫЧИСЛЯЕМ:} & v_k = \mathcal{G}_{\phi_0}(x_k, L_k).T, \quad M_k = \mathcal{G}_{\phi_0}(x_k, L_k).L. & \\ & & \\ \text{ОСТАНОВКА ЭТАПА:} & \text{Если (A):} & \|\, g_{M_k}(x_k)[\phi_0] \,\|_* \leq \frac{1}{2} \|\, g_{M_0}(u_0)[\phi_0] \,\|_*, \\ & & \\ & \text{иначе (B):} & \frac{M_k}{A_k} \cdot \left(1 + \frac{S_0}{M_0}\right) \leq \frac{1}{4}\mu^2. \end{array}} \tag{2.3.75}$$

Если этап был прерван по условию (A), то мы называем его *успешным.* В этом случае мы запускаем следующий этап, выбирая v_k как новую стартовую точку и оставляя оценку параметра выпуклости μ без изменения.

Предположим, что этап был прерван по условию (B) (это *неудачный* этап). Если бы μ было правильной *нижней* оценкой для параметра выпуклости $\mu_{\hat{f}}$, то выполнялись бы неравенства

$$\begin{aligned} \tfrac{1}{2M_k}\|g_{M_k}(x_k)[\phi_0]\|_*^2 &\overset{(2.3.20)}{\leq} \hat{f}(x_k) - \hat{f}(x^*) \\ &\overset{(2.3.48)}{\leq} \tfrac{1}{2A_k}\|x_0 - x^*\|^2 \\ &\overset{(2.3.25)}{\leq} \tfrac{1}{2A_k\mu^2} \cdot \left(1 + \tfrac{S_0}{M_0}\right) \cdot \|g_{M_0}(u_0)[\phi_0]\|_*^2. \end{aligned}$$

Поэтому в силу условия (B) в этом случае этап должен был бы прерваться по условию (A). Так как этого не случилось, мы заключаем что $\mu > \mu_{\hat{f}}$. Таким образом, мы переопределяем $\mu := \frac{1}{2}\mu$, и запускаем этап снова из старой стартовой точки x_0.

Мы опускаем подробный анализ эффективности этой стратегии. Можно показать, что для генерации ϵ-решения задачи (2.3.62) сильно выпуклой целевой функцией потребуется

$$O\left(\kappa_{\hat{f}}^{1/2} \ln \kappa_{\hat{f}}\right) + O\left(\kappa_{\hat{f}}^{1/2} \ln \kappa_{\hat{f}} \cdot \ln \tfrac{\kappa_{\hat{f}}}{\epsilon}\right), \quad \kappa_{\hat{f}} \overset{\text{def}}{=} \tfrac{L_{\hat{f}}}{\mu_{\hat{f}}},$$

вызовов оракула. Первый член в этой оценке соответствует общему числу вызовов оракула на всех неудачных этапах. Множитель $\kappa_{\hat{f}}^{1/2} \ln \kappa_{\hat{f}}$ дает верхнюю оценку на длину этапа независимо от способа его прерывания.

2.3.6 Вычислительные эксперименты

Методы, описанные в этом параграфе, численно проверялись на случайно сгенерированных задачах нахождения разреженного решения в наименьших квадратах, поставленных в форме

$$\text{найти } \phi^* = \min_{x \in \mathbb{R}^n} \left[\phi(x) \overset{\text{def}}{=} \tfrac{1}{2}\|Ax - b\|_2^2 + \|x\|_1 \right], \tag{2.3.76}$$

где $A \equiv (a_1, \ldots, a_n)$ - это $m \times n$-заполненная матрица, $m < n$. Все задачи генерировались с заранее известным оптимальным решением, которое можно получить из двойственного представления исходной *прямой* задачи (2.3.76):

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \left[\tfrac{1}{2}\|Ax - b\|_2^2 + \|x\|_1 \right] &= \min_{x \in \mathbb{R}^n} \max_{u \in \mathbb{R}^m} \left[\langle u, b - Ax \rangle - \tfrac{1}{2}\|u\|_2^2 + \|x\|_1 \right] \\ &= \max_{u \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} \left[\langle b, u \rangle - \tfrac{1}{2}\|u\|_2^2 - \langle A^T u, x \rangle + \|x\|_1 \right] \qquad (2.3.77) \\ &= \max_{u \in \mathbb{R}^m} \left[\langle b, u \rangle - \tfrac{1}{2}\|u\|_2^2 : \ \|A^T u\|_\infty \leq 1 \right]. \end{aligned}$$

Таким образом, задача, двойственная к (2.3.76), состоит в нахождении евклидовой проекции вектора $b \in \mathbb{R}^m$ на двойственный многогранник

$$\mathcal{D} \ = \ \{y \in \mathbb{R}^m : \ \|A^T y\|_\infty \leq 1\}.$$

Эта интерпретация объясняет изменяющуюся разреженность оптимального решения $x^*(\tau)$ следующей параметрической версии задачи (2.3.76):

$$\min_{x\in\mathbb{R}^n} \quad \Big[\; \phi_\tau(x) \overset{\mathrm{def}}{=} \tfrac{1}{2}\|Ax-b\|_2^2 + \tau\|x\|_1 \; \Big] \tag{2.3.78}$$

Действительно, при $\tau > 0$, имеем

$$\phi_\tau(x) \quad = \quad \tau^2 \left[\tfrac{1}{2}\|A\tfrac{x}{\tau} - \tfrac{b}{\tau}\|_2^2 + \|\tfrac{x}{\tau}\|_1 \right].$$

Поэтому в двойственной задаче мы проектируем вектор $\frac{b}{\tau}$ на многогранник $\mathcal{D}$. Ненулевые компоненты вектора $x^*(\tau)$ соответствуют активным граням многогранника $\mathcal{D}$. Таким образом, для достаточно большого τ мы получим $\frac{b}{\tau} \in \operatorname{int}\mathcal{D}$, а это означает, что $x^*(\tau) = 0$. Когда τ уменьшается, мы получаем все больше и больше ненулевых координат вектора $x^*(\tau)$. Наконец, если все грани многогранника $\mathcal{D}$ находятся в общем положении, то при $\tau \to 0$ мы получаем в $x^*(\tau)$ ровно m ненулевых компонент.

В наших вычислительных экспериментах мы сравниваем три метода минимизации. Два из них рекурсивно поддерживают соотношения (2.3.47). Это позволяет называть их прямо-двойственными алгоритмами. Действительно, обозначим

$$\phi_*(u) \quad = \quad \tfrac{1}{2}\|u\|_2^2 - \langle b, u\rangle.$$

Как мы видели в формуле (2.3.77),

$$\phi(x) + \phi_*(u) \quad \geq \quad 0 \quad \forall x \in \mathbb{R}^n, \; u \in \mathcal{D}. \tag{2.3.79}$$

Более того, эта нижняя оценка достигается только в оптимальных решениях прямой и двойственной задач. Для некоторой последовательности $\{z_i\}_{i=1}^\infty$ и стартовой точки $z_0 \in \operatorname{dom}\Psi$ соотношения (2.3.47) обеспечивают выполнение неравенства

$$A_k\phi(x_k) \quad \leq \quad \min_{x\in\mathbb{R}^n} \left\{ \sum_{i=1}^k a_i[f(z_i) + \langle \nabla f(z_i), x - z_i\rangle] + A_k\Psi(x) + \tfrac{1}{2}\|x - z_0\|^2 \right\}. \tag{2.3.80}$$

В нашей ситуации $f(x) = \frac{1}{2}\|Ax-b\|_2^2$, $\Psi(x) = \|x\|_1$, и мы выбираем $z_0 = 0$. Обозначим $u_i = b - Az_i$. Тогда $\nabla f(z_i) = -A^T u_i$ и, таким образом,

$$\begin{aligned} f(z_i) - \langle \nabla f(z_i), z_i\rangle \quad &= \quad \tfrac{1}{2}\|u_i\|^2 + \langle A^T u_i, z_i\rangle \\ &= \quad \langle b, u_i\rangle - \tfrac{1}{2}\|u_i\|_2^2 \\ &= \quad -\phi_*(u_i). \end{aligned}$$

Обозначая

$$\bar{u}_k \quad = \quad \tfrac{1}{A_k}\sum_{i=1}^k a_i u_i, \tag{2.3.81}$$

мы получаем

$$\begin{aligned} A_k[\phi(x_k) + \phi_*(\bar{u}_k)] \quad &\leq \quad A_k\phi(x_k) + \sum_{i=1}^{k} a_i\phi_*(u_i) \\ &\overset{(2.3.80)}{\leq} \quad \min_{x\in\mathbb{R}^n} \left\{ \sum_{i=1}^{k} a_i\langle\nabla f(z_i), x\rangle + A_k\Psi(x) + \tfrac{1}{2}\|x\|^2 \right\} \overset{x=0}{\leq} 0. \end{aligned}$$

В силу неравенства (2.3.79) точка u_k не может быть допустимой:

$$\phi_*(\bar{u}_k) \;\leq\; -\phi(x_k) \;\leq\; -\phi^* \;=\; \min_{u\in\mathcal{D}} \phi_*(u). \tag{2.3.82}$$

Давайте измерим уровень недопустимости этих точек. Заметим, что минимум оптимизационной задачи в неравенстве (2.3.80) достигается при $x = v_k$. Поэтому, соответствующие условия оптимальности первого порядка обеспечивают оценку

$$\| - \sum_{i=1}^{k} a_i A^T u_i + Bv_k\|_\infty \;\leq\; A_k.$$

Таким образом, $|\langle a_i, \bar{u}_k\rangle| \leq 1 + \frac{1}{A_k}|(Bv_k)^{(i)}|$, $i = 1,\ldots,n$. Предположим что матрица B в формулах (2.3.3) диагональная:

$$B^{(i,j)} = \begin{cases} d_i, & i = j, \\ 0, & \text{otherwise}. \end{cases}$$

Тогда $|\langle a_i, \bar{u}_k\rangle| - 1 \leq \frac{d_i}{A_k} \cdot |v_k^{(i)}|$ и

$$\rho(\bar{u}_k) \;\overset{\text{def}}{=}\; \left[\sum_{i=1}^{n} \tfrac{1}{d_i} \cdot (\; |\langle a_i, \bar{u}_k\rangle| - 1\;)_+^2 \right]^{1/2} \;\leq\; \tfrac{1}{A_k}\|v_k\| \;\overset{(2.3.49)}{\leq}\; \tfrac{2}{A_k}\|x^*\|, \tag{2.3.83}$$

где $(\alpha)_+ = \max\{\alpha, 0\}$. Таким образом, мы можем использовать функцию $\rho(\cdot)$ как меру двойственной недопустимости. Ввиду оценки (2.3.82) она может использоваться как естественный критерий остановки прямо-двойственных методов.

При генерации случайных тестовых задач мы пользовались следующей стратегией.

- Выбираем $m_* \leq m$ число ненулевых компонент в оптимальном решении x^* задачи (2.3.76) и параметр $\rho > 0$, ответственный за размер вектора x^*.
- Случайно генерируем матрицу $B \in \mathbb{R}^{m\times n}$, элементы которой равномерно распределены на отрезке $[-1, 1]$.
- Случайно генерируем вектор $v^* \in \mathbb{R}^m$ с элементами, равномерно распределенными в $[0, 1]$. Полагаем $y^* = v^*/\|v^*\|_2$.
- Сортируем вектор $B^T y^*$ в порядке убывания модулей его элементов. Для упрощения обозначений предположим что этот порядок совпадает с естественным.

- Для $i = 1, \dots, n$ полагаем $a_i = \alpha_i b_i$, где $\alpha_i > 0$ выбраны в соответствии со следующим правилом:

$$\alpha_i \;=\; \begin{cases} \frac{1}{|\langle b_i, y^* \rangle|} & \text{для } i = 1, \dots, m_*, \\ \\ 1, & \text{если } |\langle b_i, y^* \rangle| \leq 0,1 \text{ и } i > m_*, \\ \\ \frac{\xi_i}{|\langle b_i, y^* \rangle|} & \text{иначе,} \end{cases}$$

  причем ξ_i равномерно распределены в $[0, 1]$.

- Для $i = 1, \dots, n$ генерируем компоненты прямого решения:

$$[x^*]^{(i)} = \begin{cases} \xi_i \cdot \operatorname{sign}(\langle a_i, y^* \rangle) & \text{для } i \leq m_*, \\ \\ 0 & \text{иначе,} \end{cases}$$

  где ξ_i равномерно распределены в $\left[0, \frac{\rho}{\sqrt{m_*}}\right]$.

- Полагаем $b = y^* + Ax^*$.

Таким образом, оптимальное значение случайно сгенерированной задачи (2.3.76) вычисляется по формуле

$$\phi^* \;=\; \tfrac{1}{2}\|y^*\|_2^2 + \|x^*\|_1.$$

В первой серии тестов мы использовали это значение в критерии остановки.

Приведем результаты расчетов для двух типичных случайных задач. Первая задача является относительно простой.

Задача 1: $n = 4000,\ m = 1000,\ m_* = 100,\ \rho = 1.$

	PG			DG			AC		
Зазор	к	#Ax	Ускор.	к	#Ax	Ускор.	к	#Ax	Ускор.
1	1	4	0.21%	1	4	0.85%	1	4	0.14%
2^{-1}	3	8	0.20%	3	12	0.81%	4	28	1.24%
2^{-2}	10	29	0.24%	8	38	0.89%	8	60	2.47%
2^{-3}	28	83	0.32%	25	123	1.17%	14	108	4.06%
2^{-4}	159	476	0.88%	156	777	3.45%	40	316	17.50%
2^{-5}	557	1670	1.53%	565	2824	6.21%	74	588	29.47%
2^{-6}	954	2862	1.31%	941	4702	5.17%	98	780	25.79%
2^{-7}	1255	3765	0.86%	1257	6282	3.45%	118	940	18.62%
2^{-8}	1430	4291	0.49%	1466	7328	2.01%	138	1096	12.73%
2^{-9}	1547	4641	0.26%	1613	8080	2.13%	156	1240	8.19%
2^{-10}	1640	4920	0.14%	1743	8713	0.61%	173	1380	4.97%
2^{-11}	1722	5167	0.07%	1849	9243	0.33%	188	1500	3.01%
2^{-12}	1788	5364	0.04%	1935	9672	0.17%	202	1608	1.67%
2^{-13}	1847	5539	0.02%	2003	10013	0.09%	216	1720	0.96%
2^{-14}	1898	5693	0.01%	2061	10303	0.05%	230	1836	0.55%
2^{-15}	1944	5831	0.01%	2113	10563	0.05%	248	1968	0.31%
2^{-16}	1987	5961	0.00%	2164	10817	0.03%	265	2112	0.19%
2^{-17}	2029	6085	0.00%	2217	11083	0.02%	279	2224	0.10%
2^{-18}	2072	6215	0.00%	2272	11357	0.01%	305	2432	0.06%
2^{-19}	2120	6359	0.00%	2331	11652	0.00%	314	2504	0.03%
2^{-20}	2165	6495	0.00%	2448	12238	0.00%	319	2544	0.02%

В этой таблице колонка $\boxed{\text{Зазор}}$ показывает убывание начальной невязки. В оставшейся части таблицы мы видим результаты для трех методов:

- прямой градиентный метод (2.3.28), обозначается PG;
- двойственная версия градиентного метода (2.3.50), обозначается DG;
- быстрый градиентный метод (2.3.52), обозначается AC.

Во всех методах мы используем следующие значения параметров:

$$\gamma_u = \gamma_d = 2, \quad x_0 = 0, \quad L_0 = \max_{1 \le i \le n} \|a_i\|^2 \ \le\ L_f, \quad \mu = 0.$$

Объясним смысл оставшихся колонок в таблице. Для каждого метода колонка $\boxed{\text{к}}$ показывает число итераций, необходимых для достижения соответствующего уменьшения начальной невязки по значению функции. Колонка $\boxed{\text{Ax}}$ показывает необходимое число матрично-векторных умножений. Напомним, что для вычисления значения $f(x)$ требуется одно умножение. Если при этом надо вычислить градиент, то потребуется еще одно. Так, например, в соответствии с оценкой (2.3.56) каждая итерация метода (2.3.52) требует четырех вычислений значений для пар функция/градиент. Поэтому в данном методе мы можем ожидать восемь матрично-векторных умножений на каждой итерации. Для градиентного метода (2.3.28) в среднем на итерации требуется два вызова оракула. Однако один из них нужен в процедуре линейного поиска (2.3.26). Следовательно в нем потребуется только одно значение функции. Поэтому в данном случае нам в среднем потребуется три матрично-векторных умножений на итерацию. Интересно, что наши предсказания выполняются с хорошей точностью.

Наконец, в колонке $\boxed{\text{Ускор.}}$ представлено отношение числа итераций, потребовавшегося для достижения данной точности, к теоретической оценке (в процентах).

Мы видим, что все методы сходятся существенно лучше, чем предсказано теорией. Однако для всех методов имеется часть траектории, где теоретические предсказания довольно точны. Эта особенность еще больше заметна в нашей второй таблице, где решается более трудная задача.

Задача 2: $n = 5000,\ m = 500,\ m_* = 100,\ \rho = 1.$

	PG			DG			AC		
Зазор	к	#Ах	Ускор.	к	#Ах	Ускор.	к	#Ах	Ускор.
1	1	4	0.24%	1	4	0.96%	1	4	0.16%
2^{-1}	2	6	0.20%	2	8	0.81%	3	24	0.92%
2^{-2}	5	17	0.21%	5	24	0.81%	5	40	1.49%
2^{-3}	11	33	0.19%	11	45	0.77%	8	64	1.83%
2^{-4}	38	113	0.30%	38	190	1.21%	19	148	5.45%
2^{-5}	234	703	0.91%	238	1189	3.69%	52	416	20.67%
2^{-6}	1027	3081	1.98%	1026	5128	7.89%	106	848	43.08%
2^{-7}	2402	7206	2.31%	2387	11933	9.17%	160	1280	48.70%
2^{-8}	3681	11043	1.77%	3664	18318	7.05%	204	1628	39.54%
2^{-9}	4677	14030	1.12%	4664	23318	4.49%	245	1956	28.60%
2^{-10}	5410	16230	0.65%	5392	26958	2.61%	288	2300	19.89%
2^{-11}	5938	17815	0.36%	5879	29393	1.41%	330	2636	13.06%
2^{-12}	6335	19006	0.19%	6218	31088	0.77%	370	2956	8.20%
2^{-13}	6637	19911	0.10%	6471	32353	0.41%	402	3212	4.77%
2^{-14}	6859	20577	0.05%	6670	33348	0.21%	429	3424	2.71%
2^{-15}	7021	21062	0.03%	6835	34173	0.13%	453	3616	1.49%
2^{-16}	7161	21483	0.01%	6978	34888	0.05%	471	3764	0.83%
2^{-17}	7281	21842	0.01%	7108	35539	0.05%	485	3872	0.42%
2^{-18}	7372	22115	0.00%	7225	36123	0.03%	509	4068	0.24%
2^{-19}	7438	22313	0.00%	7335	36673	0.02%	525	4192	0.12%
2^{-20}	7492	22474	0.00%	7433	37163	0.01%	547	4372	0.07%

В этой таблице прямой градиентный метод сходится значительно быстрее теоретических предсказаний. Это не очень удивляет, так как данная схема может автоматически ускоряться при благоприятных обстоятельствах. Это происходит, когда, скажем, функция сильно выпукла (см. теорему 2.3.5). Все другие схемы для таких ускорений требует явного вмешательства.

Однако, несмотря на все эти отклонения, основные выводы нашего теоретического анализа подтверждаются: быстрый метод (2.3.52) существенно превосходит как прямой, так и двойственный градиентные методы.

Во второй серии тестов мы изучали возможности прямо-двойственных методов (2.3.50) и (2.3.52) по уменьшению двойственной меры недопустимости $\rho(\cdot)$ (см. формулу (2.3.83)). Эта задача, по крайней мере для двойственного градиентного метода (2.3.50), оказывается гораздо труднее, чем решение прямой задачи минимизации (2.3.76). Посмотрим на следующие результаты.

Задача 3: $n = 500,\ m = 50,\ m_* = 25,\ \rho = 1.$

	DG				AC			
Зазор	к	#Ax	$\Delta\phi$	Ускор.	к	#Ax	$\Delta\phi$	Ускор.
1	2	8	$2.5 \cdot 10^{0}$	8.26%	2	16	$3.6 \cdot 10^{0}$	2.80%
2^{-1}	5	25	$1.4 \cdot 10^{0}$	9.35%	7	56	$8.8 \cdot 10^{-1}$	15.55%
2^{-2}	13	64	$6.0 \cdot 10^{-1}$	13.17%	11	88	$5.3 \cdot 10^{-1}$	20.96%
2^{-3}	26	130	$3.9 \cdot 10^{-1}$	12.69%	15	120	$4.4 \cdot 10^{-1}$	19.59%
2^{-4}	48	239	$2.7 \cdot 10^{-1}$	12.32%	21	164	$3.1 \cdot 10^{-1}$	19.21%
2^{-5}	103	514	$1.6 \cdot 10^{-1}$	13.28%	35	276	$1.8 \cdot 10^{-1}$	25.83%
2^{-6}	243	1212	$8.3 \cdot 10^{-2}$	15.64%	54	432	$1.0 \cdot 10^{-1}$	31.75%
2^{-7}	804	4019	$3.0 \cdot 10^{-2}$	25.93%	86	688	$4.6 \cdot 10^{-2}$	39.89%
2^{-8}	1637	8183	$6.3 \cdot 10^{-3}$	26.41%	122	976	$1.8 \cdot 10^{-2}$	40.22%
2^{-9}	3298	16488	$4.6 \cdot 10^{-4}$	26.6%	169	1348	$5.3 \cdot 10^{-3}$	38.58%
2^{-10}	4837	24176	$1.8 \cdot 10^{-7}$	19.33%	224	1788	$7.7 \cdot 10^{-4}$	34.28%
2^{-11}	4942	24702	$1.2 \cdot 10^{-14}$	9.97%	301	2404	$8.0 \cdot 10^{-5}$	30.88%
2^{-12}	5149	25734	$-1.3 \cdot 10^{-15}$	5.16%	419	3352	$2.7 \cdot 10^{-5}$	29.95%
2^{-13}	5790	28944	$-1.3 \cdot 10^{-15}$	2.92%	584	4668	$5.3 \cdot 10^{-6}$	29.11%
2^{-14}	6474	32364	0.0	2.67%	649	5188	$4.1 \cdot 10^{-7}$	29.48%

В этой таблице мы видим цену уменьшения начального значения ρ в $2^{14} \approx 10^{4}$ раз. Заметим, что оба метода требуют для этого больше итераций, чем при решении Задачи 1, которая решалась до точности по функциональной невязке порядка $2^{-20} \approx 10^{-6}$. Более того, для достижения требуемого уровня ρ методу (2.3.50) пришлось уменьшать невязку по функции практически до машинной точности, а норму градиента до уровня 10^{-12}. Быстрый градиентный метод дает более сбалансированные результаты: конечная невязка по ϕ остается на уровне 10^{-6}, а норма градиента уменьшается только до $1,3 \cdot 10^{-3}$.

Посмотрим теперь на задачу большего размера.

Задача 4: $n = 1000,\ m = 100,\ m_* = 50,\ \rho = 1.$

	DG				AC			
Зазор	к	#Ax	$\Delta\phi$	Ускор.	к	#Ax	$\Delta\phi$	Ускор.
1	2	8	$3.7 \cdot 10^{0}$	6.41%	2	12	$4.2 \cdot 10^{0}$	1.99%
2^{-1}	5	24	$2.0 \cdot 10^{0}$	7.75%	7	56	$1.4 \cdot 10^{0}$	11.71%
2^{-2}	15	74	$1.0 \cdot 10^{0}$	11.56%	12	96	$8.7 \cdot 10^{-1}$	15.49%
2^{-3}	37	183	$6.9 \cdot 10^{-1}$	14.73%	17	132	$6.8 \cdot 10^{-1}$	16.66%
2^{-4}	83	414	$4.5 \cdot 10^{-1}$	16.49%	26	208	$4.7 \cdot 10^{-1}$	20.43%
2^{-5}	198	989	$2.4 \cdot 10^{-1}$	19.79%	42	336	$2.5 \cdot 10^{-1}$	26.76%
2^{-6}	445	2224	$7.8 \cdot 10^{-2}$	22.28%	65	520	$1.0 \cdot 10^{-1}$	32.41%
2^{-7}	1328	6639	$2.2 \cdot 10^{-2}$	33.25%	91	724	$3.6 \cdot 10^{-2}$	31.50%
2^{-8}	2675	13373	$4.1 \cdot 10^{-3}$	33.48%	125	996	$1.1 \cdot 10^{-2}$	30.07%
2^{-9}	4508	22535	$5.6 \cdot 10^{-5}$	28.22%	176	1404	$2.6 \cdot 10^{-3}$	27.85%
2^{-10}	4702	23503	$2.7 \cdot 10^{-10}$	14.7%	240	1916	$4.4 \cdot 10^{-4}$	26.08%
2^{-11}	4869	24334	$-2.2 \cdot 10^{-15}$	7.61%	328	2620	$7.7 \cdot 10^{-5}$	26.08%
2^{-12}	6236	31175	$-2.2 \cdot 10^{-15}$	4.88%	465	3716	$6.5 \cdot 10^{-6}$	26.20%
2^{-13}	12828	64136	$-2.2 \cdot 10^{-15}$	5.02%	638	5096	$2.4 \cdot 10^{-6}$	24.62%
2^{-14}	16354	81766	$-4.4 \cdot 10^{-15}$	5.24%	704	5628	$7.8 \cdot 10^{-7}$	24.62%

По сравнению с Задачей 3, в Задаче 4 все размеры удвоены. Это практически ничего не меняет для быстрого градиентного метода. Однако вычислительные затраты двойственного градиентного метода существенно возросли. Дальнейшее увеличение размеров делает последнюю схему неработоспособной. Посмотрим как работают оба метода на Задаче 1, в которой значение $\rho(\cdot)$ служит критерием остановки..

Задача 1а: $n = 4000,\ m = 1000,\ m_* = 100,\ \rho = 1.$

	DG				AC			
Зазор	к	#Ах	$\Delta\phi$	Ускор.	к	#Ах	$\Delta\phi$	Ускор.
1	2	8	$2.3 \cdot 10^1$	2.88%	2	12	$2.4 \cdot 10^1$	0.99%
2^{-1}	5	24	$1.2 \cdot 10^1$	3.44%	8	60	$8.1 \cdot 10^0$	7.02%
2^{-2}	17	83	$5.8 \cdot 10^0$	6.00%	13	100	$4.6 \cdot 10^0$	10.12%
2^{-3}	44	219	$3.5 \cdot 10^0$	7.67%	20	160	$3.5 \cdot 10^0$	11.20%
2^{-4}	100	497	$2.7 \cdot 10^0$	8.94%	28	220	$2.9 \cdot 10^0$	12.10%
2^{-5}	234	1168	$1.9 \cdot 10^0$	10.51%	44	348	$2.1 \cdot 10^0$	14.79%
2^{-6}	631	3153	$1.0 \cdot 10^0$	14.18%	78	620	$1.0 \cdot 10^0$	23.46%
2^{-7}	1914	9568	$1.0 \cdot 10^{-2}$	21.50%	117	932	$2.9 \cdot 10^{-1}$	26.44%
2^{-8}	3704	18514	$4.6 \cdot 10^{-7}$	20.77%	157	1252	$6.8 \cdot 10^{-2}$	23.88%
2^{-9}	3731	18678	$1.4 \cdot 10^{-14}$	15.77%	212	1688	$5.3 \cdot 10^{-3}$	21.63%
2^{-10}	1D	поиск	не работает		287	2288	$2.0 \cdot 10^{-4}$	19.87%
2^{-11}					391	3120	$2.5 \cdot 10^{-5}$	18.43%
2^{-12}					522	4168	$7.0 \cdot 10^{-6}$	16.48%
2^{-13}					693	5536	$4.5 \cdot 10^{-7}$	14.40%
2^{-14}					745	5948	$3.8 \cdot 10^{-7}$	13.76%

Причина неудачи двойственного градиентного метода довольно любопытна. В конце он генерирует точки с очень маленькой невязкой по значению функции. Таким образом, критерий остановки в градиентной итерации (2.3.26) уже не может справиться с ошибками округления. В быстрой схеме (2.3.52) этого не происходит, поскольку уменьшение целевой функции и двойственной меры недопустимости идет пропорционально. В некотором смысле такая ситуация естественна. Мы видели, что на данных тестовых задачах все методы ускоряются в конце. С другой стороны, скорость сходимости двойственных переменных $\bar{u}_k$ ограничена скоростью роста коэффициентов a_i в представлении (2.3.81). Для двойственного метода эти коэффициенты практически постоянны. Для быстрого метода они растут пропорционально номеру итерации.

Приведенные численные результаты демонстрируют преимущество быстрого градиентного метода (2.3.52) с адаптивной стратегией выбора шага. Интересно посмотреть, как этот метод работает в других ситуациях.

Заметим, что мы решали задачу (2.3.76) с помощью довольно общей модели (2.3.5), в которой не учитывается тот факт, что функция f *квадратична*. Характеристическим свойством квадратичной функции является постоянство гессиана. Поэтому кажется естественным связать оператор B в метрике (2.3.3) с гессианом функции f.

Положим $B = \mathrm{diag}\,(A^T A) \equiv \mathrm{diag}\,(\nabla^2 f(x))$. Тогда

$$\|e_i\|^2 \;=\; \langle Be_i, e_i\rangle \;=\; \|a_i\|_2^2 \;=\; \|Ae_i\|_2^2 \;=\; \langle \nabla^2 f(x)e_i, e_i\rangle, \quad i = 1, \ldots, n,$$

где e_i – координатный вектор в $\mathbb{R}^n$. Таким образом,

$$L_0 \overset{\mathrm{def}}{=} 1 \;\le\; L_f \;\equiv\; \max_{\|u\|=1} \langle \nabla^2 f(x) u, u \rangle \;\le\; n.$$

Итак, в данной метрике у нас есть очень хорошие верхние и нижние оценки для константы Липшица L_f. Давайте посмотрим на соответствующие вычислительные результаты. Мы решаем Задачу 1 (с $n = 4000$, $m = 1000$, и $\rho = 1$) до точности Зазор $= 2^{-20}$ для разных размеров m_* носителя оптимального вектора, который постепенно увеличивается от 100 до 1000.

Задача 1b.

	PG		AC	
m_*	к	#Ax	к	#Ax
100	42	127	58	472
200	53	160	61	496
300	69	208	70	568
400	95	286	77	624
500	132	397	84	680
600	214	642	108	872
700	330	993	139	1120
800	504	1513	158	1272
900	1149	3447	196	1576
1000	2876	8630	283	2272

Напомним, что первая строка в этой таблице соответствует ранее обсуждавшейся Задаче 1. Для удобства сравнения в следующей таблице мы повторяем конечные результаты для этой задачи, добавив результаты для $m_* = 1000$ (в обоих случаях без диагонального шкалирования).

	PG		AC	
m_*	к	#Ax	к	#Ax
100	2165	6495	319	2544
1000	42509	127528	879	7028

Таким образом, при $m_* = 100$ диагональное шкалирование делает Задачу 1 очень простой. Для простых задач простые и дешевые методы имеют ощутимое преимущество по сравнению с более сложными процедурами. Когда m_* увеличивается, отшкалированные задачи делаются все более и более сложными. Наконец, мы видим снова преимущество быстрого метода.

2.4 Приложение: барьерная проекция на симплекс

В случае $K = \mathbb{R}^n_+$, выберем

$$F(x) \;=\; -\sum\limits_{i=1}^n \ln x^{(i)}, \quad \nu = n.$$

Рассмотрим $\hat{P} = \{x \in \mathbb{R}^n : \; \langle e, x\rangle = 1\}$. Тогда на каждой итерации метода (2.2.32) нам нужно решать следующую задачу:

$$\phi^\star \overset{\mathrm{def}}{=} \max\limits_x \left\{ \langle s, x\rangle + \sum\limits_{i=1}^n \ln x^{(i)} : \; \sum\limits_{i=1}^n x^{(i)} = 1 \right\}. \tag{2.4.1}$$

Покажем, что ее сложность не зависит от размера исходных данных (коэффициентов вектора $s \in \mathbb{R}^n$).

Рассмотрим лагранжиан:

$$\mathcal{L}(x,\lambda) \;=\; \langle s, x\rangle + \sum\limits_{i=1}^n \ln x^{(i)} + \lambda \cdot \left[1 - \sum\limits_{i=1}^n x^{(i)} \right], \quad x \in \mathbb{R}^n, \; \lambda \in R.$$

Двойственная функция

$$\phi(\lambda) = \max\limits_x \left\{ \mathcal{L}(x,\lambda) : \; \sum\limits_{i=1}^n x^{(i)} = 1 \right\} \overset{\mathrm{def}}{=} \mathcal{L}(x(\lambda),\lambda)$$

определяется с помощью вектора $x(\lambda) : \; x^{(i)}(\lambda) = \frac{1}{\lambda - s^{(i)}}$, $i = 1, \dots, n$. Таким образом,

$$\begin{aligned} \phi(\lambda) \;&=\; -n + \lambda - \sum\limits_{i=1}^n \ln\left(\lambda - s^{(i)}\right), \\ \phi_\star \;&=\; \min\limits_\lambda \left\{ \phi(\lambda) : \; \lambda > \max\limits_{1 \le i \le n} s^{(i)} \right\}. \end{aligned} \tag{2.4.2}$$

Заметим, что функция $\phi(\cdot)$ является самосогласованной. Таким образом, для ее минимизации мы можем применять стандартный метод Ньютона, который сходится квадратично для любого λ из области

$$\mathcal{Q}(s) \;=\; \{\lambda : \; 4(\phi'(\lambda))^2 \le \phi''(\lambda)\}.$$

(см. параграф 1.3). Покажем, что трудоемкость нахождения такой точки не зависит от начальной информации.

Рассмотрим функцию $\psi(\lambda) = -\phi'(\lambda) = \sum\limits_{i=1}^n \frac{1}{\lambda - s^{(i)}} - 1$. Ясно, что задача (2.4.2) эквивалентна нахождению наибольшего корня $\lambda_\star$ уравнения

$$\psi(\lambda) \;=\; 0. \tag{2.4.3}$$

Обозначим $\lambda_0 = 1 + \max\limits_{1 \le i \le n} s^{(i)}$. Тогда $\psi(\lambda_0) \ge 0$ и, следовательно, $\lambda_0 \le \lambda_\star$. Рассмотрим следующий процесс:

$$\lambda_{k+1} \;=\; \lambda_k - \frac{\psi(\lambda_k)}{\psi'(\lambda_k)}, \quad k \ge 0. \tag{2.4.4}$$

Это стандартный метод Ньютона для решения уравнения (2.4.3), что эквивалентно решению задачи (2.4.2).

Лемма 2.4.1 *При всех $k \geq 0$ выполнено неравенство* $(\phi'(\lambda_k))^2 \leq n^7 \cdot \left(\frac{1}{16}\right)^k \phi''(\lambda_k)$.

Доказательство.
Заметим, что функция ψ убывающая и сильно выпуклая. Таким образом, при всех $k \geq 0$ имеем

$$\lambda_k \ < \ \lambda_{k+1} \ < \lambda_\star, \quad \psi'(\lambda_k) \ < 0 \ , \quad \psi(\lambda_k) \ > \ 0.$$

Поскольку $\psi(\lambda_k) \ \geq \ \psi(\lambda_{k+1}) + \psi'(\lambda_{k+1})(\lambda_k - \lambda_{k+1}) \ = \ \psi(\lambda_{k+1}) + \frac{\psi'(\lambda_{k+1})}{\psi'(\lambda_k)}\psi(\lambda_k)$, получаем

$$1 \ \geq \ \frac{\psi(\lambda_{k+1})}{\psi(\lambda_k)} + \frac{\psi'(\lambda_{k+1})}{\psi'(\lambda_k)} \ \geq \ 2\sqrt{\frac{\psi(\lambda_{k+1})\psi'(\lambda_{k+1})}{\psi(\lambda_k)\psi'(\lambda_k)}}.$$

Следовательно, при любом $k \geq 0$ получаем

$$\phi''(\lambda_k) \cdot |\phi'(\lambda_k)| \ \leq \left(\tfrac{1}{4}\right)^k \phi''(\lambda_0) \cdot |\phi'(\lambda_0)|. \tag{2.4.5}$$

Далее, в силу выбора λ_0 имеем

$$|\phi'(\lambda_0)| \ = \ \psi(\lambda_0) \ = \ \sum_{i=1}^{n} \frac{1}{\lambda_0 - s^{(i)}} - 1 \ < \ n - 1,$$

$$\phi''(\lambda_0) \ = \ \sum_{i=1}^{n} \frac{1}{(\lambda_0 - s^{(i)})^2} \ \leq \ n.$$

Наконец, поскольку $0 \leq \psi(\lambda_k) = \sum\limits_{i=1}^{n} \frac{1}{\lambda_k - s^{(i)}} - 1$, заключаем, что

$$\phi''(\lambda_k) \ = \ \sum_{i=1}^{n} \frac{1}{(\lambda_k - s^{(i)})^2} \ \geq \ \frac{1}{n}.$$

Используя эти оценки в неравенстве (2.4.5), получаем

$$\frac{1}{\phi''(\lambda_k)}(\phi'(\lambda_k))^2 \ \leq \ \left(\tfrac{1}{16}\right)^k \frac{(\phi''(\lambda_0))^2(\phi'(\lambda_0))^2}{(\phi''(\lambda_k))^3} \ \leq \ \left(\tfrac{1}{16}\right)^k \cdot n^7.$$

□

Сравнивая утверждение леммы 2.4.1 с определением области $\mathcal{Q}(s)$, мы видим, что процесс (2.4.4) приходит в область квадратичной сходимости после не более чем

$$\left\lceil \tfrac{1}{4}(2 + 7\log_2 n) \right\rceil \tag{2.4.6}$$

итераций. Каждая итерация требует $O(n)$ арифметических операций.

Похожая техника может использоваться для нахождения барьерной проекции в конусе неотрицательно определенных матриц:

$$\max_X \{\langle S, X\rangle + \ln\det X \ : \ \langle I, X\rangle = 1\}.$$

Наиболее прямолинейная стратегия состоит в нахождении всего спектра матрицы S и решении задачи (2.4.1) с вектором s, заполненным собственными значениями этой матрицы. Однако проще сначала представить S в форме UTU^T, где U – ортогональная матрица, а T – тридиагональная. После этого можно задать функцию

$$\phi(\lambda) \;\; = \;\; -n - \lambda - \ln\det(\lambda I - T)$$

и минимизировать ее методом Ньютона (2.4.4). Каждый шаг такой схемы потребует $O(n)$ арифметических операций.

Глава 3

Вариационные неравенства

3.1 Вариационные неравенства с гладким оператором

3.1.1 Введение

Мотивировка

Вариационные неравенства (ВН) с монотонными операторами представляют собой максимально общий класс задач, обладающих выпуклой структурой. В этом виде могут быть сформулированы выпуклые задачи оптимизации, задачи нахождения седловых точек у выпукло-вогнутых функций, различные задачи нахождения равновесия и т. д. В п. 2.1.5 мы уже рассматривали метод решения ВН с ограниченным оператором. Его оценка сложности составляет $O(\frac{1}{\epsilon^2})$ вызовов оракула, необходимых для решения задачи с точностью ϵ.

В этом параграфе мы рассматриваем методы для гладких монотонных операторов. Они основываются на простых правилах пересчета некоторой линейной функции, которая представляет собой модель нашей задачи. Это правило называется шагом *двойственной экстраполяции*. Применяя это правило последовательно, мы получим простую итеративную схему с оценкой сложности $O(\frac{1}{\epsilon})$, справедливой для ВН с липшицевым монотонным оператором. Та же схема с правильной стратегией одномерного поиска может быть применена к монотонным оператором с *ограниченным изменением* и, в частности, к негладким задачам. В этом случае оценка сложности превращается в $O(\frac{1}{\epsilon^2})$.

Содержание

Наш анализ скорости сходимости основывается на контроле изменения некоторой линейной модели на допустимом множестве. В п. 3.1.2 мы вводим все необходимые объекты и анализируем последствия одного шага двойственной экстраполяции (теорема 3.1.1). В п. 3.1.3 мы представляем две алгоритмические схемы и обосновываем их скорости сходимости на двух классах монотонных операторов: операторов, удовлетворяющих условию

Липшица, и операторов с ограниченным изменением. Мы показываем, как применить последнюю схему для решения задач негладкой минимизации.

В п. 3.1.4 мы применяем предложенные методы к задаче нахождения седловой точки гладкой выпукло-вогнутой функции. Показывается, как можно увеличить эффективность метода с помощью правильного масштабирования прямой и двойственной задачи.

В п. 3.1.5 мы приводим метод, непосредственно применимый к билинейным матричным играм. Показывается, что полученные результаты сравнимы с оценками эффективности, полученными с помощью техники сглаживания (см. гл. 5).

В завершающем п. 3.1.6 мы сравниваем полученные результаты со стандартным экстраградиентным методом.

Обозначения

Если пространство E представляет собой прямое произведение двух пространств, $E = E_1 \times E_2$, то вектор $x \in E$ разбивается на соответствующие две части:

$$x = (u, v) = (x^{(1)}, x^{(2)}) \quad \Rightarrow \quad u = x^{(1)} \in E_1, \quad v = x^{(2)} \in E_2.$$

В этой ситуации $f(x) \equiv f(u, v)$ и

$$\nabla f(x) = (\nabla_u f(u, v), \nabla_v f(u, v)) = (\nabla_1 f(x), \nabla_2 f(x)),$$

$$\nabla_u f(u, v) = \nabla_1 f(x) \in E_1^*, \quad \nabla_v f(u, v) = \nabla_2 f(x) \in E_2^*.$$

3.1.2 Двойственная экстраполяция

Пусть множество Q является выпуклым и замкнутым. Рассмотрим непрерывный оператор $g(x) : Q \to E^*$, который является *монотонным* на Q:

$$\langle g(x) - g(y), x - y \rangle \geq 0 \quad \forall x, y \in Q.$$

Нас будет интересовать следующая задача решения *вариационного неравенства*:

$$\text{найти } x^* \in Q \ : \quad \langle g(x), x - x^* \rangle \geq 0 \quad \forall x \in Q. \tag{3.1.1}$$

Иногда решение этой задачи называется *слабым* решением ВН. Поскольку оператор $g(x)$ непрерывен и монотонен, задача (3.1.1) эквивалентна нахождению *строгого решения*:

$$\text{найти } x^* \in Q \ : \quad \langle g(x^*), x - x^* \rangle \geq 0 \quad \forall x \in Q. \tag{3.1.2}$$

Для описания качества приближенных решений этих задач нам потребуются дополнительные объекты.

Пусть функция $d(x)$ является сильно выпуклой на Q:

$$d(\alpha x + (1-\alpha)y) \leq \alpha d(x) + (1-\alpha)d(y) - \alpha(1-\alpha)\tfrac{\sigma}{2}\|x - y\|^2 \quad \forall x, y \in Q, \ \alpha \in [0, 1]$$

с параметром выпуклости $\sigma > 0$. Рассмотрим ее сопряженную функцию относительно Q:

$$d_Q^*(s) = \max_x \{\langle s, x\rangle - d(x) : \; x \in Q\}.$$

Поскольку функция $d(x)$ сильно выпукла, функция $d_Q^*(s)$ определена и дифференцируема при всех $s \in E^*$. Обозначим $\tilde{Q} = \{\nabla d_Q^*(s), \; s \in E^*\}$. Ясно, что это множество выпукло и $\tilde{Q} \subseteq Q$. Нам потребуется следующее необременительное предположение.

Предположение 3.1.1 Функция $d(x)$ дифференцируема при любом $x \in \tilde{Q}$.

Поскольку функция $d(\cdot)$ сильно выпукла, при любых $x \in \tilde{Q}$ и $y \in Q$ имеем

$$d(y) \geq d(x) + \langle \nabla d(x), y - x\rangle + \tfrac{1}{2}\sigma \|y - x\|^2.$$

Определим теперь *брегмановское расстояние*:

$$\rho(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x\rangle, \quad x \in \tilde{Q}, \; y \in Q.$$

Ясно, что для $x \neq y$ имеем $\rho(x, y) > 0$ и $\rho(x, x) \equiv 0$. Заметим, что при фиксированном $x \in \tilde{Q}$ функция $\rho(x, y)$ сильно выпукла по y. Таким образом,

$$\rho(x, y) \geq \tfrac{1}{2}\sigma \|x - y\|^2, \quad x \in \tilde{Q}, \; y \in Q. \tag{3.1.3}$$

Выберем произвольную точку $\bar{x} \in \tilde{Q}$ в качестве *центра* множества Q. Заметим, что множество Q может быть неограниченным. Мы будем описывать качество произвольной точки $x \in Q$ как приближенного решения задачи (3.1.1) с помощью следующей *условной функции близости*:

$$f_D(x) = \max_{y \in Q} \{\langle g(y), x - y\rangle : \; \rho(\bar{x}, y) \leq D\},$$

где D – фиксированный положительный параметр.

Лемма 3.1.1 *Функция $f_D(x)$ выпукла на всем E. Для любого x из множества*

$$\mathcal{F}_D = \{y \in Q : \; \rho(\bar{x}, y) \leq D\}$$

выполнено неравенство $f_D(x) \geq 0$. Если x^* – решение задачи (3.1.1) и $\rho(\bar{x}, x^*) \leq D$, то $f_D(x^*) = 0$. Более того, если $f_D(\hat{x}) = 0$ при некотором $\hat{x} \in Q$, $\rho(\bar{x}, \hat{x}) < D$, то точка $\hat{x}$ является решением задачи (3.1.1).

Доказательство.
Действительно, функция $f_D(x)$ всюду определена, так как множество $\mathcal{F}_D$ ограничено. Она выпукла по x как максимум параметрического семейства линейных функций. Если $x \in \mathcal{F}_D$, то, очевидным образом, $f_D(x) \geq 0$.

Далее, пусть точка x^* является решением задачи (3.1.1). Тогда $\langle g(y), x^* - y\rangle \leq 0$ при всех y из $\mathcal{F}_D$. Включение $x^* \in \mathcal{F}_D$ означает, что $f_D(x^*) = 0$.

Наконец, пусть $f_D(\hat{x}) = 0$ при некотором $\hat{x} \in \mathcal{F}_D$. Это означает, что точка $\hat{x}$ является слабым решением вариационного неравенства

$$\langle g(y), y - \hat{x} \rangle \geq 0 \quad \forall y \in \mathcal{F}_D.$$

Поскольку функция $g(y)$ непрерывна, мы заключаем что точка $\hat{x}$ является также и строгим решением:

$$\langle g(\hat{x}), y - \hat{x} \rangle \geq 0 \quad \forall y \in Q, \ \rho(\bar{x}, y) \leq D.$$

Заметим, что минимум линейной функции $l(y) = \langle g(\hat{x}), y - \hat{x} \rangle$, $y \in \mathcal{F}_D$, достигается при $y = \hat{x}$. Более того, в этой точке ограничение $\rho(\bar{x}, y) \leq D$ не активно и его можно убрать. Мы заключаем что точка $\hat{x}$ является решением неравенств (3.1.2) и (3.1.1). □

В этом параграфе мы рассмотрим численные методы, которые генерируют точки с маленькими значениями условной функции близости $f_D(\cdot)$. Начнем со следующего наблюдения. Рассмотрим последовательность произвольных точек $\{y_i\}_{i=0}^N \subset Q$ и последовательность положительных весов $\{\lambda_i\}_{i=0}^N$. Обозначим

$$\begin{array}{rcl} S_N & = & \sum\limits_{i=0}^{N} \lambda_i, \qquad \tilde{y}_N \ = \ \frac{1}{S_N} \sum\limits_{i=0}^{N} \lambda_i y_i, \\ \Delta_N(D) & = & \max\limits_x \left\{ \sum\limits_{i=0}^{N} \lambda_i \langle g(y_i), y_i - x \rangle : \ x \in \mathcal{F}_D \right\}. \end{array} \tag{3.1.4}$$

Лемма 3.1.2 *Справедливо неравенство*

$$f_D(\tilde{y}_N) \leq \frac{1}{S_N} \Delta_N(D). \tag{3.1.5}$$

Доказательство.

Действительно, поскольку оператор $g(x)$ монотонен, имеем

$$\begin{array}{rcl} f_D(\tilde{y}_N) & = & \max\limits_x \{ \langle g(x), \tilde{y}_N - x \rangle : \ x \in \mathcal{F}_D \} \\ & = & \frac{1}{S_N} \max\limits_x \left\{ \sum\limits_{i=0}^{N} \lambda_i \langle g(x), y_i - x \rangle : \ x \in \mathcal{F}_D \right\} \\ & \leq & \frac{1}{S_N} \max\limits_x \left\{ \sum\limits_{i=0}^{N} \lambda_i \langle g(y_i), y_i - x \rangle : \ x \in \mathcal{F}_D \right\} \ \equiv \ \frac{1}{S_N} \Delta_N(D). \quad \Box \end{array}$$

Как и в параграфе 2.1, нам потребуются две вспомогательные функции множества Q опорного типа, определенные на всем E^*:

$$\begin{array}{rcl} \xi_D(s) & = & \max\limits_x \{ \langle s, x - \bar{x} \rangle : \ x \in \mathcal{F}_D \}, \\ V_\beta(z, s) & = & \max\limits_x \{ \langle s, x - z \rangle - \beta \rho(z, x) : \ x \in Q \}, \end{array} \tag{3.1.6}$$

где β и D – положительные параметры, и $z \in \tilde{Q}$ – его центр. Заметим, что функция $V_\beta(z,s)$ убывает по β. В то же время для

$$s_N = -\sum_{i=0}^{N} \lambda_i g(y_i) \tag{3.1.7}$$

имеем

$$\Delta_N(D) = \sum_{i=0}^{N} \lambda_i \langle g(y_i), y_i - \bar{x}\rangle + \xi_D(s_N). \tag{3.1.8}$$

Между этими функциями существует следующая связь:

$$\xi_D(s) \overset{(2.1.17)}{\leq} \beta D + V_\beta(\bar{x}, s). \tag{3.1.9}$$

Наконец, нам нужны оценки на скорость изменения функции $V_\beta(z,s)$ по s. Обозначим

$$T_\beta(z,s) = \arg\max_x \{\langle s, x - z\rangle - \beta\rho(z,x) :\ x \in Q\}.$$

Заметим, что $T_\beta(z,s) \in \tilde{Q}$. Поэтому в силу предположения 3.1.1 функция $d(\cdot)$ дифференцируема в точке $T_\beta(z,s)$. Обоснуем более строгий вариант неравенства (2.1.15).

Лемма 3.1.3 *Выберем произвольные* $s, \delta \in E^*$ *и произвольную точку* $z \in Q$. *Тогда*

$$V_\beta(z, s+\delta) \leq V_\beta(z,s) + \langle \delta, T_\beta(z,s) - z\rangle + V_\beta(T_\beta(z,s), \delta). \tag{3.1.10}$$

Более того, для любых $z \in Q$ *и* $s \in E^*$ *выполнено неравенство*

$$V_\beta(z,s) \leq \tfrac{1}{2\sigma\beta}\|s\|_*^2. \tag{3.1.11}$$

Доказательство.

Обозначим $T = T_\beta(z,s)$. В силу условия оптимальности первого порядка для задачи максимизации, определяющей значение $V_\beta(z,s)$, имеем

$$\langle s - \beta \cdot (\nabla d(T) - \nabla d(z)), x - T\rangle \leq 0 \quad \forall x \in Q.$$

Поэтому

$$\langle s, x - z\rangle \leq \langle s, T - z\rangle + \beta\langle \nabla d(T) - \nabla d(z), x - T\rangle \quad \forall x \in Q.$$

Таким образом,

$$\begin{array}{rcl} V_\beta(z,s+\delta) & = & \max\limits_{x\in Q} \;\; \{\langle s+\delta, x-z\rangle - \beta\cdot[d(x)-d(z)-\langle\nabla d(z), x-z\rangle]\} \\ & \leq & \max\limits_{x\in Q} \;\; \{\langle \delta, x-z\rangle + \langle s, T-z\rangle + \beta\langle\nabla d(T)-\nabla d(z), x-T\rangle \\ & & \qquad -\beta\cdot[d(x)-d(z)-\langle\nabla d(z), x-z\rangle]\} \\ & = & \max\limits_{x\in Q} \;\; \{\langle \delta, x-T\rangle - \beta\cdot[d(x)-d(T)-\langle\nabla d(T), x-T\rangle]\} \\ & & \qquad +\langle\delta, T-z\rangle + \langle s, T-z\rangle - \beta\cdot[d(T)-d(z)-\langle\nabla d(z), T-z\rangle] \\ & = & \qquad V_\beta(T,\delta)+\langle\delta, T-z\rangle + V_\beta(z,s). \end{array}$$

Остается заметить, что

$$V_\beta(z,s) = \max_{x\in Q}\{\langle s, x-z\rangle - \beta\cdot\rho(z,x)\} \leq \max_{x\in E}\{\langle s, x-z\rangle - \tfrac{1}{2}\sigma\beta\|x-z\|^2\} = \tfrac{1}{2\sigma\beta}\|s\|_*^2. \qquad \square$$

Замечание 4 Заметим, что функция $V_\beta(z,s)$ выпукла и дифференцируема по s. При этом ее градиент дается формулой $\nabla_s V_\beta(z,s) = T_\beta(z,s) - z \in E$. Таким образом, неравенство (3.1.10) может быть записано как

$$V_\beta(z,s+\delta) \leq V_\beta(z,s) + \langle\delta, \nabla_s V_\beta(z,s)\rangle + V_\beta(T_\beta(z,s),\delta).$$

Теперь ввиду неравенства (3.1.11), мы видим что градиент $\nabla_s V_\beta(z,s)$ удовлетворяет условию Липшица по s с константой $\frac{1}{\sigma\beta}$ (см., например, теорему 2.1.5 в [83]). $\square$

Зафиксируем некоторые положительные значения β и λ. Рассмотрим следующий шаг *двойственной экстраполяции* $\mathcal{E}_{\beta,\lambda}(s)$, который переводит произвольную точку $s \in E^*$ в новую точку s_+:

$$(x,y,s_+) = \mathcal{E}_{\beta,\lambda}(s) \quad \Leftrightarrow \quad \left\{ \begin{array}{lcl} x & = & T_\beta(\bar{x},s), \\ y & = & T_\beta(x,-\lambda\, g(x)), \\ s_+ & = & s-\lambda\, g(y). \end{array} \right. \tag{3.1.12}$$

Докажем основное свойство этого шага.

Лемма 3.1.4 *Пусть $s \in E^*$ и $(x,y,s_+) = \mathcal{E}_{\beta,\lambda}(s)$. Тогда*

$$\lambda\,\langle g(y), y-\bar{x}\rangle + V_\beta(\bar{x},s_+) \leq V_\beta(\bar{x},s) + \tfrac{\lambda^2}{2\sigma\beta}\|g(x)-g(y)\|_*^2 - \tfrac{1}{2}\sigma\beta\|x-y\|^2. \tag{3.1.13}$$

Доказательство.

Действительно, пользуясь неравенством (3.1.10) с $z = \bar{x}$ и $\delta = -\lambda\, g(y)$, получаем

$$V_\beta(\bar{x}, s_+) - V_\beta(\bar{x}, s) \leq \langle -\lambda\, g(y), x - \bar{x}\rangle + V_\beta(x, -\lambda\, g(y)). \tag{3.1.14}$$

Далее, пользуясь неравенством (3.1.10) с $z = x$, $s = -\lambda\, g(x)$, $\delta = \lambda \cdot (g(x) - g(y))$ и неравенством (3.1.11), получаем

$$\begin{aligned} V_\beta(x, -\lambda\, g(y)) &\leq V_\beta(x, -\lambda\, g(x)) + \lambda\langle g(x) - g(y), y - x\rangle + V_\beta(y, \lambda \cdot (g(x) - g(y))) \\ &\leq V_\beta(x, -\lambda\, g(x)) + \lambda\langle g(x) - g(y), y - x\rangle + \tfrac{\lambda^2}{2\sigma\beta}\|g(x) - g(y)\|_*^2. \end{aligned} \tag{3.1.15}$$

Заметим, что $V_\beta(x, -\lambda\, g(x)) = \langle -\lambda\, g(x), y - x\rangle - \beta\rho(x, y)$. Пользуясь этим неравенством, а также неравенствами (3.1.14), (3.1.15) и (3.1.3), получаем

$$\begin{aligned} &\lambda\, \langle g(y), y - \bar{x}\rangle + V_\beta(\bar{x}, s_+) - V_\beta(\bar{x}, s) \\ \leq\ & \lambda\, \langle g(y), y - \bar{x}\rangle - \lambda\, \langle g(y), x - \bar{x}\rangle + V_\beta(x, -\lambda\, g(y)) \\ \leq\ & \lambda\, \langle g(y), y - x\rangle + V_\beta(x, -\lambda g(x)) + \lambda\, \langle g(x) - g(y), y - x\rangle + \tfrac{\lambda^2}{2\sigma\beta}\|g(x) - g(y)\|_*^2 \\ =\ & \langle -\lambda\, g(x), y - x\rangle - \beta\rho(x, y) + \lambda\, \langle g(x), y - x\rangle + \tfrac{\lambda^2}{2\sigma\beta}\|g(x) - g(y)\|_*^2 \\ \leq\ & \tfrac{\lambda^2}{2\sigma\beta}\|g(x) - g(y)\|_*^2 - \tfrac{1}{2}\sigma\beta\|x - y\|^2. \end{aligned}$$

□

Остановимся на немедленных следствиях леммы 3.1.4.

Теорема 3.1.1 *Пусть* $s \in E^*$ *и* $(x, y, s_+) = \mathcal{E}_{\beta,\lambda}(s)$.

1. Предположим, что оператор $g(x)$ *непрерывен по Липшицу на* Q*:*

$$\|g(x_1) - g(x_2)\|_* \leq L\|x_1 - x_2\| \quad \forall x_1, x_2 \in Q \tag{3.1.16}$$

и параметры шага удовлетворяют условию $\lambda \leq \frac{\sigma}{L}\beta$*. Тогда*

$$\lambda\, \langle g(y), y - \bar{x}\rangle + V_\beta(\bar{x}, s_+) \leq V_\beta(\bar{x}, s). \tag{3.1.17}$$

2. Предположим, что оператор $g(x)$ *имеет ограниченное изменение на* Q*:*

$$\|g(x) - g(y)\|_* \leq M \quad \forall x, y \in Q. \tag{3.1.18}$$

Тогда

$$\lambda\, \langle g(y), y - \bar{x}\rangle + V_\beta(\bar{x}, s_+) \leq V_\beta(\bar{x}, s) + \tfrac{\lambda^2 M^2}{2\sigma\beta}. \tag{3.1.19}$$

3.1.3 Алгоритмические схемы

Теперь у нас есть все необходимое для анализа различных методов решения задачи (3.1.1). Начнем с вариационного неравенства с липшицевым оператором $g(x)$. Для простоты предположим, что соответствующая константа L нам известна. Рассмотрим следующую схему.

Метод решения гладких вариационных неравенств

Инициализация	Выбираем $\bar{x} \in \tilde{Q}$. Фиксируем $\beta = \frac{L}{\sigma}$ и выбираем $D > 0$. Задаем точность $\epsilon > 0$. Полагаем $s_{-1} = 0 \in E^*$.
Итерация	Вычисляем $(x_k, y_k, s_k) = \mathcal{E}_{\beta,1}(s_{k-1}),\ k \geq 0$.
Остановка	Если $\sum\limits_{i=0}^{k} \langle g(y_i), y_i - \bar{x} \rangle + \xi_D(s_k) \leq \epsilon \cdot (k+1)$, то Стоп.
Результат	$\tilde{y}_k = \frac{1}{k+1} \sum\limits_{i=0}^{k} y_i$.

(3.1.20)

Заметим, что параметр D входит только в критерий остановки метода.

Теорема 3.1.2 *Пусть оператор $g(x)$ липшицев на Q. Предположим, что последовательность $\{y_i\}_{i \geq 0}$ сформирована методом (3.1.20). Тогда для любого $k \geq 0$ выполнено неравенство*

$$f_D(\tilde{y}_k) \leq \frac{LD}{\sigma \cdot (k+1)}. \tag{3.1.21}$$

Критерий прерывания в методе (3.1.20) обеспечивает неравенство $f_D(\tilde{y}_k) \leq \epsilon$. Этот критерий будет выполнен не более чем через $1 + \lfloor \frac{LD}{\sigma\epsilon} \rfloor$ итераций.

Доказательство.

Заметим, что в методе (3.1.20) мы выбираем $\lambda_i \equiv 1$. Таким образом, $S_k = k + 1$. В силу неравенства (3.1.5), представления (3.1.8) и неравенства (3.1.9) для всех $k \geq 0$ имеем

$$\begin{array}{rcl} (k+1) \cdot f_D(\tilde{y}_k) & \leq & \sum\limits_{i=0}^{k} \langle g(y_i), y_i - \bar{x} \rangle + \xi_D(s_k) \\ \\ & \leq & a_k \equiv \sum\limits_{i=0}^{k} \langle g(y_i), y_i - \bar{x} \rangle + \beta D + V_\beta(\bar{x}, s_k). \end{array}$$

Используя неравенство (3.1.17), мы получаем

$$a_k = a_{k-1} + \langle g(y_k), y_k - \bar{x} \rangle + V_\beta(\bar{x}, s_k) - V_\beta(\bar{x}, s_{k-1}) \leq a_{k-1}, \quad k \geq 0.$$

Остается заметить, что $a_{-1} = \beta D + V_\beta(\bar{x}, s_{-1}) = \beta D$. □

Замечание 5 Критерий прерывания в методе (3.1.20) может быть заменен на такой:

$$\sum_{i=0}^{k} \langle g(y_i), y_i - \bar{x} \rangle + \beta D + V_\beta(\bar{x}, s_k) \le \epsilon \cdot (k+1).$$

В этом случае все утверждения теоремы 3.1.2 остаются верными.

Ясно, что оценка (3.1.21) содержательна, только если $D \ge \rho(\bar{x}, x^*)$, где x^* – решение задачи (3.1.1). Если, например, множество Q ограничено, то мы можем выбрать

$$D = \Theta(\bar{x}) = \max_{y \in Q} \rho(\bar{x}, y).$$

Тогда получаем следующий вариант оценки (3.1.21):

$$f_D(\tilde{y}_k) \le \tfrac{L\Theta(\bar{x})}{\sigma \cdot (k+1)}. \tag{3.1.22}$$

Заметим, что с точки зрения информационной сложности эта оценка не может быть улучшена (см. [77]).

Рассмотрим теперь задачу (3.1.1), в которой оператор $g(x)$ имеет ограниченное изменение на множестве Q (см. (3.1.18)). Для простоты предположим, что соответствующая константа M нам известна. Рассмотрим следующий метод.

Метод для ВН с ограниченным изменением оператора

Инициализация Выбираем $\bar{x} \in \tilde{Q}$ и $D > 0$. Для $k \ge 0$ положим

$$\beta_k = M\sqrt{\tfrac{k+1}{\sigma D}}.$$

Выберем точность $\epsilon > 0$. Положим $s_{-1} = 0 \in E^*$.

Итерация Вычисляем $(x_k, y_k, s_k) = \mathcal{E}_{\beta_k,1}(s_{k-1}), \ k \ge 0$.

Остановка Если $\sum\limits_{i=0}^{k} \langle g(y_i), y_i - \bar{x} \rangle + \xi_D(s_k) \le \epsilon \cdot (k+1)$,
то Стоп.

Результат $\tilde{y}_k = \frac{1}{k+1} \sum\limits_{i=0}^{k} y_i$.

(3.1.23)

Теорема 3.1.3 *Предположим, что оператор$g(x)$ имеет ограниченное изменение на множестве Q:*

$$\|g(x_1) - g(x_2)\|_* \leq M \quad \forall x_1, x_2 \in Q. \tag{3.1.24}$$

Пусть последовательность $\{y_i\}_{i\geq 0}$ сформирована методом (3.1.23). Тогда для любого $k \geq 0$ выполнено неравенство

$$f_D(\tilde{y}_k) \leq 2M\sqrt{\tfrac{D}{\sigma\cdot(k+1)}}. \tag{3.1.25}$$

Критерий прерывания в методе (3.1.23) обеспечивает выполнение неравенства $f_D(\tilde{y}_k) \leq \epsilon$. Этот критерий сработает не более чем за $1 + \lfloor \frac{4M^2D}{\sigma\epsilon^2} \rfloor$ итераций.

Доказательство.

Заметим, что в схеме (3.1.23) выбираются $\lambda_i \equiv 1$. Таким образом, $S_k = k+1$. В силу неравенства (3.1.5), представления (3.1.8) и неравенства (3.1.9) для всех $k \geq 0$ имеем

$$(k+1)\cdot f_D(\tilde{y}_k) \leq \sum_{i=0}^{k} \langle g(y_i), y_i - \bar{x}\rangle + \xi_D(s_k) \leq \beta_k D + b_k,$$

где $b_k = \sum\limits_{i=0}^{k} \langle g(y_i), y_i - \bar{x}\rangle + V_{\beta_k}(\bar{x}, s_k)$. Используя неравенство (3.1.19) и предположение (3.1.24), получаем

$$\begin{aligned} b_k - b_{k-1} &= \langle g(y_k), y_k - \bar{x}\rangle + V_{\beta_k}(\bar{x}, s_k) - V_{\beta_{k-1}}(\bar{x}, s_{k-1}) \\ &\leq \langle g(y_k), y_k - \bar{x}\rangle + V_{\beta_k}(\bar{x}, s_k) - V_{\beta_k}(\bar{x}, s_{k-1}) \\ &\leq \tfrac{M^2}{2\sigma\beta_k} = \tfrac{1}{2}M\sqrt{\tfrac{D}{\sigma\cdot(k+1)}}, \quad k \geq 1. \end{aligned}$$

Используя неравенство (3.1.19) снова, получаем

$$b_0 = \langle g(y_0), y_0 - \bar{x}\rangle + V_{\beta_0}(\bar{x}, s_0) \leq V_{\beta_0}(\bar{x}, s_{-1}) + \tfrac{M^2}{2\sigma\beta_0} = \tfrac{1}{2}M\sqrt{\tfrac{D}{\sigma}}.$$

Таким образом, пользуясь индукцией, легко видеть что

$$b_k \leq \tfrac{1}{2}M\sqrt{\tfrac{D}{\sigma}} \sum_{i=0}^{k} \tfrac{1}{\sqrt{i+1}} \leq M\sqrt{\tfrac{D}{\sigma}(k+1)} \equiv \beta_k D. \qquad \square$$

Заметим, что метод (3.1.23) может использоваться для решения задачи недифференцируемой оптимизации

$$\text{найти } f^* = \min_x \{f(x) : \ x \in Q\}, \tag{3.1.26}$$

где функция $f(x)$ выпукла и субдифференцируема в любой точке выпуклого замкнутого множества Q. Обозначим через $g(x)$ произвольный субградиент функции $f(\cdot)$ в точке $x \in Q$.

Теорема 3.1.4 *Предположим, что оператор* $g(x)$ *в задаче (3.1.26) имеет ограниченное изменение на* Q*:*

$$\|g(x) - g(y)\|_* \leq M \quad \forall x, y \in Q. \tag{3.1.27}$$

Пусть последовательность $\{y_i\}_{i\geq 0}$ *формируется методом (3.1.23). Тогда для любого* $k \geq 0$ *выполнено неравенство*

$$\min_{0\leq i\leq k} f(y_i) - \min_x \{f(x) : \, x \in \mathcal{F}_D\} \leq 2M\sqrt{\tfrac{D}{\sigma\cdot(k+1)}}. \tag{3.1.28}$$

Доказательство.
Пользуясь тем же рассуждением, что и в теореме 3.1.3, можно доказать, что

$$2M\sqrt{\tfrac{D}{\sigma}(k+1)} \geq \sum_{i=0}^{k}\langle g(y_i), y_i - \bar{x}\rangle + \xi_D(s_k) \equiv \max_x\{\sum_{i=0}^{k}\langle g(y_i), y_i - x\rangle : \, x \in \mathcal{F}_D\}.$$

Поскольку $\langle g(y_i), y_i - x\rangle \geq f(y_i) - f(x)$, мы немедленно получаем оценку (3.1.28). □

3.1.4 Вычисление седловых точек

Покажем, как применять метод (3.1.20) для поиска седловых точек. Рассмотрим задачу

$$\text{найти } f^* = \min_{u\in Q_1}\max_{v\in Q_2} f(u, v), \tag{3.1.29}$$

где множества $Q_1 \subseteq E_1$ и $Q_2 \subseteq E_2$ выпуклы и замкнуты, а гладкая функция $f(u, v)$ выпукла по u и вогнута по v. Пусть в пространствах E_i, $i = 1, 2$, заданы нормы $\|\cdot\|_i$.

Рассмотрим следующий оператор:

$$g(x) = \begin{pmatrix} \nabla_u f(u, v) \\ -\nabla_v f(u, v) \end{pmatrix}, \quad x = (u, v) \in Q \equiv Q_1 \times Q_2. \tag{3.1.30}$$

Поскольку функция $f(u, v)$ выпукла по u и вогнута по v, оператор $g(x)$ монотонен на Q. Более того, легко показать, что с этим оператором вариационное неравенство (3.1.1) имеет то же решение что и седловая задача (3.1.29). Таким образом, мы можем решать задачу (3.1.29) с помощью метода (3.1.20). Однако в силу специальной структуры оператора $g(x)$, можно получить более точные оценки.

Нам потребуется следующее предположение.

Предположение 3.1.2 Оператор $g(x)$ липшицев на Q в следующем смысле:

$$\begin{aligned} \|\nabla_u f(u + \Delta u, v + \Delta v) - \nabla_u f(u, v)\|_{1_*} &\leq L_{11}(f)\|\Delta u\|_1 + L_{12}(f)\|\Delta v\|_2, \\ \|\nabla_v f(u + \Delta u, v + \Delta v) - \nabla_v f(u, v)\|_{2_*} &\leq L_{21}(f)\|\Delta u\|_1 + L_{22}(f)\|\Delta v\|_2 \end{aligned} \tag{3.1.31}$$

для всех $u, u + \Delta u \in Q_1$ и $v, v + \Delta v \in Q_2$.

Заметим, что без ограничения общности мы можем считать, что

$$L_{11}(f) \le 1, \quad L_{12}(f) \le 1, \quad L_{21}(f) \le 1, \quad L_{22}(f) \le 1. \tag{3.1.32}$$

Действительно, легко доказывается следующий результат.

Лемма 3.1.5 *Пусть функция $f(u,v)$ удовлетворяет предположению 3.1.2. Зафиксируем произвольные константы α и γ. Тогда функция*

$$\hat{f}(u,v) = f(\alpha u, \gamma v), \quad (u,v) \in \hat{Q} = \{(u,v):\ (\alpha u, \gamma v) \in Q\},$$

удовлетворяет предположению 3.1.2 с

$$L_{11}(\hat{f}) = \alpha^2 L_{11}(f), \qquad L_{12}(\hat{f}) = |\alpha\gamma| \cdot L_{12}(f),$$

$$L_{21}(\hat{f}) = |\alpha\gamma| \cdot L_{21}(f), \quad L_{22}(\hat{f}) = \gamma^2 L_{22}(f).$$

Доказательство.
Поскольку $\nabla_u \hat{f}(u,v) = \alpha \nabla f(\alpha u, \gamma v)$, из первого неравенства (3.1.31) получаем

$$\|\nabla_u \hat{f}(u+\Delta u, v+\Delta v) - \nabla_u \hat{f}(u,v)\|_{1_*}$$

$$= |\alpha| \cdot \|\nabla_u f(\alpha\cdot(u+\Delta u), \gamma\cdot(v+\Delta v)) - \nabla_u f(\alpha u, \gamma v)\|_{1_*}$$

$$\le \alpha^2 L_{11}(f)\|\Delta u\|_1 + |\alpha\gamma| \cdot L_{12}(f)\|\Delta v\|_2.$$

Оставшаяся часть леммы доказывается аналогично. □

Предположим теперь, что функция $f(u,v)$ удовлетворяет условию (3.1.32). Введем в пространстве $E = E_1 \times E_2$ следующую норму:

$$\|x\| = \max\{\|u\|_1, \|v\|_2\}, \quad x = (u,v) \in E.$$

Тогда в двойственном пространстве $E^* = E_1^* \times E_2^*$ норма должна быть такой:

$$\|s\|_* = \|s^{(1)}\|_{1_*} + \|s^{(2)}\|_{2_*}, \quad s = (s^{(1)}, s^{(2)}) \in E^*.$$

Лемма 3.1.6 *Если функция $f(u,v)$ удовлетворяет условию (3.1.32), то*

$$\|g(x) - g(y)\|_* \le 4\|x - y\| \quad \forall x, y \in Q. \tag{3.1.33}$$

Если к тому же $L_{11}(f) = L_{22}(f) = 0$, то

$$\|g(x) - g(y)\|_* \le 2\|x - y\| \quad \forall x, y \in Q. \tag{3.1.34}$$

Доказательство.
В силу определения наших норм для любых $x = (u_1, v_1)$ и $y = (u_2, v_2)$ из Q имеем

$$\begin{aligned}\|g(x) - g(y)\|_* &= \|\nabla_u f(u_1, v_1) - \nabla_u f(u_2, v_2)\|_{1*} + \|\nabla_v f(u_1, v_1) - \nabla_v f(u_2, v_2)\|_{2*} \\ &\le (L_{11}(f) + L_{21}(f))\|u_1 - u_2\|_1 + (L_{22}(f) + L_{12}(f))\|v_1 - v_2\|_2 \\ &\le 4\|x - y\|.\end{aligned}$$

Второе утверждение леммы очевидно. □

В соответствии с п. 3.1.2 нам необходимы две функции $d_1(u)$ и $d_2(v)$, удовлетворяющие предположению 3.1.1, которые сильно выпуклыми с параметрами $\sigma_i > 0$, $i = 1, 2$:

$$\begin{aligned}d_1(u_1) &\ge d_1(u_2) + \langle \nabla d_1(u_2), u_1 - u_2\rangle_1 + \tfrac{1}{2}\sigma_1\|u_1 - u_2\|_1^2, \quad u_1 \in Q_1,\ u_2 \in \tilde{Q}_1 \\ d_2(v_1) &\ge d_2(v_2) + \langle \nabla d_2(v_2), v_1 - v_2\rangle_2 + \tfrac{1}{2}\sigma_2\|v_1 - v_2\|_2^2, \quad v_1 \in Q_2,\ v_2 \in \tilde{Q}_2.\end{aligned}$$

Тогда можно будет определить брегмановские расстояния:

$$\begin{aligned}\rho_1(u_1, u_2) &= d_1(u_2) - d_1(u_1) + \langle \nabla d_1(u_1), u_2 - u_1\rangle_1, \quad u_1 \in \tilde{Q}_1,\ u_2 \in Q_1, \\ \rho_2(v_1, v_2) &= d_2(v_2) - d_2(v_1) + \langle \nabla d_2(v_1), v_2 - v_1\rangle_2, \quad v_1 \in \tilde{Q}_2,\ v_2 \in Q_2.\end{aligned}$$

Теперь можно измерять качество приближенных решений задачи (3.1.29). Зафиксируем точку $\bar{x} = (\bar{u}, \bar{v}) \in \tilde{Q} \equiv \tilde{Q}_1 \times \tilde{Q}_2$ в качестве центра множества Q. Предположим, что существует решение (u^*, v^*) задачи (3.1.29), удовлетворяющее условиям

$$\rho_1(\bar{u}, u^*) \le D_1, \quad \rho_2(\bar{v}, v^*) \le D_2. \tag{3.1.35}$$

Тогда можно определить следующие условные меры близости:

$$\begin{aligned}F_{D_2}(u) &= \max_v \{f(u, v) :\ v \in Q_2,\ \rho_2(\bar{v}, v) \le D_2\}, \\ \Phi_{D_1}(v) &= \min_u \{f(u, v) :\ u \in Q_1,\ \rho_1(\bar{u}, u) \le D_1\}.\end{aligned}$$

Ясно, что $f^* = \min\limits_{u \in Q_1} F_{D_2}(u) = \max\limits_{v \in Q_2} \Phi_{D_1}(v)$.

Рассмотрим произвольную последовательность точек $y_i = (u_i, v_i) \in Q$, $i = 0, \dots, N$. Обозначим

$$\begin{aligned}\tilde{y}_N &= \tfrac{1}{N+1} \sum_{i=0}^{N} y_i \equiv (\tilde{u}_N, \tilde{v}_N), \\ \hat{\Delta}_N(D_1, D_2) &= \sup_y \{\sum_{i=0}^{N} \langle g(y_i), y_i - x\rangle :\ x \in \hat{\mathcal{F}}_{D_1, D_2}\},\end{aligned}$$

где $\hat{\mathcal{F}}(D_1, D_2) = \{x = (u, v) \in Q :\ \rho_1(\bar{u}, u) \le D_1,\ \rho_2(\bar{v}, v) \le D_2\}$.

Лемма 3.1.7 *Справедливы неравенства*

$$0 \le [F_{D_1}(\tilde{u}_N) - f^*] + [f^* - \Phi_{D_2}(\tilde{v}_N)] \le \tfrac{1}{N+1}\hat{\Delta}_N(D_1, D_2). \qquad (3.1.36)$$

Доказательство.

Поскольку функция $f(u,v)$ выпукла по u и вогнута по v, имеем

$$\hat{\Delta}_N(D_1, D_2) = \sup_x \{ \sum_{i=0}^{N} \langle g(y_i), y_i - x \rangle : \ x \in \hat{\mathcal{F}}_{D_1,D_2} \}$$

$$= \ \sup_{(u,v)} \{ \sum_{i=0}^{N} [\langle \nabla_u f(u_i, v_i), u_i - u \rangle_1 + \langle \nabla_v f(u_i, v_i), v - v_i \rangle_2] : \ (u,v) \in \hat{\mathcal{F}}_{D_1,D_2} \}$$

$$\ge \ \sup_{(u,v)} \{ \sum_{i=0}^{N} [(f(u_i, v_i) - f(u, v_i)) + (f(u_i, v) - f(u_i, v_i))] : \ (u,v) \in \hat{\mathcal{F}}_{D_1,D_2} \}$$

$$\ge \ (N+1) \sup_{(u,v)} \{ f(\tilde{u}_N, v) - f(u, \tilde{v}_N) : \ (u,v) \in \hat{\mathcal{F}}_{D_1,D_2} \}$$

$$= \ (N+1)(F_{D_2}(\tilde{u}_N) - \Phi_{D_1}(\tilde{v}_N)). \qquad \square$$

Как и в п. 3.1.2, нам нужно ограничить значение $\hat{\Delta}_N(D_1, D_2)$ сверху. Обозначим

$$\xi_{D_1,D_2}(s) = \max_x \{ \langle s, x - \bar{x} \rangle : \ x \in \hat{\mathcal{F}}_{D_1,D_2} \}.$$

Заметим, что для $s_N = -\sum\limits_{i=0}^{N} g(y_i)$ имеем

$$\hat{\Delta}_N(D_1, D_2) = \sum_{i=0}^{N} \langle g(y_i), y_i - \bar{x} \rangle + \xi_{D_1,D_2}(s_N).$$

Для $x_1 = (u_1, v_1) \in \tilde{Q}$ и $x_2 = (u_2, v_2) \in Q$ обозначим

$$\rho(x_1, x_2) = \tfrac{1}{D_1}\rho_1(u_1, u_2) + \tfrac{1}{D_2}\rho_2(v_1, v_2). \qquad (3.1.37)$$

Ясно что функция $\rho(x,y)$ сильно выпукла по y на Q с параметром выпуклости

$$\sigma = \min\{ \tfrac{\sigma_1}{D_1}, \tfrac{\sigma_2}{D_2} \}. \qquad (3.1.38)$$

Определим функцию $V_\beta(z,s)$ аналогично (3.1.6):

$$V_\beta(z,s) = \max_x \{ \langle s, x - z \rangle - \beta\rho(z,x) : \ x \in Q \}, \quad z \in \tilde{Q}, \ s \in E^*.$$

Лемма 3.1.8 *Для любых положительных* D_1, D_2 *и* β, *и любого* $s \in E^*$ *выполняется неравенство*

$$\xi_{D_1,D_2}(s) \le 2\beta + V_\beta(\bar{x}, s). \qquad (3.1.39)$$

Доказательство.

Доказательство очень похоже на доказательство леммы 2.1.2:

$$\begin{aligned}\xi_{D_1,D_2}(s) &= \max_x\{\langle s, x-\bar{x}\rangle :\ x=(u,v)\in Q,\ \tfrac{1}{D_1}\rho_1(\bar{u},u)\le 1,\ \tfrac{1}{D_2}\rho_2(\bar{v},v)\le 1\}\\ &\le \max_x\{\langle s, x-\bar{x}\rangle :\ x=(u,v)\in Q,\ \rho(\bar{x},x)\le 2\}\\ &\le 2\beta + V_\beta(\bar{x},s).\end{aligned}$$ □

Теперь можно обосновать применение общей схемы (3.1.20) к задаче (3.1.1).

Теорема 3.1.5 *Пусть функция $f(x,y)$ удовлетворяет условиям (3.1.32) и (3.1.35). Предположим, что метод (3.1.20) применяется к задаче (3.1.29) и при этом используются оператор $g(x)$ определенный формулой (3.1.30), прокс-функция $\rho(x,y)$ заданная формулой (3.1.37), и следующее значение параметра β:*

$$\beta = 4\max\left\{\tfrac{D_1}{\sigma_1}, \tfrac{D_2}{\sigma_2}\right\}. \tag{3.1.40}$$

Тогда скорость сходимости метода оценивается следующим образом:

$$0 \le F_{D_2}(\tilde{u}_N) - \Phi_{D_1}(\tilde{v}_N) \le \tfrac{8}{N+1}\max\left\{\tfrac{D_1}{\sigma_1}, \tfrac{D_2}{\sigma_2}\right\}. \tag{3.1.41}$$

Если к тому же $L_{11}(f) = L_{22}(f) = 0$, то можно взять $\beta = 2\max\left\{\tfrac{D_1}{\sigma_1}, \tfrac{D_2}{\sigma_2}\right\}$ и получить

$$0 \le F_{D_2}(\tilde{u}_N) - \Phi_{D_1}(\tilde{v}_N) \le \tfrac{4}{N+1}\max\left\{\tfrac{D_1}{\sigma_1}, \tfrac{D_2}{\sigma_2}\right\}. \tag{3.1.42}$$

Доказательство.

Действительно, поскольку функция $f(u,v)$ удовлетворяет условию (3.1.32), неравенство (3.1.33) дает нам значение константы Липшица $L=4$ для оператора $g(x)$. Таким образом, в силу соотношений (3.1.40) и (3.1.38) условия пункта 1 теоремы 3.1.1 выполнены. Поэтому, пользуясь неравенством (3.1.39) и таким же рассуждением, как и в доказательстве теоремы 3.1.2 получаем первое нужное утверждение.

Второе утверждение следует из второй части утверждения леммы 3.1.6. □

3.1.5 Билинейные матричные игры

Сравним полученные результаты с тем, что можно получить с помощью техники сглаживания (см. гл. 5) применительно к следующей задаче. Пусть линейный оператор A отображает пространство E_1 в E_2^*, а множества $Q_1 \subset E_1$ и $Q_2 \subset E_2$ выпуклы и замкнуты. Рассмотрим задачу

$$\text{найти } f^* = \min_{u\in Q_1}\max_{v\in Q_2} f(u,v), \tag{3.1.43}$$

где $f(u,v) = \langle Au - c, v\rangle_2 + \langle b, u\rangle_1$, $c \in E_2^*$ и $b \in E_1^*$. В соответствии с формулой (3.1.30), оператор $g(x)$ определяется следующим образом:

$$g(x) = \begin{pmatrix} b + A^T v \\ c - Au \end{pmatrix}.$$

Тогда $L_{11}(f) = L_{22}(f) = 0$ и

$$L_{12}(f) = L_{21}(f) = \|A\|_{1,2} \equiv \max_{u,v}\{\langle Au, v\rangle_2 : \ \|u\|_1 \leq 1, \ \|v\|_2 \leq 1\} \overset{\text{def}}{=} L.$$

Для того чтобы получить единичную константу Липшица, нам нужно отшкалировать аргументы функции f некоторыми положительными множителями α и γ. Ввиду леммы 3.1.5 эти множители должны удовлетворять неравенству

$$\alpha\gamma \leq \tfrac{1}{L}. \tag{3.1.44}$$

Таким образом, мы получаем следующую задачу:

$$\text{Find } f^* = \min_{\hat{u} \in \hat{Q}_1} \max_{\hat{v} \in \hat{Q}_2} \hat{f}(\hat{u}, \hat{v}), \tag{3.1.45}$$

где $\hat{f}(\hat{u}, \hat{v}) = f(\alpha\hat{u}, \gamma\hat{v})$, $\hat{Q}_1 = \{\hat{u} : \ \alpha\hat{u} \in Q_1\}$, и $\hat{Q}_2 = \{\hat{v} : \ \gamma\hat{v} \in Q_2\}$.

Пусть функция $d_1(u)$ сильно выпукла на выпуклом множестве Q_1 с параметром σ_1. Рассмотрим функцию $\hat{d}_1(\hat{u}) = d_1(\alpha\hat{u})$.

Лемма 3.1.9 *Функция $\hat{d}_1(\hat{u})$ сильно выпукла на множестве $\hat{Q}_1$ с параметром выпуклости*

$$\hat{\sigma}_1 = \sigma_1\alpha^2.$$

Доказательство.
Действительно, для любого $\hat{u}_1, \hat{u}_2 \in \hat{Q}_1$ и $\lambda \in [0,1]$ имеем

$$\begin{aligned} \lambda\hat{d}_1(\hat{u}_1) + (1-\lambda)\hat{d}_1(\hat{u}_2) &= \lambda d_1(\alpha\hat{u}_1) + (1-\lambda)d_1(\alpha\hat{u}_2) \\ &\geq d_1(\alpha \cdot (\lambda\hat{u}_1 + (1-\lambda)\hat{u}_2)) + \tfrac{1}{2}\sigma_1\lambda(1-\lambda)\|\alpha \cdot (\hat{u}_1 - \hat{u}_2)\|^2 \\ &= \hat{d}_1(\lambda\hat{u}_1 + (1-\lambda)\hat{u}_2) + \tfrac{1}{2}\sigma_1\alpha^2\lambda(1-\lambda)\|\hat{u}_1 - \hat{u}_2\|^2. \qquad \square \end{aligned}$$

Аналогично функция $\hat{d}_2(\hat{v}) \equiv d_2(\gamma\hat{v})$ сильно выпукла на множестве $\hat{Q}_2$ с параметром выпуклости $\hat{\sigma}_2 = \sigma_2\gamma^2$. Теперь мы можем задать брегмановское расстояние для отшкалированных множеств:

$$\begin{aligned} \hat{\rho}_1(\hat{u}_1, \hat{u}) &= \hat{d}_1(\hat{u}) - \hat{d}_1(\hat{u}_1) - \langle \nabla\hat{d}_1(\hat{u}_1), \hat{u} - \hat{u}_1\rangle_1 \\ &= d_1(\alpha\hat{u}) - d_1(\alpha\hat{u}_1) - \langle \nabla d_1(\alpha\hat{u}_1), \alpha\hat{u} - \alpha\hat{u}_1\rangle_1 = \rho_1(\alpha\hat{u}_1, \alpha\hat{u}), \\ \hat{\rho}_2(\hat{v}_1, \hat{v}) &= \hat{d}_2(\hat{v}) - \hat{d}_2(\hat{v}_1) - \langle \nabla\hat{d}_2(\hat{v}_1), \hat{v} - \hat{v}_1\rangle_2 \\ &= d_2(\gamma\hat{v}) - d_2(\gamma\hat{v}_1) - \langle \nabla d_2(\gamma\hat{v}_1), \gamma\hat{v} - \gamma\hat{v}_1\rangle_2 = \rho_2(\gamma\hat{v}_1, \gamma\hat{v}). \end{aligned}$$

Обозначим $D_1 = \max\limits_{u\in Q_1} \rho_1(\bar{u}, u) = \max\limits_{\hat{u}\in \hat{Q}_1} \hat{\rho}_1(\hat{\bar{u}}, \hat{u})$, $D_2 = \max\limits_{v\in Q_2} \rho_2(\bar{v}, v) = \max\limits_{\hat{v}\in \hat{Q}_2} \hat{\rho}_2(\hat{\bar{v}}, \hat{v})$. Тогда для точек $\hat{x}_1 = (\hat{u}_1, \hat{v}_1)$ и $\hat{x}_2 = (\hat{u}_2, \hat{v}_2)$ можно определить следующее расстояние:

$$\begin{aligned} \hat{\rho}(\hat{x}_1, \hat{x}_2) &= \tfrac{1}{D_1}\hat{\rho}_1(\hat{u}_1, \hat{u}_2) + \tfrac{1}{D_2}\hat{\rho}_2(\hat{v}_1, \hat{v}_2) \\ &= \tfrac{1}{D_1}\rho_1(\alpha\hat{u}_1, \alpha\hat{u}_2) + \tfrac{1}{D_2}\rho_2(\gamma\hat{v}_1, \gamma\hat{v}_2). \end{aligned}$$

Остается выписать оператор отшкалированной задачи:

$$\hat{g}(\hat{x}) = \begin{pmatrix} \nabla_{\hat{u}}\hat{f}(\hat{u}, \hat{v}) \\ -\nabla_{\hat{v}}\hat{f}(\hat{u}, \hat{v}) \end{pmatrix} = \begin{pmatrix} \alpha\nabla_u f(\alpha\hat{u}, \gamma\hat{v}) \\ -\gamma\nabla_v f(\alpha\hat{u}, \gamma\hat{v}) \end{pmatrix} = \begin{pmatrix} \alpha\cdot[b + \gamma A^T\hat{v}] \\ \gamma\cdot[c - \alpha A\hat{u}] \end{pmatrix}.$$

Теперь выпишем все операции метода (3.1.20), решающего задачу (3.1.45). Мы пометим шапкой все объекты, сформированные этим методом при решении отшкалированной задачи. Параллельно мы реконструируем соответствующие объекты исходной задачи. Заметим, что в силу теоремы 3.1.5 можно пользоваться таким значением параметра β:

$$\beta = 2\max\left\{\tfrac{D_1}{\hat{\sigma}_1}, \tfrac{D_2}{\hat{\sigma}_2}\right\} = 2\max\left\{\tfrac{D_1}{\sigma_1\alpha^2}, \tfrac{D_2}{\sigma_2\gamma^2}\right\}.$$

Ввиду ограничения (3.1.44) оптимальным способом выбора параметров будет следующий:

$$\alpha = \tfrac{1}{L^{1/2}}\left[\tfrac{D_1\sigma_2}{D_2\sigma_1}\right]^{1/4}, \quad \gamma = \tfrac{1}{L^{1/2}}\left[\tfrac{D_2\sigma_1}{D_1\sigma_2}\right]^{1/4}, \quad \beta = 2L\left[\tfrac{D_1D_2}{\sigma_1\sigma_2}\right]^{1/2}. \tag{3.1.46}$$

Первой операцией на каждой итерации является вычисление точки $\hat{x}_k = \hat{T}_\beta(\hat{\bar{x}}, \hat{s}_{k-1})$. Для этого надо решить задачу

$$\begin{aligned} &\max_{\hat{x}}\{\langle\hat{s}_{k-1}, \hat{x} - \hat{\bar{x}}\rangle - \beta\hat{\omega}(\hat{\bar{x}}, \hat{x}) :\ \hat{x}\in\hat{Q}\} \\ = \ &\max_{\hat{u},\hat{v}}\{\langle\hat{s}^{(1)}_{k-1}, \hat{u} - \hat{\bar{u}}\rangle_1 + \langle\hat{s}^{(2)}_{k-1}, \hat{v} - \hat{\bar{v}}\rangle_2 \qquad\qquad (3.1.47) \\ &-\beta\cdot[\tfrac{1}{D_1}\rho_1(\alpha\hat{\bar{u}}, \alpha\hat{u}) + \tfrac{1}{D_2}\rho_2(\gamma\hat{\bar{v}}, \gamma\hat{v})] :\ \alpha\hat{u}\in Q_1,\ \gamma\hat{v}\in Q_2\}. \end{aligned}$$

Заметим что точки $\hat{s}_k$ формируются в методе (3.1.20) следующим образом:

$$\hat{s}_k = -\sum_{i=0}^{k}\hat{g}(\hat{y}_i) = -\begin{pmatrix} \alpha\sum\limits_{i=0}^{k}[b + \gamma A^T\hat{y}^{(2)}_i] \\ \gamma\sum\limits_{i=0}^{k}[c - \alpha A\hat{y}^{(1)}_i] \end{pmatrix} \equiv \begin{pmatrix} \alpha s^{(1)}_k \\ \gamma s^{(2)}_k \end{pmatrix}.$$

Поэтому в силу соотношений (3.1.46) вычисление (3.1.47) сводится к решению двух независимых задач:

$$\begin{aligned} \alpha\hat{x}^{(1)}_k &\equiv x^{(1)}_k = \arg\max_u\{\langle s^{(1)}_{k-1}, u - \bar{u}\rangle_1 - \tfrac{2}{\sigma_1\alpha^2}\rho_1(\bar{u}, u) :\ u\in Q_1\}, \\ \gamma\hat{x}^{(2)}_k &\equiv x^{(2)}_k = \arg\max_v\{\langle s^{(2)}_{k-1}, v - \bar{v}\rangle_2 - \tfrac{2}{\sigma_2\gamma^2}\rho_2(\bar{v}, v) :\ v\in Q_2\}. \end{aligned}$$

Далее, значение оператора $\hat{g}(\cdot)$ в точке $\hat{x}_k$ определяется следующим образом:

$$\hat{g}(\hat{x}_k) = \begin{pmatrix} \alpha \cdot [b + \gamma A^T \hat{x}_k^{(2)}] \\ \gamma \cdot [c - \alpha A \hat{x}_k^{(1)}] \end{pmatrix} = \begin{pmatrix} \alpha \cdot [b + A^T x_k^{(2)}] \\ \gamma \cdot [c - A x_k^{(1)}] \end{pmatrix}.$$

Таким образом, мы получаем следующие правила вычисления $\hat{y}_k$:

$$\begin{array}{rcl} \alpha \hat{y}_k^{(1)} \equiv y_k^{(1)} & = & \arg\max\limits_u \{ -\langle b + A^T x_k^{(2)}, u - x_k^{(1)} \rangle_1 - \frac{2}{\sigma_1 \alpha^2} \rho_1(x_k^{(1)}, u) : \ u \in Q_1 \}, \\ \\ \gamma \hat{y}_k^{(2)} \equiv y_k^{(2)} & = & \arg\max\limits_v \{ -\langle c - A x_k^{(1)}, v - -x_k^{(2)} \rangle_2 - \frac{2}{\sigma_2 \gamma^2} \rho_2(x_k^{(2)}, v) : \ v \in Q_2 \}. \end{array}$$

Заметим, что

$$s_k^{(1)} = -\sum_{i=0}^{k} [b + A^T y_i^{(2)}], \quad s_k^{(2)} = -\sum_{i=0}^{k} [c - A y_i^{(1)}].$$

Запишем вышеприведенные правила в алгоритмической форме.

Метод решения билинейных матричных игр

Инициализация Выбираем $\bar{x} = (\bar{u}, \bar{v}) \in \tilde{Q}_1 \times \tilde{Q}_2$. Определяем α и γ из (3.1.46). Полагаем $s_{-1} = 0 \in E^*$.

Итерация $(k \geq 0)$ Вычисляем

$$\begin{array}{rcl} x_k^{(1)} & = & \arg\max\limits_{u \in Q_1} \left\{ \langle s_{k-1}^{(1)}, u - \bar{u} \rangle_1 - \frac{2}{\sigma_1 \alpha^2} \rho_1(\bar{u}, u) \right\}, \\ x_k^{(2)} & = & \arg\max\limits_{v \in Q_2} \left\{ \langle s_{k-1}^{(2)}, v - \bar{v} \rangle_2 - \frac{2}{\sigma_2 \gamma^2} \rho_2(\bar{v}, v) \right\}, \\ \\ y_k^{(1)} & = & \arg\max\limits_{u \in Q_1} \left\{ -\langle b + A^T x_k^{(2)}, u - x_k^{(1)} \rangle_1 - \frac{2}{\sigma_1 \alpha^2} \rho_1(x_k^{(1)}, u) \right\}, \\ y_k^{(2)} & = & \arg\max\limits_{v \in Q_2} \left\{ -\langle c - A x_k^{(1)}, v - x_k^{(2)} \rangle_2 - \frac{2}{\sigma_2 \gamma^2} \rho_2(x_k^{(2)}, v) \right\}, \\ \\ s_k^{(1)} & = & s_{k-1}^{(1)} - [b + A^T y_k^{(2)}], \quad s_k^{(2)} \ = \ s_{k-1}^{(2)} - [c - A y_k^{(1)}]. \end{array} \qquad (3.1.48)$$

Результат $\tilde{y}_k = \frac{1}{k+1} \sum\limits_{i=0}^{k} y_i$.

Теорема 3.1.6 *Пусть последовательность $\{\hat{y}_k\}$ формируется методом (3.1.48). Тогда*

$$F_{D_2}(\tilde{y}_k^{(1)}) - \Phi_{D_1}(\tilde{y}_k^{(2)}) \leq \frac{4L}{k+1} \left[\frac{D_1 D_2}{\sigma_1 \sigma_2} \right]^{1/2}. \qquad (3.1.49)$$

Доказательство.
Мы уже видели, что последовательности $\{x_k\}$, $\{y_k\}$ и $\{s_k\}$, сформированные методом (3.1.48), соответствуют последовательностям $\{\hat{x}\}$, $\{\hat{y}\}$ и $\{\hat{s}\}$, построенным методом (3.1.20) для отшкалированной задачи (3.1.45). В силу теоремы 3.1.5 и соотношения (3.1.46) имеем

$$\tfrac{4L}{k+1}\left[\tfrac{D_1D_2}{\sigma_1\sigma_2}\right]^{1/2} = \tfrac{4}{k+1}\max\left\{\tfrac{D_1}{\sigma_1\alpha^2}, \tfrac{D_2}{\sigma_2\gamma^2}\right\} \geq \hat{F}_{D_2}\left(\hat{\tilde{y}}_k^{(1)}\right) - \hat{\Phi}_{D_1}\left(\hat{\tilde{y}}_k^{(2)}\right).$$

Остается заметить, что для $u = \alpha\hat{u}$ выполняются равенства

$$\begin{aligned} \hat{F}_{D_2}(\hat{u}) &= \max_{\hat{v}}\{\hat{f}(\hat{u},\hat{v}) :\ \hat{v} \in \hat{Q}_2\} \\ &= \max_{\hat{v}}\{f(\alpha\hat{u},\gamma\hat{v}) :\ \gamma\hat{v} \in Q_2,\ \rho(\alpha\hat{u},\gamma\hat{v}) \leq D_2\} \\ &= \max_{v}\{f(\alpha\hat{u},v) :\ v \in Q_2,\ \rho(\alpha\hat{u},v) \leq D_2\} = F_{D_2}(u). \end{aligned}$$

Аналогично для $v = \gamma\hat{v}$ получаем $\hat{\Phi}_{D_1}(\hat{v}) = \Phi_{D_1}(v)$. $\square$

В завершении этого пункта рассмотрим пример приложения. Пусть $E_1 = \mathbb{R}^n$. Выберем

$$Q_1 = \Delta_n = \{u \in \mathbb{R}^n :\ u \geq 0,\ \sum\limits_{i=1}^n u^{(i)} = 1\}.$$

Тогда для множества Q_1 можно выбрать следующую *энтропийную* прокс-функцию:

$$d_1(u) = \sum\limits_{i=1}^n u^{(i)} \ln u^{(i)}.$$

Обозначим $\|u\|_1 = \sum\limits_{i=1}^n |u^{(i)}|$. Тогда в соответствии с леммой 3 из работы [84] эта функция строго выпукла с параметром $\sigma_1 = 1$. Заметим, что ее двойственная функция есть

$$(d_1)^*_{Q_1}(s) = \max_{u\in\Delta_n}\{\langle s,u\rangle_1 - d_1(u)\} = \ln\left(\sum\limits_{i=1}^n e^{s^{(i)}}\right).$$

Таким образом, для $s \in \mathbb{R}^n$ имеем

$$\nabla(d_1)^*_{Q_1}(s)^{(i)} = e^{s^{(i)}} / \left[\sum\limits_{j=1}^n e^{s^{(j)}}\right], \quad i = 1,\ldots,n.$$

Следовательно, $\tilde{Q}_1 = \{u \in \mathbb{R}^n :\ u > 0,\ \sum\limits_{i=1}^n u^{(i)} = 1\}$, и мы видим, что функция $d_1(u)$ удовлетворяет предположению 3.1.1.

Выберем $\bar{u}$ в центре симплекса Δ_n:

$$\bar{u}^{(i)} = \tfrac{1}{n}, \quad i = 1,\ldots,n.$$

Тогда

$$D_1 = \max_{u \in Q_1} \rho_1(\bar{u}, u) = \ln n.$$

Аналогично если выбрать $E_2 = \mathbb{R}^m$, $Q_2 = \Delta_m$, $\|v\|_2 = \sum\limits_{i=1}^{m} |v^{(i)}|$, $\bar{v}$ в центре симплекса Δ_m и

$$d_2(v) = \sum_{i=1}^{m} v^{(i)} \ln v^{(i)},$$

то получим $\sigma_2 = 1$ и $D_2 = \ln m$.

Таким образом, в соответствии с правилами (3.1.46) мы можем использовать в методе (3.1.48) следующие значения параметров:

$$L = \max_{i,j} |A^{(i,j)}|, \quad \alpha^2 = \tfrac{1}{L}\sqrt{\tfrac{\ln n}{\ln m}}, \quad \gamma^2 = \tfrac{1}{L}\sqrt{\tfrac{\ln m}{\ln n}}.$$

Скорость сходимости метода (3.1.48) в нашей ситуации оценивается следующим образом:

$$F_{D_2}(\tilde{y}_k^{(1)}) - \Phi_{D_1}(\tilde{y}_k^{(2)}) \le 4 \max_{i,j} |A^{(i,j)}| \cdot \tfrac{\sqrt{\ln n \ln m}}{k+1}. \tag{3.1.50}$$

Каждая итерация в схеме требует четырех матрично-векторных умножений с матрицей A.

Наконец заметим, что для точки $z \in \tilde{Q}_1$ решение $T_\mu(z, s)$ задачи

$$\max_{u \in \Delta_n} \{\langle s, u - z\rangle_1 - \mu \rho_1(z, u)\}$$

может быть найдено в явном виде:

$$T_\mu(z, s)^{(i)} = z^{(i)} e^{s^{(i)}/\mu} \left[\sum_{j=1}^{n} z^{(j)} e^{s^{(j)}/\mu}\right]^{-1}, \quad j = 1, \ldots, n.$$

3.1.6 Обсуждение

Рассмотрим двойственный экстраполяционный шаг

$$\begin{aligned} x &= T_\beta(\bar{x}, s), \\ y &= T_\beta(x, -\lambda\, g(x)), \\ s_+ &= s - \lambda\, g(y), \end{aligned} \tag{3.1.51}$$

в евклидовой постановке: $Q \subseteq E \equiv \mathbb{R}^n$, $\|x\| = \left[\sum\limits_{i=1}^{n} (x^{(i)})^2\right]^{1/2}$, и $d(x) = \frac{1}{2}\|x\|^2$. Тогда $\sigma = 1$ и $\rho(x, y) = \frac{1}{2}\|y - x\|^2$. Обозначим

$$\pi_Q(x) = \arg\min_y \{\|y - x\| : \ y \in Q\}.$$

Заметим что $E^* \equiv \mathbb{R}^n \equiv E$. Таким образом,

$$T_\beta(z,s) = \arg\max_x \{\langle s, x-z\rangle - \tfrac{1}{2}\beta\|x-z\|^2 : \ x \in Q\} = \pi_Q(z + \tfrac{1}{\beta}s).$$

Следовательно, в нашей ситуации двойственный экстраполяционный шаг может быть записан следующим образом:

$$\begin{aligned} x &= \pi_Q(\bar{x} + \tfrac{1}{\beta}s), \\ y &= \pi_Q(x - \tfrac{\lambda}{\beta}g(x)), \\ s_+ &= s - \lambda\, g(y). \end{aligned} \tag{3.1.52}$$

Сравним этот процесс со стандартным прямым экстраградиентным методом (см., например, правило (23) в работе [78]):

$$\begin{aligned} y &= \pi_Q(x - \gamma g(x)), \\ x_+ &= \pi_Q(x - \gamma g(y)). \end{aligned} \tag{3.1.53}$$

Мы видим, что в формулах (3.1.53) основной пересчет $x \to x_+$ осуществляется в прямом пространстве. Однако в схеме (3.1.52) основным шагом является пересчет $s \to s_+$, который происходит в двойственном пространстве. Даже в евклидовой постановке методы (3.1.52) и (3.1.53) формируют разные последовательности тестовых точек за исключением случая $Q \equiv \mathbb{R}^n$. Основное различие между этими методами заключается в том, что двойственный метод полностью базируется на линейной модели задачи (что есть линейный функционал s_k). Эта модель используется для формирования тестовых точек и в критерии остновки, который гарантирует нужную точность. В прямом методе (3.1.53) модель не сохраняется; она разрушается при проектировании.

3.2 Сильно монотонные операторы и квазивариационные неравенства

3.2.1 Введение

Мотивировка

В пункте 2.1.5 и параграфе 3.1 мы уже рассматривали вычислительные методы для решения вариационных неравенств (ВН), операторы которых обладают разной степенью гладкости. В этом параграфе мы покажем как решать ВН с сильно монотонным оператором. Пользуясь этими результатами, мы покажем как можно решать *квазивариационные неравенства* (КВН).

Содержание

В п. 3.2.2 мы рассматриваем стандартные ВН с липшицевым сильно монотонным оператором. Для обоснования скорости сходимости численных методов мы вводим новую *строго выпуклую* функцию близости. Для уменьшения значения этой функции мы рекурсивно пересчитываем простую квадратичную модель нашей задачи. Наш подход комбинирует идеи двойственной экстраполяции [87] с элементами оценочных функций (см. раздел 2.2 в монографии [83]). В результате мы получаем простой метод, который существенно превосходит по своей эффективности обычный градиентный метод.

В п. 3.2.3 мы рассматриваем задачу решения квазивариационного неравенства и приводим основные результаты о существовании его решения [103]. Для полноты изложения, мы приводим также анализ эффективности простейшего градиентного метода. В п. 3.2.4 мы применяем результаты п. 3.2.2 для решения КВН. Для этого мы вводим *оператор релаксации* и приводим достаточные условия для того чтобы он был оператором сжатия. Эти условия гарантируют существование единственного решения у КВН для гораздо более широкого класса задач, чем было известно ранее [103]. С другой стороны, мы показываем, что метод простой итерации с приближенно вычисленным значением оператора релаксации существенно превосходит по своей эффективности обычный градиентный метод. Такие значения оператора релаксации могут быть вычислены с помощью метода, описанного в разделе 3.2.2.

Обозначения

Пусть положительно определенный оператор $B : E \to E^*$ является самосопряженным. С помощью этого оператора введем евклидовы нормы на E и E^*:

$$\begin{aligned} \|x\| &= \langle Bx, x\rangle^{1/2}, \quad x \in E, \\ \|s\|_* &= \max_{x\in E}\{\langle s, x\rangle : \|x\| \le 1\} \\ &= \langle s, B^{-1}s\rangle^{1/2}, \quad s \in E^*. \end{aligned}$$

Обозначим через $\pi_Q(x)$ евклидову проекцию точки x на множество Q. Необходимые и достаточные характеристики проекции выглядят так:

$$\begin{aligned} \pi_Q(x) &\in Q, \\ \langle B(\pi_Q(x) - x), y - \pi_Q(x)\rangle &\ge 0 \quad \forall y \in Q. \end{aligned} \tag{3.2.1}$$

3.2.2 Решение сильно монотонных вариационных неравенств

Пусть множество Q является выпуклым и замкнутым. Рассмотрим непрерывный оператор $g(x): \ Q \to E^*$, который является *сильно монотонным*:

$$\langle g(x) - g(y), x - y \rangle \ \ \geq \ \ \mu \, \|x - y\|^2 \quad \forall x, y \in Q. \tag{3.2.2}$$

Константа $\mu \geq 0$ называется *параметром сильной монотонности* оператора g. Если $\mu = 0$, то оператор g монотонен. В дальнейшем мы всегда предполагаем, что $\mu > 0$.

В этом пункте мы будем искать решение следующего *вариационного неравенства*:

$$\text{найти } x^*(Q) \in Q \ : \quad \langle g(x^*(Q)), y - x^*(Q) \rangle \geq 0 \quad \forall y \in Q. \tag{3.2.3}$$

При отсутствии двусмысленности мы часто пользуемся обозначением $x^* \equiv x^*(Q)$. Поскольку оператор g сильно монотонен, решение x^* задачи (3.2.3) удовлетворяет неравенствам

$$\langle g(y), x^* - y \rangle + \tfrac{1}{2}\mu \|y - x^*\|^2 \ \overset{(3.2.2)}{\leq} \ \langle g(x^*), x^* - y \rangle - \tfrac{1}{2}\mu \|y - x^*\|^2 \ \overset{(3.2.3)}{\leq} \ 0, \tag{3.2.4}$$

которые справедливы при всех $y \in Q$. Нетрудно показать, что решение такого ВН всегда существует. Понятно, что оно может быть только единственным.

Для описания качества приближенного решения задачи (3.2.3) нам потребуется следующая *мера близости*:

$$f(x) = \sup_{y \in Q} \left\{ \langle g(y), x - y \rangle + \tfrac{1}{2}\mu \|y - x\|^2 \right\}. \tag{3.2.5}$$

Теорема 3.2.1 *Мера близости $f(x)$ определена и сильно выпукла на Q с параметром выпуклости μ. Более того, она неотрицательна на Q и обращается в нуль только в единственном решении вариационного неравенства (3.2.3).*

Доказательство.
Действительно, из сильной монотонности оператора g следует, что

$$\begin{aligned} \langle g(y), x - y \rangle + \tfrac{1}{2}\mu \|y - x\|^2 \ &\overset{(3.2.2)}{\leq} \ \langle g(x), x - y \rangle - \tfrac{1}{2}\mu \|y - x\|^2 \\ &\leq \ \|g(x)\|_* \cdot \|x - y\| - \tfrac{1}{2}\mu \|y - x\|^2 \\ &\leq \ \tfrac{1}{2\mu} \|g(x)\|_*^2 \end{aligned}$$

для любых точек $x, y \in Q$. Таким образом, значение $f(x)$ корректно определено. Далее, функция $f(x)$ сильно выпукла по x с параметром выпуклости μ, поскольку она задана как максимум параметрического семейства сильно выпуклых (по x) квадратичных функций:

$$f(x) \ = \ \max_{y \in Q} \phi_y(x), \quad \phi_y(x) \ = \ \langle g(y), x - y \rangle + \tfrac{1}{2}\mu \|y - x\|^2, \ y \in Q.$$

Наконец, если $x \in Q$, то ясно, что $f(x) \geq 0$. Рассмотрим теперь x^* – решение задачи (3.2.3). В силу неравенств (3.2.4) имеем $f(x^*) = 0$. С другой стороны, если $f(\hat{x}) = 0$ для некоторой точки $\hat{x} \in Q$, то эта точка является решением слабого вариационного неравенства

$$\langle g(y), y - \hat{x} \rangle \geq 0 \quad \forall y \in Q.$$

Поскольку оператор $g(y)$ непрерывен, это означает также, что $\hat{x}$ является решением задачи (3.2.3). $\Box$

Покажем, как можно генерировать точки с малым значением функции близости $f(\cdot)$. Рассмотрим произвольную последовательность точек $\{y_i\}_{i=0}^N \subset Q$ и последовательность положительных весов $\{\lambda_i\}_{i=0}^N$. Обозначим

$$S_N \;=\; \sum_{i=0}^N \lambda_i, \qquad \tilde{y}_N \;=\; \tfrac{1}{S_N} \sum_{i=0}^N \lambda_i y_i,$$

По сравнению с формулами (3.1.4), нам потребуется другое определение функции зазора:

$$\Delta_N \;=\; \max_{x \in Q} \left\{ \sum_{i=0}^N \lambda_i \left[\langle g(y_i), y_i - x \rangle - \tfrac{1}{2}\mu \|x - y_i\|^2 \right] \right\}. \tag{3.2.6}$$

Заметим, что вычисление величины Δ_N сводится к нахождению евклидовой проекции на Q.

Лемма 3.2.1 *Справедливо неравенство*

$$f(\tilde{y}_N) \leq \tfrac{1}{S_N} \Delta_N. \tag{3.2.7}$$

Доказательство.

Действительно, поскольку оператор $g(x)$ сильно монотонен, имеем

$$\begin{aligned} f(\tilde{y}_N) \quad &= \quad \sup_{x \in Q} \left\{ \langle g(x), \tilde{y}_N - x \rangle + \tfrac{1}{2}\mu \|x - \tilde{y}_N\|^2 \right\} \\ &= \quad \tfrac{1}{S_N} \sup_{x \in Q} \left\{ \sum_{i=0}^N \lambda_i \langle g(x), y_i - x \rangle + \tfrac{1}{2}\mu S_N \|x - \tilde{y}_N\|^2 \right\} \\ &\leq \quad \tfrac{1}{S_N} \sup_{x \in Q} \left\{ \sum_{i=0}^N \lambda_i \langle g(x), y_i - x \rangle + \tfrac{1}{2}\mu \sum_{i=0}^N \lambda_i \|x - y_i\|^2 \right\} \\ &\overset{(3.2.2)}{\leq} \quad \tfrac{1}{S_N} \max_{x \in Q} \left\{ \sum_{i=0}^N \lambda_i \left[\langle g(y_i), y_i - x \rangle - \tfrac{1}{2}\mu \|x - y_i\|^2 \right] \right\} \;\equiv\; \tfrac{1}{S_N} \Delta_N. \qquad \Box \end{aligned}$$

Таким образом, нашей целью является контролирование скорости роста величины Δ_N по сравнению со скоростью роста суммы S_N. Для $\beta > 0$ обозначим

$$\begin{aligned} \psi_y^\beta(x) \;&=\; \langle g(y), y - x \rangle - \tfrac{1}{2}\beta \|x - y\|^2, \\ \Psi_k(x) \;&=\; \sum_{i=0}^k \lambda_i \psi_{y_i}^\mu(x). \end{aligned}$$

Заметим, что квадратичная функция $\psi_y^\beta(x)$ сильно вогнута с параметром β, а функция $\Psi_k(x)$ сильно вогнута с параметром μS_k. Ясно, что $\Delta_k = \max\limits_{x\in Q} \Psi_k(x)$.

Рассмотрим следующую итерацию:

$$\begin{aligned} x_k &= \arg\max_{x\in Q} \Psi_k(x), \\ y_{k+1} &= \arg\max_{x\in Q} \psi_{x_k}^\beta(x). \end{aligned} \tag{3.2.8}$$

Теорема 3.2.2 *Если* $\lambda_{k+1} \le \frac{\mu}{\beta} S_k$, *то*

$$\Delta_{k+1} \;\le\; \Delta_k + \tfrac{1}{2}\lambda_{k+1}\left[\tfrac{1}{\mu+\beta}\|g(y_{k+1}) - g(x_k)\|_*^2 - \beta\|y_{k+1} - x_k\|^2\right]. \tag{3.2.9}$$

Доказательство.

Заметим, что $\Psi_{k+1}(x) = \Psi_k(x) + \lambda_{k+1}\psi_{y_{k+1}}^\mu(x)$. Поэтому

$$\begin{aligned} \Delta_{k+1} &= \max_{x\in Q}\left\{\Psi_k(x) + \lambda_{k+1}\psi_{y_{k+1}}^\mu(x)\right\} \\ &\le \Delta_k + \max_{x\in Q}\left\{\langle \nabla\Psi_k(x_k), x - x_k\rangle - \tfrac{1}{2}\mu S_k\|x - x_k\|^2 + \lambda_{k+1}\psi_{y_{k+1}}^\mu(x)\right\} \\ &\le \Delta_k + \max_{x\in Q}\left\{-\tfrac{1}{2}\mu S_k\|x - x_k\|^2 + \lambda_{k+1}\left[\langle g(y_{k+1}), y_{k+1} - x\rangle - \tfrac{1}{2}\mu\|x - y_{k+1}\|^2\right]\right\}. \end{aligned}$$

В силу определения точки y_{k+1} имеем

$$\langle -g(x_k) - \beta B(y_{k+1} - x_k), x - y_{k+1}\rangle \;\le\; 0 \quad \forall x \in Q.$$

Таким образом,

$$\begin{aligned} &\langle g(y_{k+1}), y_{k+1} - x\rangle - \tfrac{1}{2}\mu\|x - y_{k+1}\|^2 \\ &= \langle g(y_{k+1}) - g(x_k), y_{k+1} - x\rangle - \tfrac{1}{2}\mu\|x - y_{k+1}\|^2 + \langle g(x_k), y_{k+1} - x\rangle \\ &\le \|g(y_{k+1}) - g(x_k)\|_* \cdot \|y_{k+1} - x\| - \tfrac{1}{2}\mu\|x - y_{k+1}\|^2 + \beta\langle B(y_{k+1} - x_k), x - y_{k+1}\rangle. \end{aligned}$$

Заметим, что

$$2\langle B(y_{k+1} - x_k), x - y_{k+1}\rangle \;=\; \|x - x_k\|^2 - \|y_{k+1} - x_k\|^2 - \|x - y_{k+1}\|^2.$$

Поэтому

$$\begin{aligned} &\langle g(y_{k+1}), y_{k+1} - x\rangle - \tfrac{1}{2}\mu\|x - y_{k+1}\|^2 \\ &\le \tfrac{1}{2(\mu+\beta)}\|g(y_{k+1}) - g(x_k)\|_*^2 + \tfrac{1}{2}\beta\|x - x_k\|^2 - \tfrac{1}{2}\beta\|y_{k+1} - x_k\|^2. \end{aligned}$$

Учитывая все эти неравенства и пользуясь верхней оценкой для λ_{k+1}, мы приходим к неравенству (3.2.9). □

Следствие 3.2.1 *Предположим, что оператор $g(x)$ липшицев на Q:*

$$\|g(x) - g(y)\|_* \ \leq \ L\|x - y\|, \quad \forall x, y \in Q. \tag{3.2.10}$$

Тогда при

$$\beta \ = \ L, \quad \lambda_{k+1} \ = \ \tfrac{\mu}{L} S_k,$$

выполняется неравенство $\Delta_{k+1} \leq \Delta_k$.

Теперь мы готовы выписать метод решения вариационного неравенства (3.2.3) с сильно монотонным липшицевым оператором. Этот метод представляет собой комбинацию двойственного экстраполяционного метода [87] с двойственным методом, построенным с помощью оценочных функций (см. раздел 2.2 в [83]).

Для простоты предположим, что константы μ и L нам известны. Обозначим через $\gamma = \frac{L}{\mu} \geq 1$ *число обусловленности* оператора g.

Метод для сильно монотонных ВН

Инициализация Выбираем $\bar{x} \in Q$. Полагаем $\lambda_0 = 1$ и $y_0 = \bar{x}$.

Итерация **($k \geq 0$):** Вычисляем

$$\begin{array}{rcl} x_k & = & \arg\max\limits_{x \in Q} \Psi_k(x), \\ y_{k+1} & = & \arg\max\limits_{x \in Q} \psi^L_{x_k}(x), \\ \lambda_{k+1} & = & \frac{1}{\gamma} \cdot S_k. \end{array} \tag{3.2.11}$$

Результат $\tilde{y}_k = \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i y_i$.

Теорема 3.2.3 *При условиях (3.2.2) и (3.2.10), для любого $k \geq 0$ выполняется неравенство*

$$\begin{array}{rcl} \frac{\mu}{2} \cdot \|\tilde{y}_k - x^*\|^2 \ \leq \ f(\tilde{y}_k) & \leq & \left[f(\bar{x}) + \frac{\mu \cdot (\gamma^2 - 1)}{2} \cdot \|\bar{x} - x^*\|^2 \right] \cdot \exp\left\{ -\frac{k}{\gamma+1} \right\} \\ & \leq & f(\bar{x}) \cdot \gamma^2 \cdot \exp\left\{ -\frac{k}{\gamma+1} \right\}. \end{array} \tag{3.2.12}$$

Доказательство.

Первое и последнее из неравенств (3.2.12) следуют из сильной выпуклости функции $f(x)$ с параметром μ. Докажем среднее неравенство.

В силу следствия 3.2.1 имеем $\Delta_{k+1} \le \Delta_k$ при всех $k \ge 0$. Заметим, что $S_0 = \lambda_0 = 1$, и

$$S_{k+1} = S_k + \lambda_{k+1} = \left(1 + \tfrac{1}{\gamma}\right) S_k.$$

Поэтому

$$\begin{aligned} f(\tilde{y}_k) &\overset{(3.2.7)}{\le} \tfrac{\Delta_k}{S_k} \le \Delta_0 \cdot \left(1 - \tfrac{1}{\gamma+1}\right)^k \\ &\le \Delta_0 \cdot \exp\left\{-\tfrac{k}{\gamma+1}\right\}. \end{aligned}$$

Осталось оценить величину Δ_0. Заметим, что

$$\begin{aligned} \Delta_0 &= \max_{x\in Q} \Psi_0(x) \\ &= \max_{x\in Q} \psi^{\mu}_{y_0}(x) \\ &= \max_{x\in Q} \left\{\langle g(\bar{x}), \bar{x} - x\rangle - \tfrac{\mu}{2} \cdot \|x - \bar{x}\|^2\right\} \\ &= \max_{x\in Q} \left\{\langle g(\bar{x}) - g(x^*), \bar{x} - x\rangle + \langle g(x^*), \bar{x} - x\rangle - \tfrac{\mu}{2} \cdot \|x - \bar{x}\|^2\right\} \\ &\overset{(3.2.3)}{\le} \langle g(x^*), \bar{x} - x^*\rangle + \max_{x\in Q} \left\{\langle g(\bar{x}) - g(x^*), \bar{x} - x\rangle - \tfrac{\mu}{2} \cdot \|x - \bar{x}\|^2\right\} \\ &\overset{(3.2.10)}{\le} \langle g(x^*), \bar{x} - x^*\rangle + \tfrac{L^2}{2\mu} \cdot \|\bar{x} - x^*\|^2. \end{aligned}$$

Поскольку $\langle g(x^*), \bar{x} - x^*\rangle + \frac{\mu}{2} \cdot \|x^* - \bar{x}\|^2 \overset{(3.2.5)}{\le} f(\bar{x})$, получаем второе из неравенство (3.2.12). □

Заметим, что задача (3.2.3) может быть решена стандартным методом градиентного типа:

$$\begin{aligned} x_0 &= \bar{x} \in Q, \\ x_{k+1} &= \pi_Q(x_k - \lambda B^{-1} g(x_k)), \quad k \ge 0. \end{aligned} \tag{3.2.13}$$

Однако хорошо известно, что у этого метода медленная сходимость. Приведем соответствующую оценку. Поскольку $x^* \overset{(3.2.3)}{=} \pi_Q(x^* - \lambda B^{-1} g(x^*))$, выбирая оптимальный шаг

$\lambda = \frac{\mu}{L^2}$, получаем

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &\leq \|x_k - \lambda B^{-1} g(x_k) - (x^* - \lambda B^{-1} g(x^*))\|^2 \\
&= \|x_k - x^*\|^2 - 2\lambda \langle g(x_k) - g(x^*), x_k - x^* \rangle + \lambda^2 \|g(x_k) - g(x^*)\|_*^2 \\
&\overset{(3.2.2),(3.2.10)}{\leq} (1 - 2\lambda\mu + \lambda^2 L^2) \cdot \|x_k - x^*\|^2 \\
&\leq \|x_k - x^*\|^2 \cdot \exp\left\{ -\frac{k}{\gamma^2} \right\}.
\end{aligned}$$

При больших значениях числа обусловленности эта оценка гораздо хуже чем (3.2.12). Заметим, что скорость сходимости (3.2.12) не может быть улучшена никаким черноящичным методом, применяемым к задаче (3.2.2), (3.2.10) (см. [79]). В то же время с точки зрения сложности реализации метод (3.2.11) сравним с (3.2.13): на каждой итерации требуется вычислить две проекции на множество и два значения оператора вместо одной проекции и одного значения в методе (3.2.13).

3.2.3 Квазивариационные неравенства

Пусть точечно-множественное отображение $\mathcal{Q} : E \to 2^E$ принимает выпуклые и замкнутые значения. Рассмотрим следующую задачу *квазивараиционного неравенства* (КВН):

$$\text{найти } x_* \in \mathcal{Q}(x_*) : \quad \langle g(x_*), y - x_* \rangle \geq 0 \quad \forall y \in \mathcal{Q}(x_*). \tag{3.2.14}$$

Известен следующий результат о существовании его решения (см. теорема 9 в работе [103]).

Теорема 3.2.4 *Предположим, что выполнены следующие предположения:*

(a) оператор g липшицев с константой L и сильно монотонен на E с параметром $\mu > 0$;

(b) существует такое $\alpha < \frac{1}{\gamma(\gamma + \sqrt{\gamma^2 - 1})}$, что

$$\|\pi_{\mathcal{Q}(x)}(z) - \pi_{\mathcal{Q}(y)}(z)\| \leq \alpha \|x - y\| \quad \forall x, y, z \in E. \tag{3.2.15}$$

Тогда задача (3.2.14) имеет единственное решение.

Далее мы получим этот результат как следствие теоремы 3.2.5, обосновывающей скорость сходимости градиентного метода, применяемого к задаче (3.2.14). Пока только отметим, что предположение (3.2.15) является усилением свойства сжимаемости для точечно-множественного отображения $\mathcal{Q}(x)$. Приведем пример такого отображения.

Лемма 3.2.2 *Пусть функция $c(x) : E \to E$ является липшицевой:*

$$\|c(x) - c(y)\| \le \alpha \|x - y\|, \quad x, y \in E,$$

и пусть множество $\bar{Q}$ выпукло и замкнуто. Тогда отображение сдвига

$$\mathcal{Q}(x) \overset{\mathrm{def}}{=} c(x) + \bar{Q}$$

удовлетворяет условию (3.2.15) с тем же значением α.

Доказательство.

Действительно, при любых $x, z \in E$ имеем

$$\pi_{c(x)+\bar{Q}}(z) = c(x) + \pi_{\bar{Q}}(z - c(x)).$$

Обозначим $z_1 = z - c(x)$, $z_2 = z - c(y)$. Поскольку проекция берется в евклидовой норме, имеем

$$\begin{array}{rl} & \|\pi_{c(x)+\bar{Q}}(z) - \pi_{c(y)+\bar{Q}}(z)\|^2 \\ \\ = & \|c(x) + \pi_{\bar{Q}}(z - c(x)) - c(y) - \pi_{\bar{Q}}(z - c(y))\|^2 \\ \\ = & \|z_2 - \pi_{\bar{Q}}(z_2) - z_1 + \pi_{\bar{Q}}(z_1)\|^2 \\ \\ = & \|z_2 - z_1\|^2 - 2\langle B(z_2 - z_1), \pi_{\bar{Q}}(z_2) - \pi_{\bar{Q}}(z_1)\rangle + \|\pi_{\bar{Q}}(z_2) - \pi_{\bar{Q}}(z_1)\|^2. \end{array}$$

Заметим, что

$$\begin{array}{rl} & \langle B(z_2 - z_1), \pi_{\bar{Q}}(z_2) - \pi_{\bar{Q}}(z_1)\rangle \\ \\ = & \langle B(z_2 - \pi_{\bar{Q}}(z_2) + \pi_{\bar{Q}}(z_2) - z_1), \pi_{\bar{Q}}(z_2) - \pi_{\bar{Q}}(z_1)\rangle \\ \\ \overset{(3.2.1)}{\ge} & \langle B(\pi_{\bar{Q}}(z_2) - \pi_{\bar{Q}}(z_1) + \pi_{\bar{Q}}(z_1) - z_1), \pi_{\bar{Q}}(z_2) - \pi_{\bar{Q}}(z_1)\rangle \\ \\ \overset{(3.2.1)}{\ge} & \|\pi_{\bar{Q}}(z_2) - \pi_{\bar{Q}}(z_1)\|^2. \end{array}$$

Поэтому $\|\pi_{c(x)+\bar{Q}}(z) - \pi_{c(y)+\bar{Q}}(z)\|^2 \le \|z_2 - z_1\|^2 \le \alpha^2 \|x - y\|^2$. □

Заметим, что задача (3.2.14) может решаться стандартным методом градиентного типа:

$$x_{k+1} = \pi_{\mathcal{Q}(x_k)}(x_k - \lambda B^{-1} g(x_k)), \quad k \ge 0. \tag{3.2.16}$$

Оценим его скорость сходимости:

$$\begin{aligned}
\|x_{k+1} - x_*\| &= \|\pi_{\mathcal{Q}(x_k)}(x_k - \lambda B^{-1} g(x_k)) - \pi_{\mathcal{Q}(x_*)}(x_* - \lambda B^{-1} g(x_*))\| \\
&= \|\pi_{\mathcal{Q}(x_k)}(x_k - \lambda B^{-1} g(x_k)) - \pi_{\mathcal{Q}(x_*)}(x_k - \lambda B^{-1} g(x_k)) \\
&\quad + \pi_{\mathcal{Q}(x_*)}(x_k - \lambda B^{-1} g(x_k)) - \pi_{\mathcal{Q}(x_*)}(x_* - \lambda B^{-1} g(x_*))\| \\
&\overset{(3.2.15)}{\leq} \alpha \|x_k - x_*\| + \|\pi_{\mathcal{Q}(x_*)}(x_k - \lambda B^{-1} g(x_k)) - \pi_{\mathcal{Q}(x_*)}(x_* - \lambda B^{-1} g(x_*))\|.
\end{aligned}$$

Поскольку оператор g сильно монотонен и липшицев, имеем

$$\begin{aligned}
&\|\pi_{\mathcal{Q}(x_*)}(x_k - \lambda B^{-1} g(x_k)) - \pi_{\mathcal{Q}(x_*)}(x_* - \lambda B^{-1} g(x_*))\|^2 \\
\leq\ & \|(x_k - \lambda B^{-1} g(x_k)) - (x_* - \lambda B^{-1} g(x^*))\|^2 \\
=\ & \|x_k - x_*\|^2 - 2\lambda \langle g(x_k) - g(x_*), x_k - x^* \rangle + \lambda^2 \|g(x_k) - g(x_*\|^2 \\
\leq\ & (1 - 2\lambda\mu + \lambda^2 L^2) \|x_k - x_*\|^2.
\end{aligned}$$

Таким образом, мы доказали следующую теорему.

Теорема 3.2.5 *Если оператор g сильно монотонен и липшицев с константами L и μ, а точечно-множественное отображение $\mathcal{Q}(x)$ удовлетворяет условию (3.2.15) с $\alpha < \frac{1}{\gamma(\gamma+\sqrt{\gamma^2-1})}$, то градиентный метод (3.2.16) с оптимальным шагом $\lambda = \frac{\mu}{L^2}$ сходится к единственному решению задачи (3.2.14) со следующей скоростью:*

$$\|x_k - x_*\| \leq \exp\left\{-k \cdot \left(\tfrac{1}{\gamma(\gamma+\sqrt{\gamma^2-1})} - \alpha\right)\right\} \cdot \|x_0 - x_*\|. \tag{3.2.17}$$

Мы показали, что квазивариационное неравенство (3.2.14) может быть решено градиентным методом (3.2.16) только в том случае, когда скорость изменения допустимого множества $\mathcal{Q}(x)$ достаточно мала по сравнению с обратным числом обусловленности оператора g:

$$\alpha < \tfrac{1}{\gamma(\gamma+\sqrt{\gamma^2-1})} \approx \tfrac{1}{2\gamma^2}. \tag{3.2.18}$$

В следующем пункте мы покажем, что это ограничение может быть существенно ослаблено.

3.2.4 Оператор релаксации для квазивариационного неравенства

Для задачи (3.2.14) нам потребуется *оператор релаксации* $T(x) \overset{\text{def}}{=} x^*(\mathcal{Q}(x))$. В случае сильно монотонного оператора g оператор $T(x)$ полностью определен следующими соотношениями:

$$\begin{aligned} T(x) &\in \mathcal{Q}(x), \\ \langle g(T(x)), y - T(x)\rangle &\geq 0 \quad \forall y \in \mathcal{Q}(x). \end{aligned} \tag{3.2.19}$$

Ясно, что решение задачи (3.2.14) является неподвижной точкой этого оператора:

$$x_* = T(x_*).$$

Оказывается, при стандартных предположениях оператор $T(x)$ липшицев.

Теорема 3.2.6 *Предположим что оператор g липшицев и сильно монотонен с константами L и $\mu > 0$ соответственно. Пусть существует такое $\alpha \geq 0$, что*

$$\|\pi_{\mathcal{Q}(x)}(z) - \pi_{\mathcal{Q}(y)}(z)\| \;\leq\; \alpha\|x - y\| \quad \forall x, y, z \in E. \tag{3.2.20}$$

Тогда оператор $T(x)$ удовлетворяет условию Липшица с константой $\alpha\gamma$, где $\gamma = L/\mu$.

Доказательство.

Зафиксируем две произвольные точки $x_1, x_2 \in E$. Обозначим

$$T_i = T(x_i), \quad g_i = g(T_i), \quad Q_i = \mathcal{Q}(x_i), \quad i = 1, 2.$$

Зафиксируем некоторое $\lambda > 0$. По определению

$$T_1 \overset{(3.2.1),(3.2.19)}{=} \pi_{Q_1}(T_1 - \lambda B^{-1} g_1).$$

Обозначим $y_2 = \pi_{Q_2}(T_1 - \lambda B^{-1} g_1)$. В силу условия (3.2.20) имеем

$$\|y_2 - T_1\| \;\leq\; \alpha\|x_1 - x_2\|. \tag{3.2.21}$$

С другой стороны, $\langle B(y_2 - (T_1 - \lambda B^{-1} g_1)), T_2 - y_2\rangle \overset{(3.2.1)}{\geq} 0$. Таким образом,

$$\begin{aligned} \langle B(y_2 - T_1), T_2 - y_2\rangle &\geq \lambda\langle g_1, y_2 - T_2\rangle \\ &= \lambda\langle g_1, y_2 - T_1\rangle + \lambda\langle g_1, T_1 - T_2\rangle \\ &\overset{(3.2.2)}{\geq} \lambda\langle g_1, y_2 - T_1\rangle + \lambda\langle g_2, T_1 - T_2\rangle + \lambda\mu\|T_1 - T_2\|^2 \\ &\overset{(3.2.19)}{\geq} \lambda\langle g_1 - g_2, y_2 - T_1\rangle + \lambda\mu\|T_1 - T_2\|^2. \end{aligned}$$

Поэтому

$$\begin{aligned} \lambda\mu\|T_1 - T_2\|^2 &\leq \lambda\langle g_1 - g_2, T_1 - y_2\rangle + \langle B(y_2 - T_1), T_2 - y_2\rangle \\ &\leq \lambda\langle g_1 - g_2, T_1 - y_2\rangle + \langle B(y_2 - T_1), T_2 - T_1\rangle \\ &\overset{(3.2.10)}{\leq} (1 + \lambda L) \cdot \|T_1 - T_2\| \cdot \|y_2 - T_1\|. \end{aligned}$$

Поскольку величина λ может быть сколь угодно большой, из неравенства (3.2.21) получаем $\|T_1 - T_2\| \le \frac{\alpha L}{\mu}\|x_1 - x_2\|$. □

В качестве простого следствия мы получили следующее условие существования решения.

Следствие 3.2.2 *Если $\alpha < \gamma^{-1}$, то существует единственное решение квазивариационного неравенства (3.2.14).*

Доказательство.
Действительно, из условия следует, что оператор $T(x)$ сжимающий. □

Заметим, что последнее утверждение существенно усиливает результат теоремы 3.2.4. Более того, соединив его с результатами п. 3.2.2, можно получить эффективный численный метод решения задачи (3.2.14). Начнем с одного вспомогательного утверждения. В последующем мы всегда предполагаем, что оператор g липшицев и монотонен с константами L и μ соответственно.

Лемма 3.2.3 *Пусть точка $x^*(Q)$ является решением вариационного неравенства (3.2.3). Для точки $\hat{x} \in Q$ положим $\bar{x} = \arg\max\limits_{y \in Q} \psi^{\beta}_{\hat{x}}(y)$, где $\beta > 0$. Тогда*

$$\mu\|\hat{x} - x^*(Q)\|^2 + \beta\|\hat{x} - \bar{x}\|^2 \;\;\le\;\; (\beta + L)\cdot\|\hat{x} - x^*(Q)\|\cdot\|\hat{x} - \bar{x}\|. \tag{3.2.22}$$

Следовательно,

$$\tfrac{\mu}{\beta+L}\cdot\|\hat{x} - x^*(Q)\| \;\;\le\;\; \|\hat{x} - \bar{x}\| \;\;\le\;\; \tfrac{\beta+L}{\beta}\cdot\|\hat{x} - x^*(Q)\|. \tag{3.2.23}$$

Более того,

$$f(\bar{x}) \;\;\le\;\; \tfrac{\beta^2+L^2}{2\mu}\|\hat{x} - \bar{x}\|^2 \;\;\le\;\; \tfrac{(\beta^2+L^2)\cdot(\beta+L)^2}{2\mu\beta^2}\|\hat{x} - x^*(Q)\|^2. \tag{3.2.24}$$

Доказательство.
В силу определения

$$\langle g(\hat{x}) + \beta B(\bar{x} - \hat{x}), y - \bar{x}\rangle \;\;\ge\;\; 0 \quad \forall y \in Q. \tag{3.2.25}$$

Таким образом,

$$\begin{aligned}
\beta\langle B(\bar{x} - \hat{x}), x^*(Q) - \bar{x}\rangle \quad &\ge \quad \langle g(\hat{x}), \bar{x} - x^*(Q)\rangle \\
&= \quad \langle g(\hat{x}), \bar{x} - \hat{x}\rangle + \langle g(\hat{x}), \hat{x} - x^*(Q)\rangle \\
&\overset{(3.2.2)}{\ge} \quad \langle g(\hat{x}), \bar{x} - \hat{x}\rangle + \langle g(x^*(Q)), \hat{x} - x^*(Q)\rangle + \mu\|\hat{x} - x^*(Q)\|^2 \\
&\overset{(3.2.3)}{\ge} \quad \langle g(\hat{x}) - g(x^*(Q)), \bar{x} - \hat{x}\rangle + \mu\|\hat{x} - x^*(Q)\|^2
\end{aligned}$$

Поэтому

$$\begin{aligned}\mu\|\hat{x}-x^*(Q)\|^2+\beta\|\hat{x}-\bar{x}\|^2 &\le \beta\langle B(\bar{x}-\hat{x}), x^*(Q)-\hat{x}\rangle+\langle g(\hat{x})-g(x^*(Q)),\hat{x}-\bar{x}\rangle\\ &\overset{(3.2.10)}{\le} (\beta+L)\cdot\|\hat{x}-x^*(Q)\|\cdot\|\hat{x}-\bar{x}\|,\end{aligned}$$

и мы получаем оценку (3.2.23).

Далее,

$$\begin{aligned}f(\bar{x}) &= \max_{y\in Q}\{\langle g(y),\bar{x}-y\rangle+\tfrac{1}{2}\mu\|y-\bar{x}\|^2\}\\ &\overset{(3.2.2)}{\le} \max_{y\in Q}\{\langle g(\bar{x}),\bar{x}-y\rangle-\tfrac{1}{2}\mu\|y-\bar{x}\|^2\}\\ &\overset{(3.2.25)}{\le} \max_{y\in Q}\{\langle g(\bar{x})-g(\hat{x})-\beta B(\bar{x}-\hat{x}),\bar{x}-y\rangle-\tfrac{1}{2}\mu\|y-\bar{x}\|^2\}.\end{aligned}$$

Поскольку оператор g монотонен, имеем

$$\begin{aligned}\|g(\bar{x})-g(\hat{x})-\beta B(\bar{x}-\hat{x})\|^2 &= \|g(\bar{x})-g(\hat{x})\|^2-2\beta\langle g(\bar{x})-g(\hat{x}),\bar{x}-\hat{x}\rangle+\|\bar{x}-\hat{x}\|^2\\ &\overset{(3.2.10)}{\le} (L^2+\beta^2)\|\bar{x}-\hat{x}\|^2.\end{aligned}$$

Поэтому $f(\bar{x})\le\frac{\beta^2+L^2}{2\mu}\|\hat{x}-\bar{x}\|^2$, и оставшееся неравенство следует из второго неравенства в формуле (3.2.23). □

Приведем некоторую модификацию метода (3.2.11). Основное ее отличие состоит в наличии предварительного шага в стиле [82]. Это позволяет получить скорость сходимости нового метода в терминах расстояния до решения.

```
Метод для сильно монотонных ВН

Инициализация  Фиксируем число шагов $N \ge 1$ и выбираем $\hat{x} \in Q$.
               Полагаем $\lambda_0 = 1$ и $y_0 = \bar{x} \overset{\mathrm{def}}{=} \arg\max_{x \in Q} \psi^L_{\hat{x}}(x)$.

Для $k = 0$ to $N - 1$ вычисляем:                                          (3.2.26)

    $x_k = \arg\max_{x \in Q} \Psi_k(x)$,
    $y_{k+1} = \arg\max_{x \in Q} \psi^L_{x_k}(x)$,
    $\lambda_{k+1} = \frac{1}{\gamma} \cdot S_k$.

Результат  $\tilde{y}_N(Q, \hat{x}) \overset{\mathrm{def}}{=} \frac{1}{S_N} \sum_{i=0}^{N} \lambda_i y_i$.
```

Теорема 3.2.7 *Пусть оператор g удовлетворяет условиям (3.2.2), (3.2.10). Тогда метод (3.2.26), применяемый к вариационному неравенству (3.2.3) сходится следующим образом:*

$$\| \tilde{y}_N(Q, \hat{x}) - x^*(Q) \| \;\le\; 3\gamma \exp\left\{ -\frac{N}{2(\gamma+1)} \right\} \cdot \| \hat{x} - x^*(Q) \|, \quad N \ge 1. \tag{3.2.27}$$

Доказательство.
Действительно,

$$\begin{aligned} \| \tilde{y}_N(Q, \hat{x}) - x^*(Q) \|^2 &\overset{(3.2.12)}{\le} \frac{2}{\mu} \left[f(\bar{x}) + \frac{\mu \cdot (\gamma^2 - 1)}{2} \cdot \| \bar{x} - x^*(Q) \|^2 \right] \cdot \exp\left\{ -\frac{N}{\gamma+1} \right\} \\ &\overset{(3.2.24)}{\le} \frac{2}{\mu} \left[\frac{4L^2}{\mu} + \frac{\mu \cdot (\gamma^2 - 1)}{2} \right] \cdot \exp\left\{ -\frac{N}{\gamma+1} \right\} \cdot \| \hat{x} - x^*(Q) \|^2 \\ &\le 9\gamma^2 \exp\left\{ -\frac{N}{\gamma+1} \right\} \cdot \| \hat{x} - x^*(Q) \|^2. \qquad \Box \end{aligned}$$

Рассмотрим эффективную стратегию решения слабо меняющегося квазивариационного неравенства. Сложность этой задачи зависит от числа обусловленности $\gamma = \frac{L}{\mu}$ оператора g и от *зазора сжатия*

$$\delta \;\overset{\mathrm{def}}{=}\; 1 - \alpha\gamma, \tag{3.2.28}$$

который предполагается положительным. Если скорость изменения α у допустимых выпуклых множеств $\mathcal{Q}(x)$ очень мала, то зазор сжатия будет близок к единице. Например, для КВН, решаемого градиентным методом (3.2.16), этот зазор будет по крайней мере $\frac{1}{2}$ (см. (3.2.18)).

Принимая во внимание оценку (3.2.27), определим минимальное число шагов $\hat{N} = N(\alpha,\gamma)$, удовлетворяющее условию

$$3\gamma \exp\left\{-\tfrac{\hat{N}}{2(\gamma+1)}\right\} \leq \tfrac{\delta}{4} \quad \Rightarrow \quad \hat{N} = \lfloor 2(\gamma+1)\ln\tfrac{12\gamma}{1-\alpha\gamma}\rfloor + 1. \tag{3.2.29}$$

Рассмотрим следующую двухуровневую схему.

$$\boxed{\begin{array}{l} \text{Выбираем } u_0 \in E. \text{ Для } k \geq 0 \text{ повторяем} \\ \\ \hat{x}_k = \pi_{\mathcal{Q}(u_k)}(u_k), \quad u_{k+1} = \tilde{y}_{\hat{N}}(\mathcal{Q}(u_k), \hat{x}_k). \end{array}} \tag{3.2.30}$$

Теорема 3.2.8 *Пусть параметр $\delta > 0$. Тогда существует единственное решение x_* квазивариационного неравенства (3.2.14) и метод (3.2.30) сходится со следующей скоростью:*

$$\|u_k - x_*\| \;\leq\; \tfrac{1}{\delta}\cdot\exp\left\{-\tfrac{\delta}{2}k\right\}\cdot\|u_0 - T(u_0)\|. \tag{3.2.31}$$

Доказательство.

Обозначим $r_k = \|u_k - T(u_k)\|$. Заметим, что

$$\begin{array}{rcl} \|u_{k+1} - T(u_k)\| & \overset{(3.2.27)}{\leq} & \frac{\delta}{4}\cdot\|\hat{x}_k - T(u_k)\| \\ \\ & \overset{(3.2.30)}{\leq} & \frac{\delta}{4}\cdot\|u_k - T(u_k)\|, \\ \\ \|T(u_{k+1}) - T(u_k)\| & \overset{\text{T.3.2.6}}{\leq} & \alpha\gamma\cdot\|u_{k+1} - u_k\|. \end{array}$$

Таким образом,

$$\begin{array}{rcl} r_{k+1} & \leq & \|u_{k+1} - T(u_k)\| + \|T(u_{k+1}) - T(u_k)\| \\ \\ & \leq & \frac{\delta}{4}\cdot r_k + \alpha\gamma\cdot\|u_{k+1} - u_k\| \\ \\ & \leq & \left(\alpha\gamma + \frac{\delta}{4}\right)\cdot r_k + \alpha\gamma\cdot\|u_{k+1} - T(u_k)\| \\ \\ & \leq & \left(1 - \frac{\delta}{2}\right)\cdot r_k. \end{array}$$

Следовательно, $r_k \leq \exp\left\{-\frac{\delta}{2}k\right\}\cdot r_0$. С другой стороны,

$$\begin{array}{rcl} r_k & = & \|u_k - T(u_k)\| \\ \\ & \geq & \|u_k - x_*\| - \|T(x_*) - T(u_k)\| \\ \\ & \overset{\text{T.3.2.6}}{\geq} & \delta\cdot\|u_k - x_*\|. \qquad \Box \end{array}$$

Следствие 3.2.3 *Для получения такой точки* u_k*, что* $\|u_k - x_*\| \le \epsilon$*, методу (3.2.30) потребуется не более чем*

$$\tfrac{4(\gamma+1)}{1-\alpha\gamma} \cdot \ln \tfrac{12\gamma}{1-\alpha\gamma} \cdot \ln \tfrac{\|u_0 - T(u_0)\|}{\epsilon\cdot(1-\alpha\gamma)} \tag{3.2.32}$$

вычислений значений оператора $g(x)$.

Нетрудно видеть, что эта оценка намного лучше соответствующего результата для градиентного метода (3.2.16).

Глава 4

Методы второго порядка

4.1 Кубическая регуляризация метода Ньютона

4.1.1 Введение

Мотивировка

Начиная с основополагающих работ Беннета [34] и Канторовича [61] метод Ньютона давно уже превратился в одно из важнейших средств для решения нелинейных задач. В простейшем случае, для задачи безусловной минимизации функции многих переменных

$$\min_{x \in \mathbb{R}^n} f(x),$$

стандартный метод Ньютона выглядит так:

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

Однако, несмотря на свою предельно естественную интерпретацию, метод Ньютона имеет несколько скрытых недостатков. Прежде всего, может случиться что в текущей точке этого процесса гессиан целевой функции является вырожденным. Тогда в методе происходит аварийная остановка. Во-вторых, может случиться, что этот метод расходится, или сходится к седловой точке, или, что еще хуже, сходится к точке локального максимума.

За последние пятьдесят лет было предложено огромное количество модификаций этого метода, так или иначе смягчающих вышеупомянутые недостатки. Количество ссылок в последней исчерпывающей монографии по этому вопросу [45] переваливает за тысячу. В большинстве подходов предлагается та или иная комбинация следующих приемов.

- *Регуляризация Левенберга – Маркварта.* Как было предложено в работах [69, 72], если матрица $\nabla^2 f(x)$ не является положительно определенной, ее надо регуляризовать с помощью единичной матрицы. А именно, для осуществления шага надо использовать направление, полученное с помощью матрицы $\nabla^2 f(x) + \gamma I \succ 0$:

$$x_{k+1} = x_k - [\nabla^2 f(x_k) + \gamma I]^{-1} \nabla f(x_k).$$

Иногда эта стратегия интерпретируется как комбинация градиентного метода с методом Ньютона.

- *Одномерный поиск.* Поскольку мы решаем задачу минимизации, целесообразно ввести в процесс шаговый множитель h_k:

  $$x_{k+1} = x_k - h_k[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$$

  (эта модификация называется *редуцированным* методом Ньютона [104]). Это помогает сформировать релаксационную последовательность: $f(x_{k+1}) \leq f(x_k)$.

- *Доверительные области* [52, 48, 46, 45]. В соответствии с этим подходом вокруг точки x_k надо зафиксировать окрестность, в которой аппроксимация второго порядка обязана быть достаточно хорошей. Эта окрестность $\Delta(x_k)$ называется доверительной областью. Можно, например, взять $\Delta(x_k) = \{x : ||x - x_k|| \leq \epsilon\}$ с некоторым $\epsilon > 0$. Тогда следующая точка x_{k+1} должна выбираться с помощью решения задачи

  $$\min_{x \in \Delta(x_k)} [\langle \nabla f(x_k), x - x_k \rangle + \tfrac{1}{2}\langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle].$$

  Заметим, что при $\Delta(x_k) \equiv \mathbb{R}^n$ это в точности стандартный ньютоновский шаг.

Однако практически никакие из этих подходов не затрагивают вопросы глобальной эффективности методов второго порядка. Исключение составляют оценки скорости сходимости редуцированного метода Ньютона для простого класса гладких сильно выпуклых функций (см. [95, 83]). Однако трудно сравнивать эти результаты с оценками сложности градиентных методов. На самом деле до сих пор отношения между градиентным методом и методом Ньютона не вполне выяснены. Конечно же, эти методы работают на разных классах задач (например, для метода Ньютона нам нужны более строгие условия на гладкость целевой функции). Да и трудоемкость итерации у них весьма различна (в методе Ньютона необходимо вычислять, хранить и обращать матрицы). Однако существуют нетривиальные приложения, где все эти различия отсутствуют и стоимость итерации у обоих методов сравнима. Тогда часто советуют начинать с градиентного метода и потом переходить на более быстрый метод Ньютона. В этой главе мы покажем, что такая практика не подкрепляется теоретическим анализом: можно показать, что метод Ньютона всегда (в смысле числа итераций) оказывается лучше градиентного.

В этом параграфе мы рассматриваем новую модификация метода Ньютона, которая устроена точно так же, как *градиентное отображение* (см. фомулу (1.2.2)). Предположим, что функция f имеет липшицев гессиан:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

Тогда нетрудно видеть, что вспомогательная функция

$$\xi_{2,x}(y) = f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y - x), y - x \rangle + \tfrac{L}{6}\|y - x\|^3$$

является *верхней оценкой второго порядка* для целевой функции:

$$f(y) \leq \xi_{2,x}(y) \quad \forall y \in \mathbb{R}^n.$$

Таким образом, можно попытаться формировать следующую тестовую точку в методе второго порядка с помощью задачи минимизации

$$x_{k+1} \in \operatorname{Arg}\min_{y} \ \xi_{2,x_k}(y) \tag{4.1.1}$$

(здесь *Argmin* обозначает глобальный минимум). В этом параграфе мы проведем детальный анализ этого подхода, который будем называть *кубической регуляризацией* метода Ньютона. Заметим, что задача (4.1.1) является невыпуклой и может иметь много локальных минимумов. Однако наш подход реализуем за полиномиальное время, так как мы показываем, что вспомогательная задача эквивалентна задачи минимизации специальной *выпуклой функции* от одной переменной.

Содержание

В п. 4.1.2 мы вводим кубическую регуляризацию и доказываем ее основные свойства. В п. 4.1.3 анализируются общие предельные свойства этого процесса. Мы доказываем, что при очень слабых предположениях все предельные точки траектории удовлетворяют необходимым условиям экстремума второго порядка. Для общей постановки мы обосновываем скорость сходимости норм градиентов к нулю, которая оказывается лучше, чем у градиентного метода. Мы также обосновываем локальную квадратичную сходимость процесса. В п. 4.1.4 приводятся глобальные оценки эффективности нашего метода на различных классах задач. Мы показываем, что во всех случаях скорость сходимости оказывается очень высокой (типа $O(\frac{1}{k^2})$ для звездно-выпуклых функций, где k – счетчик итераций). Более того, при весьма слабых условиях невырожденности мы доказываем локальную сверхлинейную скорость сходимости процесса либо порядка $\frac{4}{3}$, либо порядка $\frac{3}{2}$. Это происходит, даже если гессиан на оптимальном множестве является вырожденным.

В п. 4.1.5 мы показываем как найти решения задачи нахождения глобального минимума кубической регуляризации, и обсуждаем эффективные стратегии для оценки константы Липшица гессиана. И в конце параграфа, в п. 4.1.6, обсуждаются полученные результаты.

Обозначения

Обозначим через $\langle \cdot, \cdot \rangle$ стандартное скалярное произведение в $\mathbb{R}^n$:

$$\langle x, y \rangle = \sum_{i=1}^{n} x^{(i)} y^{(i)}, \quad x, y \in \mathbb{R}^n,$$

а через $\|x\|$ – стандартную евклидову норму:

$$\|x\| = \langle x, x\rangle^{1/2}.$$

Для симметрической $n \times n$-матрицы H ее спектр обозначается через $\{\lambda_i(H)\}_{i=1}^n$. Мы считаем, что все собственные числа занумерованы в порядке убывания:

$$\lambda_1(H) \geq \ldots \geq \lambda_n(H).$$

Поэтому $H \succeq 0$ тогда и только тогда, когда $\lambda_n(H) \geq 0$. Для прямоугольной матрицы A мы используем стандартную сингулярную норму:

$$\|A\| = \lambda_1(AA^T)^{1/2}.$$

Наконец, через I обозначается единичная $n \times n$-матрица.

4.1.2 Кубическая регуляризация квадратичной аппроксимации

Пусть выпуклое замкнутое множество $\mathcal{F} \subseteq \mathbb{R}^n$ имеет непустую внутренность. Рассмотрим дважды непрерывно дифференцируемую функцию $f(x)$, $x \in \mathcal{F}$. Пусть $x_0 \in \operatorname{int} \mathcal{F}$ является стартовой точкой нашего процесса. Предположим, что множество $\mathcal{F}$ достаточно большое: оно содержит в своей внутренности по крайней мере множество уровня

$$\mathcal{L}(f(x_0)) \equiv \{x \in \mathbb{R}^n : \ f(x) \leq f(x_0)\}.$$

Более того, в этом параграфе мы всегда предполагаем следующее.

Предположение 4.1.1 Гессиан функции f является липшицевым: $\mathcal{F}$:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathcal{F} \tag{4.1.2}$$

при некотором $L > 0$.

Для удобства изложения приведем здесь следующий вариант леммы 1.1.4.

Лемма 4.1.1 *Для любых точек x и y из $\mathcal{F}$ выполняются неравенства*

$$\|\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x)\| \overset{(1.1.15)}{\leq} \tfrac{1}{2}L\|y - x\|^2, \tag{4.1.3}$$

$$|f(y) - f(x) - \langle \nabla f(x), y - x\rangle - \tfrac{1}{2}\langle \nabla^2 f(x)(y - x), y - x\rangle| \overset{(1.1.16)}{\leq} \tfrac{L}{6}\|y - x\|^3. \tag{4.1.4}$$

Пусть M – положительный параметр. Зададим модифицированный ньютоновский шаг, пользуясь следующей *кубической регуляризацией* квадратичной аппроксимации функции $f(x)$:

$$T_M(x) \ \in \ \operatorname{Arg}\min_y \left[\langle \nabla f(x), y - x\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y - x), y - x\rangle + \tfrac{M}{6}\|y - x\|^3\right], \tag{4.1.5}$$

где "Arg" означает, что $T_M(x)$ выбирается из множества глобальных минимумов соответствующей задачи минимизации. Мы отложим дискуссию о сложности того вычисления до п. 4.1.5.

Заметим, что точка $T_M(x)$ удовлетворяет следующей системе нелинейных уравнений:

$$\nabla f(x) + \nabla^2 f(x)(y-x) + \tfrac{1}{2}M\|y-x\| \cdot (y-x) = 0. \tag{4.1.6}$$

Обозначим $r_M(x) = \|x - T_M(x)\|$. Выбирая в формуле (4.1.6) $y = T_M(x)$ и умножая его на $T_M(x) - x$, получаем уравнение

$$\langle \nabla f(x), T_M(x) - x\rangle + \langle \nabla^2 f(x)(T_M(x)-x), T_M(x)-x\rangle + \tfrac{1}{2}Mr_M^3(x) = 0. \tag{4.1.7}$$

В нашем анализе процесса (4.1.16) потребуется следующий факт.

Предложение 1 *При всех $x \in \mathcal{F}$ выполняется соотношение*

$$\nabla^2 f(x) + \tfrac{1}{2}Mr_M(x)I \succeq 0. \tag{4.1.8}$$

Это утверждение следует из теоремы 4.1.10, которую мы докажем в п. 4.1.5. А теперь представим основные свойства отображения $T_M(A)$.

Лемма 4.1.2 *Для любого $x \in \mathcal{F}$, $f(x) \le f(x_0)$, выполнено соотношение*

$$\langle \nabla f(x), x - T_M(x)\rangle \ge 0. \tag{4.1.9}$$

Если $M \ge \frac{2}{3}L$ и $x \in \operatorname{int}\mathcal{F}$, то $T_M(x) \in \mathcal{L}(f(x)) \subseteq \mathcal{F}$.

Доказательство.

Действительно, умножая соотношение (4.1.8) дважды на $x - T_M(x)$, получаем

$$\langle \nabla^2 f(x)(T_M(x)-x), T_M(x)-x\rangle + \tfrac{1}{2}Mr_M^3(x) \ge 0.$$

Таким образом, неравенство (4.1.9) следует из (4.1.7).

Далее, пусть $M \ge \frac{2}{3}L$. Предположим, что $T_M(x) \notin \mathcal{F}$. Тогда $r_M(x) > 0$. Рассмотрим следующие точки:

$$y(\alpha) = x + \alpha(T_M(x) - x), \quad \alpha \in [0,1].$$

Поскольку $y(0) \in \operatorname{int}\mathcal{F}$, значение

$$\bar{\alpha} : \ y(\bar{\alpha}) \in \partial\mathcal{F}$$

корректно определено. В соответствии с нашим предположением $\bar{\alpha} < 1$ и $y(\alpha) \in \mathcal{F}$ при любых $\alpha \in [0,\bar{\alpha}]$. Таким образом, пользуясь оценкой (4.1.4), соотношением (4.1.7) и неравенством (4.1.9), получаем

$$\begin{aligned} f(y(\alpha)) &\le f(x) + \langle \nabla f(x), y(\alpha)-x\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y(\alpha)-x), y(\alpha)-x\rangle + \tfrac{\alpha^3 L}{6}r_M^3(x) \\ &\le f(x) + \langle \nabla f(x), y(\alpha)-x\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y(\alpha)-x), y(\alpha)-x\rangle + \tfrac{\alpha^3 M}{4}r_M^3(x) \\ &= f(x) + (\alpha - \tfrac{\alpha^2}{2})\langle \nabla f(x), T_M(x)-x\rangle - \tfrac{\alpha^2(1-\alpha)}{4}Mr_M^3(x) \\ &\le f(x) - \tfrac{\alpha^2(1-\alpha)}{4}Mr_M^3(x). \end{aligned}$$

Итак, $f(y(\bar{\alpha})) < f(x)$. Следовательно, $y(\bar{\alpha}) \in \operatorname{int} \mathcal{L}(f(x)) \subseteq \operatorname{int} \mathcal{F}$. Мы получили противоречие. Поэтому $T_M(x) \in \mathcal{F}$. Пользуясь такими же аргументами, мы доказываем, что $f(T_M(x)) \leq f(x)$. □

Лемма 4.1.3 *Если $T_M(x) \in \mathcal{F}$, то*

$$\|\nabla f(T_M(x))\| \leq \tfrac{1}{2}(L+M)r_M^2(x). \tag{4.1.10}$$

Доказательство.
Из уравнения (4.1.6) получаем

$$\|\nabla f(x) + \nabla^2 f(x)(T_M(x) - x)\| = \tfrac{1}{2}Mr_M^2(x).$$

С другой стороны, в силу (4.1.3) имеем

$$\|\nabla f(T_M(x)) - \nabla f(x) - \nabla^2 f(x)(T_M(x) - x)\| \leq \tfrac{1}{2}Lr_M^2(x).$$

Сопоставляя эти соотношения, получаем неравенство (4.1.10). □

Положим

$$\bar{f}_M(x) = \min_y \left[f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle + \tfrac{M}{6}\|y-x\|^3 \right].$$

Лемма 4.1.4 *При любом $x \in \mathcal{F}$ выполняются неравенства*

$$\bar{f}_M(x) \leq \min_{y \in \mathcal{F}} \left[f(y) + \tfrac{L+M}{6}\|y-x\|^3 \right], \tag{4.1.11}$$

$$f(x) - \bar{f}_M(x) \geq \tfrac{M}{12}\, r_M^3(x). \tag{4.1.12}$$

Более того, если $M \geq L$, то $T_M(x) \in \mathcal{F}$ и

$$f(T_M(x)) \leq \bar{f}_M(x). \tag{4.1.13}$$

Доказательство.
Действительно, пользуясь нижней оценкой в неравенстве (4.1.4), для любого $y \in \mathcal{F}$ получаем

$$f(x) + \langle \nabla f(x), y-x\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle \leq f(y) + \tfrac{L}{6}\|y-x\|^3,$$

и неравенство (4.1.11) следует из определения величины $\bar{f}_M(x)$.

Далее, в силу определения точки $T_M(x)$, соотношения (4.1.7) и неравенства (4.1.9) имеем

$$\begin{aligned} f(x) - \bar{f}_M(x) &= \langle \nabla f(x), x - T_M(x)\rangle - \tfrac{1}{2}\langle \nabla^2 f(x)(T_M(x)-x), T_M(x)-x\rangle - \tfrac{M}{6}r_M^3(x) \\ &= \tfrac{1}{2}\langle \nabla f(x), x - T_M(x)\rangle + \tfrac{M}{12}\, r_M^3(x) \geq \tfrac{M}{12}\, r_M^3(x). \end{aligned}$$

Наконец, если $M \geq L$, то $T_M(x) \in \mathcal{F}$ ввиду леммы 4.1.2. Таким образом, неравенство (4.1.13) следует из верхней оценки в неравенстве (4.1.4). □

4.1.3 Общие результаты о сходимости

Рассмотрим следующую задачу:

$$\min_{x \in \mathbb{R}^n} f(x), \tag{4.1.14}$$

где целевая функция $f(x)$ удовлетворяет предположению 4.1.1. Напомним, что необходимое условие того, что точка x^* является локальным решением задачи (4.1.14), записывается так:

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succeq 0. \tag{4.1.15}$$

Таким образом, для произвольного $x \in \mathcal{F}$ можно ввести следующую меру локальной оптимальности:

$$\mu_M(x) = \max \left\{ \sqrt{\tfrac{2}{L+M} \|\nabla f(x)\|}, -\tfrac{2}{2L+M} \lambda_n(\nabla^2 f(x)) \right\},$$

где M – положительный параметр. Ясно, что для любого $x \in \mathcal{F}$ мера $\mu_M(x)$ неотрицательна и равна нулю только в точках, удовлетворяющих условиям (4.1.15). Аналитическая форма этой меры объясняется следующим результатом.

Лемма 4.1.5 *Для любого $x \in \mathcal{F}$ выполняется неравенство $\mu_M(T_M(x)) \le r_M(x)$.*

Доказательство.

Доказательство немедленно следует из неравенства (4.1.10) и сотношения (4.1.8), поскольку

$$\nabla^2 f(T_M(x)) \ge \nabla^2 f(x) - L r_M(x) I \ge -(\tfrac{1}{2} M + L) r_M(x) I. \qquad \Box$$

Пусть $L_0 \in (0, L]$ – положительный параметр. Рассмотрим следующую схему регуляризованного метода Ньютона.

Кубическая регуляризация метода Ньютона

Инициализация Выбираем $x_0 \in \mathbb{R}^n$.

Итерация k, ($k \ge 0$) (4.1.16)

1. Находим такое $M_k \in [L_0, 2L]$, что $f(T_{M_k}(x_k)) \le \bar{f}_{M_k}(x_k)$.

2. Полагаем $x_{k+1} = T_{M_k}(x_k)$.

Поскольку $\bar{f}_M(x) \leq f(x)$, этот процесс является монотонным:

$$f(x_{k+1}) \leq f(x_k).$$

Если константа L известна, на шаге 1 можно взять $M_k \equiv L$. В противном случае можно воспользоваться простой процедурой поиска, которую мы обсудим в п. 4.1.5. А сейчас остановимся на следующем факте.

Теорема 4.1.1 *Пусть последовательность* $\{x_i\}$ *формируется методом Ньютона (4.1.16). Предположим, что целевая функция* $f(x)$ *ограничена снизу:*

$$f(x) \geq f^* \quad \forall x \in \mathcal{F}.$$

Тогда $\sum\limits_{i=0}^{\infty} r_{M_i}^3(x_i) \leq \frac{12}{L_0}(f(x_0) - f^*)$. *Более того,* $\lim\limits_{i\to\infty} \mu_L(x_i) = 0$, *и для любого* $k \geq 1$ *выполняется неравенство*

$$\min_{1\leq i\leq k} \mu_L(x_i) \leq \frac{8}{3} \cdot \left(\frac{3(f(x_0)-f^*)}{2k\cdot L_0}\right)^{1/3}. \tag{4.1.17}$$

Доказательство.
В силу неравенства (4.1.12) имеем

$$f(x_0) - f^* \geq \sum_{i=0}^{k-1}[f(x_i) - f(x_{i+1}) \geq \sum_{i=0}^{k-1} \frac{M_i}{12} r_{M_i}^3(x_i) \geq \frac{L_0}{12} r_{M_i}^3(x_i).$$

Остается воспользоваться утверждением леммы 4.1.5 и верхней оценкой на величину M_k в методе (4.1.16):

$$r_{M_i}(x_i) \geq \mu_{M_i}(x_{i+1}) \geq \frac{3}{4}\mu_L(x_{i+1}). \qquad \square$$

Заметим, что неравенство (4.1.17) означает, что

$$\min_{1\leq i\leq k} \|\nabla f(x_i)\| \leq O(k^{-2/3}).$$

Мы видели, что для градиентного метода скорость убывания правой части в таком неравенстве имеет порядок $O\left(k^{-1/2}\right)$ (см. неравенство (2.3.33)).

Из теоремы 4.1.1 можно выводить разные результаты о сходимости метода. Приведем один из них.

Теорема 4.1.2 *Пусть последовательность* $\{x_i\}$ *формируется методом (4.1.16). Предположим, что при некотором* $i \geq 0$ *множество* $\mathcal{L}(f(x_i))$ *ограничено. Тогда существует предельное значение*

$$\lim_{i\to\infty} f(x_i) = f^*.$$

Множество X^* *предельных точек этой последовательности непусто. Более того, это множество связно, и для любой точки* $x^* \in X^*$ *выполнены условия*

$$f(x^*) = f^*, \quad \nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succeq 0.$$

Доказательство.
Это утверждение непосредственно выводится из теоремы 4.1.1 с помощью стандартных рассуждений. □

Опишем теперь поведение процесса (4.1.16) в окрестности невырожденной стационарной точки, которая не является точкой локального минимума.

Лемма 4.1.6 *Пусть точка* $\bar{x} \in \operatorname{int} \mathcal{F}$ *является невырожденной седловой точкой или точкой локального максимума функции* $f(x)$:

$$\nabla f(\bar{x}) = 0, \quad \lambda_n(\nabla^2 f(\bar{x})) < 0.$$

Тогда найдутся такие константы ϵ, $\delta > 0$, *что как только точка* x_i *попадет внутрь множества* $Q = \{x : \|x - \bar{x}\| \le \epsilon, f(x) \ge f(\bar{x})\}$ *(например,* $x_i = \bar{x}$*), следующая точка* x_{i+1} *немедленно его покинет:*

$$f(x_{i+1}) \le f(\bar{x}) - \delta.$$

Доказательство.
Пусть для некоторого направления d, $\|d\| = 1$, и числа $\bar{\tau} > 0$ выполняются соотношения

$$\langle \nabla^2 f(\bar{x}) d, d\rangle \equiv -\sigma < 0, \quad \bar{x} \pm \bar{\tau} d \in \mathcal{F}.$$

Положим $\epsilon = \min\left\{\frac{\sigma}{2L}, \bar{\tau}\right\}$ и $\delta = \frac{\sigma}{6}\epsilon^2$. Тогда в силу неравенства (4.1.11), верхней оценки на M_i и неравенства (4.1.4) для всех $|\tau| \le \bar{\tau}$ получаем следующую оценку:

$$\begin{aligned} f(x_{i+1}) &\le f(\bar{x} + \tau d) + \tfrac{L}{2}\|\bar{x} + \tau d - x_i\|^3 \\ &\le f(\bar{x}) - \sigma\tau^2 + \tfrac{L}{6}|\tau|^3 + \tfrac{L}{2}\left[\epsilon^2 + 2\tau\langle d, \bar{x} - x_i\rangle + \tau^2\right]^{3/2}. \end{aligned}$$

Поскольку мы свободны в выборе знака числа τ, можно гарантировать, что

$$f(x_{i+1}) \le f(\bar{x}) - \sigma\tau^2 + \tfrac{L}{6}|\tau|^3 + \tfrac{L}{2}\left[\epsilon^2 + \tau^2\right]^{3/2}, \quad |\tau| \le \bar{\tau}.$$

Выберем $\tau = \epsilon \le \bar{\tau}$. Тогда

$$f(x_{i+1}) \le f(\bar{x}) - \sigma\tau^2 + \tfrac{5L}{3}\tau^3 \le f(\bar{x}) - \sigma\tau^2 + \tfrac{5L}{3} \cdot \tfrac{\sigma}{2L} \cdot \tau^2 = f(\bar{x}) - \tfrac{1}{6}\sigma\tau^2.$$

Поскольку процесс (4.1.16) монотонен по целевой функции, он никогда снова не попадет в множество Q. □

Рассмотрим теперь поведение регуляризированного метода Ньютона (4.1.16) в окрестности невырожденной точки локального минимума. В такой ситуации предположение $L_0 > 0$ не является больше необходимым. Рассмотрим упрощенную версию метода (4.1.16):

$$\boxed{x_{k+1} = T_{M_k}(x_k), \ k \ge 0,} \tag{4.1.18}$$

где $M_k \in (0, 2L]$. Обозначим

$$\delta_k = \frac{L\|\nabla f(x_k)\|}{\lambda_n^2(\nabla^2 f(x_k))}.$$

Теорема 4.1.3 *Пусть* $\nabla^2 f(x_0) \succ 0$ *и* $\delta_0 \le \frac{1}{4}$. *Если последовательность точек* $\{x_k\}$ *сформирована методом (4.1.18), то*

1) при всех $k \ge 0$ *значения* δ_k *корректно определены и сходятся к нулю с квадратичной скоростью:*

$$\delta_{k+1} \le \frac{3}{2}\left(\frac{\delta_k}{1-\delta_k}\right)^2 \le \frac{8}{3}\delta_k^2 \le \frac{2}{3}\delta_k, \quad k \ge 0; \tag{4.1.19}$$

2) минимальные собственные значения гессианов $\nabla^2 f(x_k)$ *лежат в следующих пределах:*

$$e^{-1}\lambda_n(\nabla^2 f(x_0)) \le \lambda_n(\nabla^2 f(x_k)) \le e^{3/4}\lambda_n(\nabla^2 f(x_0)); \tag{4.1.20}$$

3) вся последовательность $\{x_i\}$ *квадратично сходится к точке* x^*, *которая является невырожденной точкой минимума функции* $f(x)$; *в частности, для любого* $k \ge 1$ *выполнено неравенство*

$$\|\nabla f(x_k)\| \le \lambda_n^2(\nabla^2 f(x_0))\,\frac{9e^{3/2}}{16L}\left(\frac{1}{2}\right)^{2^k}. \tag{4.1.21}$$

Доказательство.

Пусть для некоторого $k \ge 0$ выполнено условие $\nabla^2 f(x_k) \succ 0$. Тогда соответствующее δ_k корректно определено. Предположим, что $\delta_k \le \frac{1}{4}$. В силу уравнения (4.1.6) имеем

$$r_{M_k}(x_k) = \|T_{M_k}(x_k) - x_k\| = \|(\nabla^2 f(x_k) + r_{M_k}(x_k)\tfrac{M_k}{2}I)^{-1}\nabla f(x_k)\| \le \frac{\|\nabla f(x_k)\|}{\lambda_n(\nabla^2 f(x_k))}. \tag{4.1.22}$$

Заметим, что $\nabla^2 f(x_{k+1}) \succeq \nabla^2 f(x_k) - r_{M_k}(x_k)LI$. Таким образом,

$$\begin{aligned}\lambda_n(\nabla^2 f(x_{k+1})) &\ge \lambda_n(\nabla^2 f(x_k)) - r_{M_k}(x_k)L \\ &\ge \lambda_n(\nabla^2 f(x_k)) - \frac{L\|\nabla f(x_k)\|}{\lambda_n(\nabla^2 f(x_k))} = (1-\delta_k)\lambda_n(\nabla^2 f(x_k)).\end{aligned}$$

Следовательно, матрица $\nabla^2 f(x_{k+1})$ также положительно определена. Более того, пользуясь неравенством (4.1.10) и верхней оценкой на M_k, мы получаем

$$\delta_{k+1} = \frac{L\|\nabla f(x_{k+1})\|}{\lambda_n^2(\nabla^2 f(x_{k+1}))} \le \frac{3L^2 r_{M_k}^2(x_k)}{2\lambda_n^2(\nabla^2 f(x_{k+1}))} \le \frac{3L^2\|\nabla f(x_k)\|^2}{2\lambda_n^4(\nabla^2 f(x_k))(1-\delta_k)^2} = \frac{3}{2}\left(\frac{\delta_k}{1-\delta_k}\right)^2 \le \frac{8}{3}\delta_k^2.$$

Таким образом, $\delta_{k+1} \le \frac{1}{4}$, и неравенство (4.1.19) доказывается по индукции. Заметим, что мы также получаем $\delta_{k+1} \le \frac{2}{3}\delta_k$.

Далее, как мы уже видели,

$$\ln\frac{\lambda_n(\nabla^2 f(x_k))}{\lambda_n(\nabla^2 f(x_0))} \ge \sum_{i=0}^{\infty}\ln(1-\delta_i) \ge -\sum_{i=0}^{\infty}\frac{\delta_i}{1-\delta_i} \ge -\frac{1}{1-\delta_0}\sum_{i=0}^{\infty}\delta_i \ge -1.$$

Для получения верхней оценки заметим, что $\nabla^2 f(x_{k+1}) \preceq \nabla^2 f(x_k) + r_{M_k}(x_k)LI$. Поэтому

$$\lambda_n(\nabla^2 f(x_{k+1}) \le \lambda_n(\nabla^2 f(x_k)) + r_{M_k}(x_k)L \le (1+\delta_k)\lambda_n(\nabla^2 f(x_k)).$$

Таким образом,

$$\ln \tfrac{\lambda_n(\nabla^2 f(x_k))}{\lambda_n(\nabla^2 f(x_0))} \leq \sum\limits_{i=0}^{\infty} \ln(1+\delta_i) \leq \sum\limits_{i=0}^{\infty} \delta_i \leq \tfrac{3}{4}.$$

Остается доказать п. 3 теоремы. В силу неравенств (4.1.22) и (4.1.20) имеем

$$r_{M_k}(x_k) \leq \tfrac{1}{L}\lambda_n(\nabla^2 f(x_k))\delta_k \leq \tfrac{e^{3/4}}{L}\lambda_n(\nabla^2 f(x_0))\delta_k.$$

Таким образом, последовательность $\{x_i\}$ удовлетворяет критерию Коши и, следовательно, имеет единственную предельную точку x^*. Поскольку собственные значения матрицы $\nabla^2 f(x)$ являются непрерывными функциями от x, из неравенства (4.1.20) вытекает, что $\nabla^2 f(x^*) > 0$.

Далее, из неравенства (4.1.19) получаем

$$\delta_{k+1} \leq \tfrac{\delta_k^2}{(1-\delta_0)^2} \leq \tfrac{16}{9}\delta_k^2.$$

Обозначая $\hat{\delta}_k = \frac{16}{9}\delta_k$, получаем $\hat{\delta}_{k+1} \leq \hat{\delta}_k^2$. Таким образом, для любого $k \geq 1$ имеем

$$\delta_k = \tfrac{9}{16}\hat{\delta}_k \leq \tfrac{9}{16}\hat{\delta}_0^{2^k} < \tfrac{9}{16}\left(\tfrac{1}{2}\right)^{2^k}.$$

Используя эту верхнюю границу в неравенстве (4.1.20), получаем соотношение (4.1.21). □

4.1.4 Глобальные оценки эффективности для специальных классов задач

В предыдущем пункте мы видели, что для модифицированного метода Ньютона возможно устанавливать глобальные оценки эффективности типа (4.1.17), справедливые для общих классов невыпуклых задач. В этом пункте мы покажем, что для некоторых специальных классов невыпуклых задач оценки эффективности метода (4.1.16) могут быть гораздо лучше. Хорошей особенностью метода (4.1.16) является то, что он приспосабливается к специфике задачи автоматически.

Звездно-выпуклые функции

Начнем со следующего определения.

Определение 4.1.1 Функция $f(x)$ называется *звездно-выпуклой*, если множество ее глобальных минимумов непусто и для любых точек $x^* \in X^*$ и $x \in \mathbb{R}^n$ выполняется неравенство

$$f(\alpha x^* + (1-\alpha)x) \leq \alpha f(x^*) + (1-\alpha)f(x) \quad \forall x \in \mathcal{F},\ \forall \alpha \in [0,1]. \qquad (4.1.23)$$

Такой функцией является, например, обычная выпуклая функция. Однако в общем случае звездно-выпуклые функции не являются выпуклыми даже в одномерной ситуации. Например, функция $f(x) = |x|(1-e^{-|x|}), x \in R$, звездно-выпукла, но не выпукла. Звездно-выпуклые функции часто возникают в задачах, связанных с суммами квадратов. Например, этому классу принадлежит функция $f(x,y) = x^2y^2 + x^2 + y^2$.

Теорема 4.1.4 *Пусть целевая функция в задаче (4.1.14) является звездно-выпуклой, а множество $\mathcal{F}$ – ограниченным:* $\mathrm{diam}\, \mathcal{F} = D < \infty$, *и пусть последовательность $\{x_k\}$ формируется методом (4.1.16).*

1. Если $f(x_0) - f^* \geq \frac{3}{2}LD^3$, то $f(x_1) - f^* \leq \frac{1}{2}LD^3$.

2. Если же $f(x_0) - f^* \leq \frac{3}{2}LD^3$, то метод (4.1.16) будет сходиться со следующей скоростью:

$$f(x_k) - f(x^*) \leq \frac{3LD^3}{2(1+\frac{1}{3}k)^2}, \quad k \geq 0. \tag{4.1.24}$$

Доказательство.
Действительно, в силу неравенства (4.1.11), верхней оценки на параметры M_k и определения (4.1.23) для любого $k \geq 0$ имеем

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\leq \min_y \left[f(y) - f(x^*) + \tfrac{L}{2}\|y - x_k\|^3 : \ y = \alpha x^* + (1-\alpha)x_k, \alpha \in [0,1] \right] \\ &\leq \min_{\alpha \in [0,1]} \left[f(x_k) - f(x^*) - \alpha(f(x_k) - f(x^*)) + \tfrac{L}{2}\alpha^3 \|x^* - x_k\|^3 \ \right] \\ &\leq \min_{\alpha \in [0,1]} \left[f(x_k) - f(x^*) - \alpha(f(x_k) - f(x^*)) + \tfrac{L}{2}\alpha^3 D^3 \ \right]. \end{aligned}$$

Минимум целевой функции в последней задаче минимизации по $\alpha \geq 0$ достигается при

$$\alpha_k = \sqrt{\tfrac{2(f(x_k) - f(x^*))}{3LD^3}}.$$

Если $\alpha_k \geq 1$, то на самом деле мы должны взять $\alpha = 1$. В этом случае

$$f(x_{k+1}) - f(x^*) \leq \tfrac{1}{2}LD^3.$$

Поскольку процесс (4.1.16) монотонен, такое может случиться только на первой итерации метода.

Предположим, что $\alpha_k \leq 1$. Тогда

$$f(x_{k+1}) - f(x^*) \leq f(x_k) - f(x^*) - \left[\tfrac{2}{3}(f(x_k) - f(x^*))\right]^{3/2} \tfrac{1}{\sqrt{LD^3}}.$$

В более удобных обозначениях это перепишется так: $\alpha_{k+1}^2 \leq \alpha_k^2 - \frac{2}{3}\alpha_k^3 < \alpha_k^2$. Следовательно,

$$\tfrac{1}{\alpha_{k+1}} - \tfrac{1}{\alpha_k} = \tfrac{\alpha_k - \alpha_{k+1}}{\alpha_k \alpha_{k+1}} = \tfrac{\alpha_k^2 - \alpha_{k+1}^2}{\alpha_k \alpha_{k+1}(\alpha_k + \alpha_{k+1})} \geq \tfrac{\alpha_k^2 - \alpha_{k+1}^2}{2\alpha_k^3} \geq \tfrac{1}{3}.$$

Таким образом, $\frac{1}{\alpha_k} \geq \frac{1}{\alpha_0} + \frac{k}{3} \geq 1 + \frac{k}{3}$, и мы получаем неравенство (4.1.24). □

Введем теперь обобщение понятия невырожденного глобального минимума.

Определение 4.1.2 Мы говорим, что оптимальное множество X^* функции $f(x)$ *глобально невырождено*, если существует такая константа $\gamma > 0$, что для любого $x \in \mathcal{F}$ выполнено неравенство

$$f(x) - f^* \geq \tfrac{\gamma}{2}\, \rho^2(x, X^*), \tag{4.1.25}$$

где f^* – глобальное минимальное значение функции $f(x)$ и $\rho(x, X^*)$ – евклидово расстояние от точки x до множества X^*.

Конечно же, этим свойством обладают сильно выпуклые функции (в этом случае X^* состоит из единственной точки). Однако им могут обладать и невыпуклые функции. В качестве примера можно привести $f(x) = (\|x\|^2 - 1)^2, X^* = \{x : \|x\| = 1\}$. Более того, если множество X^* имеет связную компоненту, гессианы целевой функции в этих точках *не могут* быть невырожденными. Тем не менее, мы увидим что даже в такой ситуации модифицированная схема метода Ньютона сходится сверхлинейно.

Обозначим

$$\bar{\omega} = \tfrac{1}{L^2}\left(\tfrac{\gamma}{2}\right)^3.$$

Теорема 4.1.5 *Пусть функция $f(x)$ является звездно-выпуклой. Предположим также, что она имеем глобально невырожденное оптимальное множество. Тогда метод (4.1.16) обладает следующими свойствами.*

1. Если $f(x_0) - f(x^*) \geq \frac{4}{9}\bar{\omega}$, то на первой фазе этого процесса имеется следующая скорость сходимости:

$$f(x_k) - f(x^*) \leq \left[(f(x_0) - f(x^*))^{1/4} - \tfrac{k}{6}\sqrt{\tfrac{2}{3}}\bar{\omega}^{1/4}\right]^4. \tag{4.1.26}$$

Эта фаза прерывается, как только $f(x_{k_0}) - f(x^*) \leq \frac{4}{9}\bar{\omega}$ при некотором $k_0 \geq 0$.

2. При $k \geq k_0$ эта последовательность сходится сверхлиенейно:

$$f(x_{k+1}) - f(x^*) \leq \tfrac{1}{2}(f(x_k) - f(x^*))\sqrt{\tfrac{f(x_k)-f(x^*)}{\bar{\omega}}}. \tag{4.1.27}$$

Доказательство.

Обозначим через x_k^* проекцию точки x_k на оптимальное множество X^*. В силу неравенства (4.1.11), верхней оценки на параметры M_k и определений (4.1.23), (4.1.25) для любого $k \geq 0$ имеем

$$\begin{aligned} f(x_{k+1}) - f(x^*) \; &\leq \; \min_{\alpha \in [0,1]} \left[f(x_k) - f(x^*) - \alpha(f(x_k) - f(x^*)) + \tfrac{L}{2}\alpha^3 \|x_k^* - x_k\|^3\, \right] \\ &\leq \; \min_{\alpha \in [0,1]} \left[f(x_k) - f(x^*) - \alpha(f(x_k) - f(x^*)) + \tfrac{L}{2}\alpha^3 \left(\tfrac{2}{\gamma}(f(x_k) - f(x^*))\right)^{3/2}\right]. \end{aligned}$$

Обозначая $\Delta_k = (f(x_k) - f(x^*))/\bar{\omega}$ мы получаем неравенство

$$\Delta_{k+1} \leq \min_{\alpha \in [0,1]} \left[\Delta_k - \alpha \Delta_k + \tfrac{1}{2}\alpha^3 \Delta_k^{3/2}\right]. \tag{4.1.28}$$

Заметим, что условия оптимальности первого порядка по $\alpha \geq 0$ для этой задачи записываются так:

$$\alpha_k = \sqrt{\tfrac{2}{3}\Delta_k^{-1/2}}.$$

Таким образом, если $\Delta_k \geq \frac{4}{9}$, то

$$\Delta_{k+1} \leq \Delta_k - \left(\tfrac{2}{3}\right)^{3/2} \Delta_k^{3/4}.$$

Обозначая $u_k = \frac{9}{4}\Delta_k$, получаем более простое соотношение

$$u_{k+1} \leq u_k - \tfrac{2}{3} u_k^{3/4},$$

которое применимо для $u_k \geq 1$. Поскольку правая часть этого неравенства возрастает по $u_k \geq \frac{1}{16}$, докажем по индукции, что

$$u_k \leq \left[u_0^{1/4} - \tfrac{k}{6}\right]^4.$$

Действительно, неравенство

$$\left[u_0^{1/4} - \tfrac{k+1}{6}\right]^4 \geq \left[u_0^{1/4} - \tfrac{k}{6}\right]^4 - \tfrac{2}{3}\left[u_0^{1/4} - \tfrac{k}{6}\right]^3$$

эквивалентно неравенству

$$\tfrac{2}{3}\left[u_0^{1/4} - \tfrac{k}{6}\right]^3 \geq \left[u_0^{1/4} - \tfrac{k}{6}\right]^4 - \left[u_0^{1/4} - \tfrac{k+1}{6}\right]^4 = \tfrac{1}{6}\Big[\left[u_0^{1/4} - \tfrac{k}{6}\right]^3$$

$$+ \left[u_0^{1/4} - \tfrac{k}{6}\right]^2 \left[u_0^{1/4} - \tfrac{k+1}{6}\right] + \left[u_0^{1/4} - \tfrac{k}{6}\right] \left[u_0^{1/4} - \tfrac{k+1}{6}\right]^2 + \left[u_0^{1/4} - \tfrac{k+1}{6}\right]^3\Big],$$

которое очевидным образом выполнено.

Наконец, если $u_k \leq 1$, то оптимальное значение для α в задаче (4.1.28) есть единица и мы приходим к неравенству (4.1.27). □

Градиентно доминируемые функции

Рассмотрим другой интересный класс задач.

Определение 4.1.3 Функция $f(x)$ называется *градиентно-доминируемой* функцией степени $p \in [1, 2]$, если она достигает глобального минимума в некоторой точке x^* и для любого $x \in \mathcal{F}$ выполняется неравенство

$$f(x) - f(x^*) \leq \tau_f \|\nabla f(x)\|^p, \tag{4.1.29}$$

где τ_f – положительная константа. Параметр p называется *степенью* доминирования.

Заметим, что в этом определении не предполагается, что глобальный минимум функции f является единственным.

Приведем несколько примеров таких функций.

Пример 4.1.1 *Выпуклые функции.* Пусть функция f является выпуклой на $\mathbb{R}^n$. Предположим, что она достигает минимума в точке x^*. Тогда для любого $x \in \mathbb{R}^n$, $\|x - x^*\| \leq R$, имеем

$$f(x) - f(x^*) \leq \langle \nabla f(x), x - x^* \rangle \leq \|\nabla f(x)\| \cdot R.$$

Таким образом, на множестве $\mathcal{F} = \{x : \ \|x - x^*\| \leq R\}$ функция f является градиентно доминируемой степени единица с $\tau_f = R$. ☐

Пример 4.1.2 *Сильно выпуклые функции.* Пусть функция f является дифференцируемой и сильно выпуклой на $\mathbb{R}^n$. Это означает, что существует такая константа $\gamma > 0$, что

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2}\gamma \|y - x\|^2, \tag{4.1.30}$$

при всех $x, y \in \mathbb{R}^n$. Тогда, минимизируя обе части этого неравенства по y, получаем

$$f(x) - f(x^*) \leq \tfrac{1}{2\gamma} \|\nabla f(x)\|^2 \quad \forall x \in \mathbb{R}^n.$$

Таким образом, на множестве $\mathcal{F} = \mathbb{R}^n$ функция f является градиентно доминируемой степени два с $\tau_f = \frac{1}{2\gamma}$. ☐

Пример 4.1.3 *Сумма квадратов.* Рассмотрим систему нелинейных уравнений

$$g(x) = 0, \tag{4.1.31}$$

где функция $g(x) = (g_1(x), \ldots, g_m(x))^T : \mathbb{R}^n \to \mathbb{R}^m$ дифференцируема. Предположим, что $m \leq n$ и что у этой системы существует решение x^*. При этом предположим, что якобиан этой системы

$$J(x) = (g_1'(x), \ldots, g_m'(x))$$

является равномерно невырожденным на некотором выпуклом множестве $\mathcal{F}$, содержащем x^*. Это означает, что значение

$$\sigma \equiv \inf_{x \in \mathcal{F}} \lambda_n \left(J^T(x) J(x) \right)$$

положительно. Рассмотрим функцию

$$f(x) = \tfrac{1}{2} \sum_{i=1}^{m} g_i^2(x).$$

Ясно, что $f(x^*) = 0$. Заметим, что $\nabla f(x) = J(x)g(x)$. Таким образом,

$$\|\nabla f(x)\|^2 = \langle \left(J^T(x)J(x)\right) g(x), g(x) \rangle \geq \sigma \|g(x)\|^2 = 2\sigma(f(x) - f(x^*)).$$

Следовательно, функция f является градиентно доминируемой на $\mathcal{F}$ степени два с $\tau_f = \frac{1}{2\sigma}$. Заметим, что при $m < n$ множество решений системы (4.1.31) *не может* быть одноточечным. Таким образом, гессианы этой функции обязательно будут вырожденными на множестве решений. □

Для дальнейшего нам потребуется следующий вспомогательный результат.

Лемма 4.1.7 *На каждом шаге метода (4.1.16) можно гарантировать следующее убывание целевой функции:*

$$f(x_k) - f(x_{k+1}) \geq \tfrac{L_0 \cdot \|\nabla f(x_{k+1})\|^{3/2}}{3\sqrt{2} \cdot (L+L_0)^{3/2}}, \quad k \geq 0. \tag{4.1.32}$$

Доказательство.
В силу неравенств (4.1.12) и (4.1.10) получаем

$$f(x_k) - f(x_{k+1}) \geq \tfrac{M_k}{12} r^3_{M_k}(x_k) \geq \tfrac{M_k}{12} \left(\tfrac{2\|\nabla f(x_{k+1})\|}{L+M_k} \right)^{3/2} = \tfrac{M_k \|\nabla f(x_{k+1})\|^{3/2}}{3\sqrt{2} \cdot (L+M_k)^{3/2}}.$$

Остается заметить, что правая часть этого неравенства возрастает при $M_k \leq 2L$. Таким образом, можно заменить M_k на ее нижнюю оценку L_0. □

Начнем с анализа сложности минимизации градиентно доминируемых функций степени единица. В следующей теореме утверждается, что процесс минимизации делится на две фазы. Первая фаза (с большими значениями целевой функции) заканчивается достаточно быстро. На второй фазе наблюдается сходимость со скоростью $O(1/k^2)$.

Теорема 4.1.6 *Применим метод (4.1.16) к решению задачи минимизации градиентно доминируемой функции $f(x)$ степени $p = 1$.*

1. Если начальное значение целевой функции достаточно большое:

$$f(x_0) - f(x^*) \geq \hat{\omega} \equiv \tfrac{18}{L_0^2} \tau_f^3 \cdot (L + L_0)^3,$$

то метод демонстрирует сверхлинейную сходимость к области $\mathcal{L}(\hat{\omega})$:

$$\ln \left(\tfrac{1}{\hat{\omega}} (f(x_k) - f(x^*)) \right) \leq \left(\tfrac{2}{3} \right)^k \ln \left(\tfrac{1}{\hat{\omega}} (f(x_0) - f(x^*)) \right). \tag{4.1.33}$$

2. Если $f(x_0) - f(x^*) \leq \gamma^2 \hat{\omega}$ при некотором $\gamma > 1$, то справедлива следующая оценка скорости сходимости:

$$f(x_k) - f(x^*) \leq \hat{\omega} \cdot \frac{\gamma^2 \left(2+\frac{3}{2}\gamma\right)^2}{\left(2+\left(k+\frac{3}{2}\right)\cdot\gamma\right)^2}, \quad k \geq 0. \tag{4.1.34}$$

Доказательство.

Пользуясь неравенствами (4.1.32) и (4.1.29) с $p = 1$, получаем

$$f(x_k) - f(x_{k+1}) \geq \tfrac{L_0 \cdot (f(x_{k+1}) - f(x^*))^{3/2}}{3\sqrt{2} \cdot (L+L_0)^{3/2} \cdot \tau_f^{3/2}} = \hat{\omega}^{-1/2} (f(x_{k+1}) - f(x^*))^{3/2}.$$

Обозначив $\delta_k = (f(x_k) - f(x^*))/\hat{\omega}$, получим

$$\delta_k - \delta_{k+1} \geq \delta_{k+1}^{3/2}. \tag{4.1.35}$$

Поэтому, поскольку $\delta_k \geq 1$, имеем

$$\ln \delta_k \leq \left(\tfrac{2}{3}\right)^k \ln \delta_0,$$

а это и есть неравенство (4.1.33).

Докажем теперь неравенство (4.1.34). Пользуясь оценкой (4.1.35), получаем

$$\begin{aligned}
\tfrac{1}{\sqrt{\delta_{k+1}}} - \tfrac{1}{\sqrt{\delta_k}} \quad & \geq \quad \tfrac{1}{\sqrt{\delta_{k+1}}} - \tfrac{1}{\sqrt{\delta_{k+1} + \delta_{k+1}^{3/2}}} = \tfrac{\sqrt{\delta_{k+1} + \delta_{k+1}^{3/2}} - \sqrt{\delta_{k+1}}}{\sqrt{\delta_{k+1}}\sqrt{\delta_{k+1} + \delta_{k+1}^{3/2}}} \\
& = \quad \tfrac{1}{\sqrt{1+\sqrt{\delta_{k+1}}} \cdot \left(1 + \sqrt{1+\sqrt{\delta_{k+1}}}\right)} = \tfrac{1}{1 + \sqrt{\delta_{k+1}} + \sqrt{1 + \sqrt{\delta_{k+1}}}} \\
& \geq \quad \tfrac{1}{2 + \frac{3}{2}\sqrt{\delta_{k+1}}} \geq \tfrac{1}{2 + \frac{3}{2}\sqrt{\delta_0}}.
\end{aligned}$$

Таким образом, $\frac{1}{\delta_k} \geq \frac{1}{\gamma} + \frac{k}{2 + \frac{3}{2}\gamma}$, а это и есть неравенство (4.1.34). □

Не стоит переоценивать эффективность сверлинейной сходимости, установленной в неравенстве (4.1.33). Эта оценка верна только на первой фазе процесса и описывает сходимость к множеству $\mathcal{L}(\hat{\omega})$. Например, первая фаза процесса, описанного в теореме 4.1.4, даже короче: ее длина составляет одну итерацию.

Рассмотрим теперь градиентно доминируемые функции степени два. Здесь тоже имется две фазы.

Теорема 4.1.7 *Применим метод (4.1.16) для решения задачи минимизации градиентно доминируемой функции $f(x)$ степени $p = 2$.*

1. Если начальное значение целевой функции большое:

$$f(x_0) - f(x^*) \geq \tilde{\omega} \equiv \tfrac{L_0^4}{324(L+L_0)^6 \, \tau_f^3}, \tag{4.1.36}$$

то метод на первой фазе сходится следующим образом:

$$f(x_k) - f(x^*) \leq (f(x_0) - f(x^*)) \cdot e^{-k \cdot \sigma}, \tag{4.1.37}$$

где $\sigma = \frac{\tilde{\omega}^{1/4}}{\tilde{\omega}^{1/4} + (f(x_0) - f(x^*))^{1/4}}$. *Эта фаза заканчивается на первой итерации k_0, для которой неравенство (4.1.36) оказывается невыполненным.*

2. При $k \geq k_0$ скорость сходимости будет сверхлинейной:

$$f(x_{k+1}) - f(x^*) \leq \tilde{\omega} \cdot \left(\frac{f(x_k)-f(x^*)}{\tilde{\omega}} \right)^{4/3}. \qquad (4.1.38)$$

Доказательство.

Пользуясь неравенствами (4.1.32) и (4.1.29) с $p = 2$, получаем

$$f(x_k) - f(x_{k+1}) \geq \frac{L_0 \cdot (f(x_{k+1})-f(x^*))^{3/4}}{3\sqrt{2} \cdot (L+L_0)^{3/2} \cdot \tau_f^{3/4}} = \tilde{\omega}^{1/4}(f(x_{k+1}) - f(x^*))^{3/4}.$$

Обозначив $\delta_k = (f(x_k) - f(x^*))/\tilde{\omega}$, имеем

$$\delta_k \geq \delta_{k+1} + \delta_{k+1}^{3/4}. \qquad (4.1.39)$$

Поэтому

$$\frac{\delta_k}{\delta_{k+1}} \geq 1 + \delta_k^{-1/4} \geq 1 + \delta_0^{-1/4} = \frac{1}{1-\sigma} \geq e^{\sigma},$$

и мы получаем оценку (4.1.37). Наконец, из неравенства (4.1.39) вытекает, что $\delta_{k+1} \leq \delta_k^{4/3}$. Это и есть неравенство (4.1.38). □

Сравнивая утверждение теоремы 4.1.7 с другими теоремами этого раздела, можно увидеть важное различие: мы в первый раз получили полиномиальное вхождение начальной невязки $f(x_0) - f(x^*)$ в оценку сложности первой фазы процесса. Во всех других оценках зависимость от этой невязки гораздо слабее.

Заметим, что класс градиентно доминируемых функций степени два вкладывается в класс градиентно доминируемых функций степени единица. Однако это может только ухудшить оценки эфективности, доказанные в теореме 4.1.7.

Нелинейные преобразования выпуклых функций

Пусть оператор $u(x) : \mathbb{R}^n \to \mathbb{R}^n$ является невырожденным. Обозначим через $v(u)$ обратный оператор:

$$v(u) : \mathbb{R}^n \to \mathbb{R}^n, \quad v(u(x)) \equiv x.$$

Рассмотрим следующую функцию:

$$f(x) = \phi(u(x)),$$

где выпуклая функция $\phi(u)$ имеет ограниченные множества уровня. Такие функции типичны для задач минимизации составных функций. Обозначим через $x^* \equiv v(u^*)$ ее минимум. Зафиксируем произвольную точку $x_0 \in \mathbb{R}^n$. Обозначим

$$\begin{aligned} \sigma &= \max_u \{ \| v'(u) \| : \phi(u) \leq f(x_0) \}, \\ D &= \max_u \{ \| u - u^* \| : \phi(u) \leq f(x_0) \}. \end{aligned}$$

Докажем следующий простой результат.

Лемма 4.1.8 *При любых* $x, y \in \mathcal{L}(f(x_0))$ *выполнено неравенство*

$$\|x - y\| \le \sigma \|u(x) - u(y)\|. \qquad (4.1.40)$$

Доказательство.
Действительно, для $x, y \in \mathcal{L}(f(x_0))$ имеем $\phi(u(x)) \le f(x_0)$ и $\phi(u(y)) \le f(x_0)$. Рассмотрим траекторию $x(t) = v(tu(y) + (1-t)u(x))$, $t \in [0,1]$. Тогда

$$y - x = \int\limits_0^1 x'(t)dt = \left(\int\limits_0^1 v'(tu(y) + (1-t)u(x))dt \right) \cdot (u(y) - u(x))$$

и мы получаем оценку (4.1.40). □

Следующее утверждение очень похоже на теорему 4.1.4.

Теорема 4.1.8 *Предположим, что у функция* f *ее гессиан липшицев на множестве* $\mathcal{F} \supseteq \mathcal{L}(f(x_0))$ *с константой* L. *Пусть последовательность* $\{x_k\}$ *формируется методом (4.1.16). Тогда*

1. если $f(x_0) - f^* \ge \frac{3}{2}L(\sigma D)^3$, *то* $f(x_1) - f^* \le \frac{1}{2}L(\sigma D)^3$;

2. если же $f(x_0) - f^* \le \frac{3}{2}L(\sigma D)^3$, *то метод (4.1.16) сходится следующим образом:*

$$f(x_k) - f(x^*) \le \tfrac{3L(\sigma D)^3}{2(1+\frac{1}{3}k)^2}, \quad k \ge 0. \qquad (4.1.41)$$

Доказательство.
Действительно, в силу неравенства (4.1.11), верхней границы на параметры M_k и определения (4.1.23) для любого $k \ge 0$ выполнено неравенство

$$\begin{aligned} f(x_{k+1}) - f(x^*) \le \min_y [\quad & f(y) - f(x^*) + \tfrac{L}{2}\|y - x_k\|^3 : \\ & y = v(\alpha u^* + (1-\alpha)u(x_k)), \alpha \in [0,1] \;]. \end{aligned}$$

В силу определения точек y в вышеприведенной задаче минимизации и оценки (4.1.40) имеем

$$\begin{aligned} f(y) - f(x^*) &= \phi(\alpha u^* + (1-\alpha)u(x_k)) - \phi(u^*) \le (1-\alpha)(f(x_k) - f(x^*)), \\ \|y - x_k\| &\le \alpha\sigma\|u(x_k) - u^*\| \le \alpha\sigma D. \end{aligned}$$

Это означает, что здесь проходит рассуждение теоремы 4.1.4 с заменой D на σD. □

Теперь докажем утверждение для сильно выпуклой функции ϕ. Обозначим $\breve{\omega} = \frac{1}{L^2}\left(\frac{\gamma}{2\sigma^2}\right)^3$.

Теорема 4.1.9 *Пусть функция ϕ сильно выпукла с параметром выпуклости $\gamma > 0$. Тогда в предположениях теоремы 4.1.8 поведение метода (4.1.16) описывается следующим образом.*

1. Если $f(x_0) - f(x^*) \geq \frac{4}{9}\breve{\omega}$, то на превой фазе процесса наблюдается следующая скорость сходимости:

$$f(x_k) - f(x^*) \leq \left[(f(x_0) - f(x^*))^{1/4} - \tfrac{k}{6}\sqrt{\tfrac{2}{3}}\breve{\omega}^{1/4}\right]^4. \tag{4.1.42}$$

Эта фаза прерывается, как только $f(x_{k_0}) - f(x^*) \leq \frac{4}{9}\breve{\omega}$ для некоторого $k_0 \geq 0$.

2. При $k \geq k_0$ последовательность сходится сверхлинейно:

$$f(x_{k+1}) - f(x^*) \leq \tfrac{1}{2}(f(x_k) - f(x^*))\sqrt{\tfrac{f(x_k)-f(x^*)}{\breve{\omega}}}. \tag{4.1.43}$$

Доказательство.

Действительно, в силу неравенства (4.1.11), верней границы на параметры M_k и определения (4.1.23) для любого $k \geq 0$ имеем

$$f(x_{k+1}) - f(x^*) \leq \min_y [\ f(y) - f(x^*) + \tfrac{L}{2}\|y - x_k\|^3 :$$

$$y = v(\alpha u^* + (1-\alpha)u(x_k)), \alpha \in [0,1]\ ].$$

В силу определения точек y в последней задаче минимизации и оценки (4.1.40) выполнены соотношения

$$\begin{array}{rcl} f(y) - f(x^*) & = & \phi(\alpha u^* + (1-\alpha)u(x_k)) - \phi(u^*) \leq (1-\alpha)(f(x_k) - f(x^*)), \\ \|y - x_k\| & \leq & \alpha\sigma\|u(x_k) - u^*\| \leq \alpha\sigma\sqrt{\tfrac{2}{\gamma}(f(x_0) - f(x^*))}. \end{array}$$

Это означает что можно воспользоваться рассуждением при доказательстве теоремы 4.1.5 с заменой L на $\sigma^3 L$. □

Заметим, что функции рассмотренные в этом разделе, часто используются для тестирования методов невыпуклой оптимизации.

4.1.5 Вычислительные детали

Минимизируя кубическую регуляризацию

Заметим, что вспомогательная задача минимизации (4.1.5), которую надо решить чтобы вычислить точку $T_M(x)$, а именно,

$$\min_{h\in\mathbb{R}^n} \left[\langle g, h\rangle + \tfrac{1}{2}\langle Hh, h\rangle + \tfrac{M}{6}\|h\|^3\right], \tag{4.1.44}$$

является, вообще говоря, невыпуклой. Она может иметь большое количество изолированных локальных минимумов, в то время как нам обязательно нужен глобальный минимум.

Тем не менее, в этом разделе мы покажем, что эта цель может быть достигнута с помощью решения выпуклой задачи одномерной минимизации.

Прежде чем приводить алгоритмическое доказательство этого факта, приведем некоторое объяснение этого феномена. Введем следующие объекты:

$$\begin{array}{rcl} \xi_1(h) & = & \langle g,h\rangle + \frac{1}{2}\langle Hh,h\rangle, \quad \xi_2(h) \; = \; \|h\|^2, \\ Q & = & \left\{ \xi = (\xi^{(1)},\xi^{(2)})^T : \; \xi^{(1)} = \xi_1(h), \; \xi^{(2)} = \xi_2(h), \; h \in \mathbb{R}^n \right\} \subset R^2, \\ \varphi(\xi) & = & \xi^{(1)} + \frac{M}{6} \left(\xi^{(2)} \right)_+^{3/2}. \end{array}$$

где $(a)_+ = \max\{a,0\}$. Тогда

$$\begin{array}{rcl} \min\limits_{h\in\mathbb{R}^n} \left[\langle g,h\rangle + \frac{1}{2}\langle Hh,h\rangle + \frac{M}{6}\|h\|^3 \right] & \equiv & \min\limits_{h\in\mathbb{R}^n} \left[\xi_1(h) + \frac{M}{6}\xi_2^{3/2}(h) \right] \\ & = & \min\limits_{\xi\in Q} \varphi(\xi). \end{array}$$

Теорема 2.2 в [111] гарантирует что при $n \geq 2$ множество Q является *выпуклым и замкнутым.* Таким образом, мы свели исходную задачу невыпуклой минимизации в $\mathbb{R}^n$ к выпуклой условной задаче минимизации в $\mathbb{R}^2$. Пока это сведение неконструктивно, так как множество Q задается в неявной форме. Однако следующий результат показывает, что описание этого множества не слишком сложное.

Обозначим

$$v_u(h) = \langle g,h\rangle + \frac{1}{2}\langle Hh,h\rangle + \frac{M}{6}\|h\|^3, \quad h \in \mathbb{R}^n,$$

и

$$v_l(r) = -\frac{1}{2}\langle \left(H + \frac{Mr}{2}I\right)^{-1} g, g\rangle - \frac{M}{12}r^3.$$

Для первой функции мы иногда пользуемся обозначением $v_u(g;h)$. Обозначим

$$\mathcal{D} = \{ r \in R : \; H + \frac{M}{2}rI \succ 0, \; r \geq 0 \}.$$

Теорема 4.1.10 *При любом $M > 0$ справедливо следующее соотношение:*

$$\min_{h\in\mathbb{R}^n} v_u(h) = \sup_{r\in\mathcal{D}} v_l(r). \tag{4.1.45}$$

Для любого $r \in \mathcal{D}$ направление $h(r) = -\left(H + \frac{Mr}{2}I\right)^{-1} g$ удовлетворяет уравнению

$$0 \leq v_u(h(r)) - v_l(r) = \frac{M}{12}(r + 2\|h(r)\|)(\|h(r)\| - r)^2 = \frac{4}{3M} \cdot \frac{r+2\|h(r)\|}{(r+\|h(r)\|)^2} \cdot v_l'(r)^2. \tag{4.1.46}$$

Доказательство.

Обозначим левую часть соотношения (4.1.45) через v_u^*, а ее правую часть – через v_l^*. Покажем, что $v_u^* \geq v_l^*$. Действительно,

$$\begin{aligned}
v_u^* &= \min_{h\in\mathbb{R}^n} \left[\langle g,h\rangle + \tfrac{1}{2}\langle Hh,h\rangle + \tfrac{M}{6}\|h\|^3\right] \\
&= \min_{\substack{h\in\mathbb{R}^n,\\ \tau=\|h\|^2}} \left[\langle g,h\rangle + \tfrac{1}{2}\langle Hh,h\rangle + \tfrac{M}{6}(\tau)_+^{3/2}\right] \\
&= \min_{\substack{h\in\mathbb{R}^n,\\ \tau\in R}} \sup_{r\in R} \left[\langle g,h\rangle + \tfrac{1}{2}\langle Hh,h\rangle + \tfrac{M}{6}(\tau)_+^{3/2} + \tfrac{M}{4}r\left(\|h\|^2-\tau\right)\right] \\
&\geq \sup_{r\in\mathcal{D}} \min_{\substack{h\in\mathbb{R}^n,\\ \tau\in R}} \left[\langle g,h\rangle + \tfrac{1}{2}\langle Hh,h\rangle + \tfrac{M}{6}(\tau)_+^{3/2} + \tfrac{M}{4}r\left(\|h\|^2-\tau\right)\right] \equiv v_l^*.
\end{aligned}$$

Рассмотрим теперь произвольное $r \in \mathcal{D}$. Тогда

$$g = -Hh(r) - \tfrac{M}{2}rh(r).$$

Таким образом,

$$\begin{aligned}
v_u(h(r)) &= \langle g,h(r)\rangle + \tfrac{1}{2}\langle Hh(r),h(r)\rangle + \tfrac{M}{6}\|h(r)\|^3 \\
&= -\tfrac{1}{2}\langle Hh(r),h(r)\rangle - \tfrac{M}{2}r\|h(r)\|^2 + \tfrac{M}{6}\|h(r)\|^3 \\
&= -\tfrac{1}{2}\langle\left(H+\tfrac{Mr}{2}I\right)h(r),h(r)\rangle - \tfrac{M}{4}r\|h(r)\|^2 + \tfrac{M}{6}\|h(r)\|^3 \\
&= v_l(r) + \tfrac{M}{12}r^3 - \tfrac{M}{4}r\|h(r)\|^2 + \tfrac{M}{6}\|h(r)\|^3 \\
&= v_l(r) + \tfrac{M}{12}(r+2\|h(r)\|)\cdot(\|h(r)\|-r)^2.
\end{aligned}$$

Следовательно, мы получаем уравнение (4.1.46).

Заметим, что

$$v_l'(r) = \tfrac{M}{4}(\|h(r)\|^2 - r^2).$$

Таким образом, если оптимальное значение v_l^* достигается при некотором $r^* > 0$ из $\mathcal{D}$, то $v_l'(r^*) = 0$ и с помощью уравнения (4.1.46) мы заключаем, что $v_r^* = v_l^*$. Если $r^* = \frac{2}{M}(-\lambda_n(H))_+$, то равенство (4.1.45) можно обосновать из соображений непрерывности (поскольку функция $v_u^* \equiv v_u^*(g)$ вогнута по $g \in \mathbb{R}^n$; этот вопрос мы еще обсудим). □

Заметим что Предложение 1 (см. (4.1.8)) следует из определения множества $\mathcal{D}$.

Теорема 4.1.10 показывает, что в невырожденной ситуации решение задачи (4.1.45) может быть найдено с помощью решения одномерного уравнения

$$r = \|\left(H+\tfrac{Mr}{2}I\right)^{-1}g\|, \quad r \geq \tfrac{2}{M}(-\lambda_n(H))_+. \qquad (4.1.47)$$

Техника решения таких уравнений очень хорошо разработана в рамках методов доверительных областей (для исчерпывающей презентации имеющихся подходов см. гл. 7 в монографии [45]). По сравнению с (4.1.47), уравнения возникающие в этих методах имеют постоянную левую часть. Но конечно же основные вычислительные сложности при решении таких уравнений связаны с нелинейной выпуклой правой частью.

Для полноты изложения обсудим структуру уравнения (4.1.47). Переходя к базису собственных векторов матрицы H, это уравнение можно записать как

$$r^2 = \sum_{i=1}^{n} \frac{\tilde{g}_i^2}{(\lambda_i + \frac{M}{2} r)^2}, \quad r \geq \frac{2}{M}(-\lambda_n)_+, \tag{4.1.48}$$

где λ_i - это собственные значения матрицы H, а $\tilde{g}_i$ - координаты вектора g в новом базисе.

Если $\tilde{g}_n \neq 0$, то решение r^* уравнения (4.1.48) принадлежит внутренности области определения:

$$r > \frac{2}{M}(-\lambda_n)_+,$$

и мы можем вычислить смещение $h(r^*)$ по явной формуле

$$h(g; r^*) = -\left(H + \frac{Mr^*}{2} I\right)^{-1} g.$$

Если же $\tilde{g}_n = 0$, то эта формула не работает и надо рассматривать различные случаи. Чтобы избежать все эти осложнения, можно воспользоваться следующим простым результатом.

Лемма 4.1.9 *Пусть* $\tilde{g}_n = 0$. *Положим* $g(\epsilon) = \tilde{g} + \epsilon e_n$, *где* e_n *- это* n*-й координатный вектор. Обозначим через* $r^*(\epsilon)$ *решение уравнения (4.1.48) с* $\tilde{g} = g(\epsilon)$. *Тогда любая предельная точка траектории*

$$h(g(\epsilon); r^*(\epsilon)), \quad \epsilon \to 0,$$

является глобальным минимумом по h *функции* $v_u(g; h)$.

Доказательство.

Действительно, функция $v_u^*(g)$ вогнута по $g \in \mathbb{R}^n$. Следовательно, она непрерывна. Поэтому

$$v_u^*(g) = \lim_{\epsilon \to 0} v_u^*(g(\epsilon)) = \lim_{\epsilon \to 0} v_u(g(\epsilon); h(g(\epsilon); r^*(\epsilon))).$$

Остается заметить, что функция $v_u(g; h)$ непрерывна по совокупности аргументов. □

Для того, чтобы проиллюстрировать возможные сложности, связанные с двойственной задачей, посмотрим на следующий пример.

Пример 4.1.4 Пусть $n = 2$ и

$$\tilde{g} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \lambda_1 = 0, \quad \lambda_2 = -1, \quad M = 1.$$

Таким образом, наша прямая задача записывается так:

$$\min_{h\in\mathbb{R}^2}\left\{\psi(h)\equiv -h^{(1)}-\tfrac{1}{2}\left(h^{(2)}\right)^2+\tfrac{1}{6}\left[\sqrt{\left(h^{(1)}\right)^2+\left(h^{(2)}\right)^2}\right]^3\right\}.$$

Ввиду (4.1.6), нам нужно решить следующую систему нелинейных уравнений:

$$\begin{aligned}\tfrac{h^{(1)}}{2}\sqrt{\left(h^{(1)}\right)^2+\left(h^{(2)}\right)^2} &= 1,\\ \tfrac{h^{(2)}}{2}\sqrt{\left(h^{(1)}\right)^2+\left(h^{(2)}\right)^2} &= h^{(2)}.\end{aligned}$$

Таким образом, у нас есть три кандидата на глобальный минимум:

$$h_1^*=\begin{pmatrix}\sqrt{2}\\ 0\end{pmatrix},\quad h_2^*=\begin{pmatrix}1\\ \sqrt{3}\end{pmatrix},\quad h_3^*=\begin{pmatrix}1\\ -\sqrt{3}\end{pmatrix}.$$

С помощью непосредственной подстановки убеждаемся, что

$$\psi(h_1^*)=-\tfrac{2\sqrt{2}}{3}\ >\ -\tfrac{7}{6}=\psi(h_2^*)=\psi(h_3^*).$$

Таким образом, оба вектора h_2^* и h_3^* являются глобальными минимумами.

Посмотрим теперь на двойственную задачу:

$$\sup_r\left[\phi(r)\equiv -\tfrac{r^3}{12}-\tfrac{1}{2}\cdot\tfrac{1}{0+\frac{1}{2}r}=-\tfrac{r^3}{12}-\tfrac{1}{r}\ :\ -1+\tfrac{1}{2}r>0\right].$$

Заметим, что $\phi'(r)=-\frac{r^2}{4}+\frac{1}{r^2}$. Таким образом, $\phi'(2)=-\frac{3}{4}<0$ и мы заключаем, что

$$r^*=2,\quad \phi^*=-\tfrac{7}{6}.$$

Однако r^* не удовлетворяет уравнению $\phi'(r)=0$ и значение $h(r^*)$ не определено. □

В заключение этого раздела выпишем явное решение прямой задачи (4.1.45) в терминах собственных значений матрицы H. Обозначим через $\{s_i\}_{i=1}^n$ ортогональный базис собственных векторов матрицы H, и пусть индекс $\hat{k}$ удовлетворяет условиям

$$\begin{aligned}\tilde{g}^{(i)} &\neq 0 \ \text{ for } i<\hat{k},\\ \tilde{g}^{(i)} &= 0 \ \text{ for } i\geq\hat{k}.\end{aligned}$$

Пусть r^* – решение двойственной задачи (4.1.45). Тогда решение прямой задачи дается вектором

$$h^*=-\sum_{i=1}^{\hat{k}-1}\frac{\tilde{g}^{(i)}s_i}{\lambda_i+\frac{M}{2}r^*}+\sigma s_n,$$

где величина σ выбрана в соответствии с условием $\|h^*\|=r^*$. Заметим, что это правило работает как в случае $\hat{k}=1$, так и в случае $\hat{k}=n+1$.

Стратегии одномерного поиска

Обсудим возможную стоимость шага 1 в методе (4.1.16), который состоит в нахождении величины $M_k \in [L_0, 2L]$, удовлетворяющей условию

$$f(T_{M_k}(x)) \le \bar{f}_{M_k}(x_k).$$

Заметим, что для $M_k \ge L$ это неравенство выполнено. Рассмотрим следующую стратегию.

$$\textbf{Пока } f(T_{M_k}(x)) > f(x_k) \textbf{ повторяем } \; M_k := 2M_k; \quad \textbf{Выбираем } M_{k+1} := M_k. \tag{4.1.49}$$

Если метод (4.1.16) стартует с любым значением $M_0 \in [L_0, 2L]$, то эта процедура, применяемая на каждой итерации, имеет следующие преимущества:

- $M_k \le 2L$.
- Общее число дополнительных вычислений отображения $T_{M_k}(x)$ в течение всего процесса (4.1.16) ограничено величиной

  $$\log_2 \tfrac{2L}{L_0}.$$

  Эта граница не зависит от числа итераций основного процесса.

Однако может случиться, что правило (4.1.49) будет слишком консервативным. Действительно, в нем мы можем только увеличивать нашу оценку для константы L и никогда не можем ее уменьшать. Это может привести к сильному замедлению всего процесса. Более гибкая стратегия выглядит так:

$$\begin{aligned} &\textbf{Пока } f(T_{M_k}(x)) > f(x_k) \textbf{ повторяем } \; M_k := 2M_k; \\ &\textbf{Выбираем } x_{k+1} := T_{M_k}(x_k); \quad M_{k+1} := \max\{\tfrac{1}{2}M_k, L_0\}. \end{aligned} \tag{4.1.50}$$

Тогда несложно доказать по индукции (см. лемму 2.3.4), что N_k, общее число вычислений точек $T_M(x)$ сделанных в методе (4.1.50) в течение первых k итераций, удовлетворяет следующему неравенству:

$$N_k \le 2k + \log_2 \tfrac{M_k}{L_0}.$$

Таким образом, если N – число итераций в процессе, то надо вычислить не более

$$2N + \log_2 \tfrac{2L}{L_0}$$

точек $T_M(x)$. Эти затраты выглядят вполне умеренными по сравнению с выигрышем, достигаемым за счет использования длинных шагов.

4.1.6 Обсуждение

Сравним результаты этого раздела с известными фактами об оценках эффективности других методов. Поскольку для невыпуклых функций такие результаты практически отсутствуют, ограничимся выпуклыми задачами.

Пусть функция $f(x)$ является сильно выпуклой на $\mathbb{R}^n$ с параметром $\gamma > 0$ (см. (4.1.30)). В этом случае у нее существует единственный минимум x^* и условие (4.1.25) выполняется при всех $x \in \mathbb{R}^n$ (см., например, [83], Section 2.1.3). Предположим также что гессиан функции $f(x)$ липшицев:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L\|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

Получим для таких функций оценки эффективности метода (4.1.16), воспользовавшись результатами теорем 4.1.4 и 4.1.5.

Зафиксируем произвольную точку $x_0 \in \mathbb{R}^n$. Обозначим через D *радиус* соответствующего множества уровня:

$$D = \max_x \{\|x - x^*\| : \ f(x) \le f(x_0)\}.$$

Из соотношения (4.1.25) получаем

$$D \le \left[\tfrac{2}{\gamma}(f(x_0) - f(x^*))\right]^{1/2}.$$

Покажем, что естественной мерой качества стартовой точки x_0 является следующая характеристика:

$$\kappa \equiv \kappa(x_0) = \tfrac{LD}{\gamma}.$$

Введем три уровня переключения:

$$\omega_0 = \tfrac{\gamma^3}{18L^2} \equiv \tfrac{4}{9}\bar{\omega}, \quad \omega_1 = \tfrac{3}{2}\gamma D^2, \quad \omega_2 = \tfrac{3}{2}LD^3.$$

В силу Теремы 4.1.4 уровень $f(x_0) - f(x^*) \le \frac{1}{2}LD^3$ достигается за один шаг. Таким образом, без ограничения общности можно считать, что

$$f(x_1) - f(x^*) \le \omega_2.$$

Предположим также, что нам нужна очень высокая точность решения. Заметим, что случай $\kappa \le 1$ не представляет никаких трудностей поскольку первая же итерация метода (4.1.16) приводит нас в непосредственную близость к области сверхлинейной сходимости (см. п. 2 теоремы 4.1.5).

Поэтому рассмотрим случай $\kappa \ge 1$. Тогда $\omega_0 \le \omega_1 \le \omega_2$. Оценим продолжительность следующих фаз:

$$\begin{array}{rcccl} \text{Фаза 1:} \ \omega_1 & \le & f(x_i) & \le & \omega_2, \\[2em] \text{Фаза 2:} \ \omega_0 & \le & f(x_i) & \le & \omega_1, \\[2em] \text{Фаза 3:} \ \ \epsilon & \le & f(x_i) & \le & \omega_0. \end{array}$$

В силу теоремы 4.1.4 продолжительность k_1 первой фазы ограничена следующим образом:

$$\omega_1 \leq \frac{3LD^3}{2(1+\frac{1}{3}k_1)^2}.$$

Таким образом, $k_1 \leq 3\sqrt{\kappa}$. Далее, в силу п. 1 теоремы 4.1.5 мы можем оценить сверху продолжительность k_2 второй фазы:

$$\omega_0^{1/4} \leq (f(x_{k_1+1}) - f(x^*))^{1/4} - \frac{k_2}{6}\omega_0^{1/4} \leq (\frac{1}{2}\gamma D^2)^{1/4} - \frac{k_2}{6}\omega_0^{1/4}.$$

Это дает следующую оценку: $k_2 \leq 3^{3/4}\, 2^{1/2}\sqrt{\kappa} \leq 3,25\sqrt{\kappa}$.

Наконец, обозначим $\delta_k = \frac{1}{4\omega_0}(f(x_k) - f(x^*))$. В силу неравенства (4.1.27) имеем

$$\delta_{k+1} \leq \delta_k^{3/2}, \quad k \geq \bar{k} \equiv k_1 + k_2 + 1.$$

В то же время $f(x_{\bar{k}}) - f(x^*) \leq \omega_0$. Таким образом, $\delta_{\bar{k}} \leq \frac{1}{4}$ и верхняя граница на продолжительность k_3 последней фазы может быть найдена из неравенства

$$4^{\left(\frac{3}{2}\right)^{k_3}} \leq \frac{4\omega_0}{\epsilon}.$$

Это дает $k_3 \leq \log_{\frac{3}{2}} \log_4 \frac{2\gamma^3}{9\epsilon L^2}$. Объединяя вместе все границы, получаем, что общее число шагов N метода (4.1.16) ограничено следующим образом:

$$N \leq 6.25\sqrt{\frac{LD}{\gamma}} + \log_{\frac{3}{2}}\left(\log_4 \frac{1}{\epsilon} + \log_4 \frac{2\gamma^3}{9L^2}\right). \qquad (4.1.51)$$

Интересно, что в оценке (4.1.51) параметры нашей задачи аддитивно взаимодействуют с требуемой точностью. Напомним, что обычное такое взаимодействие мультипликативно. Давайте, например, оценим сложность нашей задачи для быстрого градиентного метода (1.2.29), применяемого для минимизации сильно выпуклой функции с липшицевым градиентом. Обозначим через $\hat{L}$ наибольшее собственное значение матрицы $\nabla^2 f(x^*)$. Тогда можно гарантировать, что

$$\gamma I \leq \nabla^2 f(x) \leq (\hat{L} + LD)I \quad \forall x,\ \|x - x^*\| \leq D.$$

Таким образом, мы получим оценку эффективности оптимального метода порядка

$$O\left(\sqrt{\frac{\hat{L}+LD}{\gamma}} \ln \frac{(\hat{L}+LD)D^2}{\epsilon}\right)$$

итераций. Для градиентного метода (1.2.7) оценка гораздо хуже:

$$O\left(\frac{\hat{L}+LD}{\gamma} \ln \frac{(\hat{L}+LD)D^2}{\epsilon}\right).$$

В результате мы получаем, что оценки эффективности модифицированного метода Ньютона (4.1.16) намного лучше оценок градиентных методов. В то же время, необходимо конечно же помнить о существенной разнице в трудоемкости одной итерации этих методов.

Заметим, что аналогичные оценки могут быть получены для некоторых классов невыпуклых задач. Например, для задач, получающихся с помощью нелинейных преобразований выпуклых функций (см. раздел 4.1.4), оценка эффективности будет следующей:

$$N \leq 6.25\sqrt{\tfrac{\sigma}{\gamma}LD} + \log_{\frac{3}{2}}\left(\log_4 \tfrac{1}{\epsilon} + \log_4 \tfrac{2\gamma^3}{9\sigma^6 L^2}\right). \qquad (4.1.52)$$

В заключение укажем, что в методе (4.1.16) можно найти элементы схемы Левенберга-Маркварта (см. соотношение (4.1.8)), или элементы доверительных областей (см. теорему 4.1.10), или стратегию одномерного поиска (см. правила шага 1 в (4.1.16)). Однако все эти элементы возникают естественным образом как *следствия* основной идеи метода, заключающейся в том что следующая точка процесса выбирается из условия глобального минимума верхней кубической аппроксимации целевой функции.

4.2 Ускоренная кубическая регуляризация метода Ньютона

4.2.1 Введение

Мотивировка

В параграфе 4.1 мы рассмотрели кубическую регуляризацию метода Ньютона (КРН), которая может применяться к задачам нахождения локального минимума общих невыпуклых функций. Основное достоинство этой схемы состоит в ее естественной геометрической интерпретации: на каждой итерации мы минимизируем *кубическую модель*, которая оказывается *глобальной верхней оценкой* для целевой функции. Это основополагающее свойство обеспечивает предсказуемое и вполне естественное поведение всего процесса.

До сих пор мы уделяли основное внимание различным классам *невыпуклых задач*. Для выпуклых функций многие из наших утверждений могут быть существенно усилены. Более того, можно задействовать стандартную технику ускорения методов первого порядка для того чтобы ускорить КРН.

В этом параграфе мы предполагаем, что целевая функция задачи безусловной минимизации имеет липшицев гессиан. Как было показано в параграфе 4.1, глобальная скорость сходимости КРН на этом классе задач имеет порядок $O(\frac{1}{k^2})$, где k – счетчик числа итераций. Однако заметим, что КРН - это *локальный одношаговый* метод второго порядка. Из теории методов первого порядка для гладких выпуклых функций мы знаем, что скорость сходимости таких методов (например, простейший *градиентный метод*, см. п. 1.2.1) может быть улучшена с $O(\frac{1}{k})$ до $O(\frac{1}{k^2})$ с помощью введения *многошаговых* стратегий. В этом параграфе мы покажем, что аналогичный прием применим и к КРН. В результате мы получим многошаговую схему второго порядка со скоростью сходимости $O(\frac{1}{k^3})$.

Содержание

В п. 4.2.2 мы обосновываем несколько полезных свойств КРН применительно к выпуклым функциям. В п. 4.2.3 мы получаем ускоренную многошаговую версию КРН. На нашем классе задач она сходится как $O(\frac{1}{k^3})$. Это ускорение достигается засчет модификации техники *оценочных последовательностей* (см. п. 1.2.2)

В п. 4.2.4 мы рассматриваем новый класс задач, которые могу трактоваться как невырожденные задачи для методов второго порядка. Этот класс составляют функции из нашего базового класса, которые к тому же являются равномерно выпуклыми *степени три*. На этом классе задач, как КРН, так и его ускоренная версия показываю глобальную линейную скорость сходимости, которая приводит к оценке трудоемкости, пропорциональной произведению некоторой степени *числа обусловленности второго порядка* и логарифма требуемой точности. Мы показываем, что ускоренная схема с подходящей стратегией обновления существенно лучше, чем КРН. Результаты этого пункта подтверждают, что понятие невырожденности или “плохой обусловленности” зависит от типа применяемых методов. Стандартная классификация в случае методов второго порядка просто не работает.

В п. 4.2.5 мы анализируем эффективность предложенных методов на классе сильно выпуклых функций. Показывается, что основные вычислительные затраты приходятся на первую фазу работы методов, которая необходима для нахождения точки из области квадратичной сходимости метода. В некотором смысле, сложность таких задач практически не зависит от требуемой точности. Такая ситуация легко может привести к ошибочным заключениям об эффективности некоторых стратегий. В п. 4.2.6 мы приводим пример такого метода, работающего с сильно выпуклыми функциями. Формально он сходится со скоростью $O(\frac{1}{k^8})$. Однако его реальная эффективность хуже чем у ускоренного КРН.

Наконец, в последнем п. 4.2.7 мы обсуждаем полученные результаты.

Обозначения

Зафиксируем некоторый положительно определенный самосопряженный линейный оператор $B : E \to E^*$. Зададим евклидовы нормы:

$$\begin{aligned} \|h\| &= \langle Bh, h\rangle^{1/2}, \quad h \in E, \\ \|s\|_* &= \langle s, B^{-1}s\rangle^{1/2}, \quad s \in E^*, \\ \|A\| &= \max_{\|h\|\le 1} \|Ah\|_*, \quad A : E \to E^*. \end{aligned}$$

Для линейного оператора $A = A^*$ ту же норма можно определить и так:

$$\|A\| = \max_{\|h\|\le 1} |\langle Ah, h\rangle|. \tag{4.2.1}$$

Заметим, что любой вектор $s \in E^*$ порождает одноранговый самосопряженный оператор $ss^* : E \to E^*$, действующий по правилу

$$ss^* \cdot x \ = \ \langle s, x \rangle \cdot s, \quad x \in E.$$

Мы определяем значение оператора $A(s) = \frac{ss^*}{\|s\|_*}$ в нуле по непрерывности: $A(0) = 0$.

В этом параграфе, как правило, рассматриваются функции с липшицевым гессианом:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \ \leq \ L_3 \|x - y\|, \quad x, y \in E, \tag{4.2.2}$$

где $L_3 \overset{\text{def}}{=} L_3(f)$. Следовательно, при всех x и y из E выполнено неравенство

$$\|\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x)\|_* \ \overset{(1.1.15)}{\leq} \ \tfrac{1}{2} L_3 \|y - x\|^2. \tag{4.2.3}$$

Более того, у квадратичной модели

$$f_2(x; y) \overset{\text{def}}{=} f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle$$

можно ограничить невязку по функции

$$|f(y) - f_2(x; y)| \ \overset{(1.1.16)}{=} \ \tfrac{L_3}{6} \|y - x\|^3, \quad x, y \in E. \tag{4.2.4}$$

Мы будем часто использовать равномерно выпуклую функцию $d_3(x) = \frac{1}{3}\|x - x_0\|^2$, у которой $\nabla d_3(x) = \|x - x_0\| \cdot B(x - x_0)$. В силу леммы 1.1.3, имеем $L_3 = L_3(d_3) = 2$, и в силу неравенства (1.1.24) и леммы 1.1.6 заключаем, что $\sigma_3(d_3) = \frac{1}{2}$.

В этом параграфе мы часто описываем сложность различных классов задач с помощью *числа обусловленности* определенного порядка:

$$\gamma_p(f) \ \overset{\text{def}}{=} \ \tfrac{\sigma_p(f)}{L_p(f)}, \quad p \geq 2. \tag{4.2.5}$$

Понятно, например, что $\gamma_2(d_2) = 1$. С другой стороны, мы видели, что $\gamma_3(d_3) = \frac{1}{4}$.

4.2.2 Итерация кубической регуляризации метода Ньютона

В этом пункте приводятся наиболее важные свойства кубической регуляризации метода Ньютона, учитывающие выпуклость целевой функции.

Рассмотрим следующую задачу:

$$\min_{x \in E} \ f(x), \tag{4.2.6}$$

где выпуклая дважды дифференцируемая функция f имеет липшицев гессиан. Как и в параграфе 4.1, введем следующее отображение

$$T_M(x) \ \overset{\text{def}}{=} \ \operatorname{Arg}\min_{y \in E} \left[\hat{f}_M(x; y) \overset{\text{def}}{=} f_2(x; y) + \tfrac{M}{6} \|y - x\|^3 \right]. \tag{4.2.7}$$

Заметим, что точка $T = T_M(x)$ является решением следующей нелинейной системы:

$$\nabla f(x) + \nabla^2 f(x)(T - x) + \tfrac{1}{2} M \cdot \|T - x\| \cdot B(T - x) = 0. \tag{4.2.8}$$

Обозначим $r_M(x) = \|x - T_M(x)\|$. Тогда

$$\begin{array}{rcl} \|\nabla f(T)\|_* & \overset{(4.2.8)}{=} & \|\nabla f(T) - \nabla f(x) - \nabla^2 f(x)(T-x) - \frac{M}{2} r_M(x) B(T-x)\|_* \\ & \overset{(4.2.3)}{\leq} & \frac{L_3+M}{2} r_M^2(x). \end{array} \tag{4.2.9}$$

Далее, умножая (4.2.8) на $T - x$, получаем

$$\langle \nabla f(x), x - T\rangle \;=\; \langle \nabla^2 f(x)(T-x), T-x\rangle + \tfrac{1}{2} M r_M^3(x). \tag{4.2.10}$$

Пусть $M \geq L_3$. Тогда в силу (4.2.4) выполнены соотношения

$$\begin{array}{rcl} f(x) - f(T) & \geq & f(x) - \hat{f}_M(x;T) \\ & = & \langle \nabla f(x), x-T\rangle - \frac{1}{2}\langle \nabla^2 f(x)(T-x), T-x\rangle - \frac{M}{6} r_M^3(x) \\ & = & \frac{1}{2}\langle \nabla^2 f(x)(T-x), T-x\rangle + \frac{M}{3} r_M^3(x). \end{array} \tag{4.2.11}$$

В частности, поскольку функция f выпукла, имеем

$$f(x) - f(T) \overset{(4.2.11)}{\geq} \tfrac{M}{3} r_M^3(x) \overset{(4.2.9)}{\geq} \tfrac{M}{3}\left(\tfrac{2}{L_3+M}\|\nabla f(T)\|_*\right)^{3/2}. \tag{4.2.12}$$

Иногда удобнее использовать следующую функциональную интерпретацию этого шага:

$$f(T) \overset{(M \geq L_3)}{\leq} \min_y \left[f_2(x;y) + \tfrac{M}{6}\|y-x\|^3\right] \overset{(4.2.4)}{\leq} \min_y \left[f(y) + \tfrac{L_3+M}{6}\|y-x\|^3\right]. \tag{4.2.13}$$

Наконец, докажем следующий результат.

Лемма 4.2.1 *Если $M \geq 2L_3$, то выполнено неравенство*

$$\langle \nabla f(T), x - T\rangle \;\geq\; \sqrt{\tfrac{2}{L_3+M}} \cdot \|\nabla f(T)\|_*^{3/2}. \tag{4.2.14}$$

Доказательство.

Положим $T = T_M(x)$ и $r = r_M(x)$. Тогда

$$\begin{array}{rcl} \frac{1}{4}L_3^2 r^4 \;=\; \left(\frac{L_3}{2}\|T-x\|^2\right)^2 & \overset{(4.2.3)}{\geq} & \|\nabla f(T) - \nabla f(x) - \nabla^2 f(x)(T-x)\|_*^2 \\ & \overset{(4.2.8)}{=} & \|\nabla f(T) + \frac{1}{2} M \cdot r \cdot B(T-x)\|_*^2 \\ & = & \|\nabla f(T)\|_*^2 + Mr\langle \nabla f(T), T-x\rangle + \frac{1}{4}M^2 r^4. \end{array}$$

Поэтому

$$\langle \nabla f(T), x - T\rangle \;\geq\; \tfrac{1}{Mr}\|\nabla f(T)\|_*^2 + \tfrac{1}{4M}(M^2 - L_3^2) r^3. \tag{4.2.15}$$

Пользуясь предположениям леммы, можно оценить производную по r правой части последнего неравенства:

$$-\tfrac{1}{Mr^2}\|\nabla f(T)\|_*^2 + \tfrac{3r^2}{4M}(M^2 - L_3^2) \ \geq\ -\tfrac{1}{Mr^2}\|\nabla f(T)\|_*^2 + \left(\tfrac{L_3+M}{2}\right)^2 \tfrac{r^2}{M} \ \overset{(4.2.9)}{\geq}\ 0.$$

Таким образом, его минимум достигается в граничной точке $r = \left[\frac{2}{L_3+M}\|\nabla f(T)\|_*\right]^{1/2}$ допустимого луча, задаваемого неравенством (4.2.9). Подставив это значение в оценку (4.2.15), получаем неравенство (4.2.14). $\Box$

В заключение этого пункта, оценим скорость сходимости метода КРН применительно к нашей основной задаче (4.2.6). Мы предполагаем, что ее решение x^* существует и что константа Липшица L_3 для гессиана целевой функции нам известна. Таким образом, в методе повторяются следующие действия:

$$x_{k+1} \ =\ T_{L_3}(x_k), \quad k = 0, 1, \ldots. \tag{4.2.16}$$

Используя те же аргументы, что и в предыдущем разделе, можно доказать следующее утверждение.

Теорема 4.2.1 *Предположим что множество уровня задачи (4.2.6) ограничено:*

$$\|x - x^*\| \leq D \quad \forall x:\ f(x) \leq f(x_0). \tag{4.2.17}$$

Если последовательность $\{x_k\}_{k=1}^{\infty}$ *сформирована методом (4.2.16), то справедлива оценка*

$$f(x_k) - f(x^*) \ \leq\ \tfrac{9L_3D^3}{(k+4)^2}, \quad k \geq 1. \tag{4.2.18}$$

Доказательство.
В силу неравенства (4.2.11) имеем $f(x_{k+1}) \leq f(x_k)$, $k \geq 0$. Таким образом, $\|x_k - x^*\| \leq D$ при всех $k \geq 0$. Далее, в силу (4.2.13) получаем, что

$$f(x_1) \ \leq\ f(x^*) + \tfrac{L_3}{3}D^3. \tag{4.2.19}$$

Рассмотрим теперь произвольное $k \geq 1$. Обозначим $x_k(\tau) = x^* + (1-\tau)(x_k - x^*)$. В силу (4.2.13) для любого $\tau \in [0,1]$ имеем

$$f(x_{k+1}) \ \leq\ f(x_k(\tau)) + \tau^3\tfrac{L_3}{3}\|x_k - x^*\|^3 \ \leq\ f(x_k) - \tau(f(x_k) - f(x^*)) + \tau^3\tfrac{L_3D^3}{3}$$

Минимум правой части этого неравенства достигается при

$$\tau \ =\ \sqrt{\tfrac{f(x_k)-f(x^*)}{L_3D^3}} \ \leq\ \sqrt{\tfrac{f(x_1)-f(x^*)}{L_3D^3}} \ \overset{(4.2.19)}{<}\ 1.$$

Таким образом, для любого $k \geq 1$ имеем

$$f(x_{k+1}) \ \leq\ f(x_k(\tau)) - \tfrac{2}{3}\cdot\tfrac{(f(x_k)-f(x^*))^{3/2}}{\sqrt{L_3D^3}}. \tag{4.2.20}$$

Обозначим $\delta_k = f(x_k) - f(x^*)$. Тогда

$$\frac{1}{\sqrt{\delta_{k+1}}} - \frac{1}{\sqrt{\delta_k}} \;=\; \frac{\delta_k - \delta_{k+1}}{\sqrt{\delta_k \delta_{k+1}}(\sqrt{\delta_k} + \sqrt{\delta_{k+1}})} \;\overset{(4.2.20)}{\geq}\; \frac{2}{3\sqrt{L_3 D^3}} \cdot \frac{\delta_k}{\sqrt{\delta_{k+1}}(\sqrt{\delta_k} + \sqrt{\delta_{k+1}})} \;\geq\; \frac{1}{3\sqrt{L_3 D^3}}.$$

Таким образом, для любого $k \geq 1$ выполнены неравенства

$$\frac{1}{\sqrt{\delta_k}} \;\geq\; \frac{1}{\sqrt{\delta_1}} + \frac{k-1}{3\sqrt{L_3 D^3}} \;\overset{(4.2.19)}{\geq}\; \frac{1}{\sqrt{L_3 D^3}} \cdot \left(\sqrt{3} + \frac{k-1}{3}\right) \;\geq\; \frac{k+4}{3\sqrt{L_3 D^3}}. \qquad \square$$

4.2.3 Ускоренный метод

Для ускорения метода (4.2.16) мы применим модифицированную технику *оценочных последовательностей*, которая использовалась в п. 1.2.2 как средство ускорения обычного градиентного метода. В нашей ситуации эта идея может быть использована двумя различными способами.

1. Линейные оценочные функции. Для решения задачи (4.2.6), мы рекурсивно пересчитываем следующие последовательности:

- последовательность оценочных функций

  $$\psi_k(x) \;=\; l_k(x) + \frac{N}{6}\|x - x_0\|^3, \quad k = 1, 2, \ldots,$$

  где $l_k(x)$ – линейные функции от $x \in E$ и N – положительный параметр;

- минимизирующая последовательность точек $\{x_k\}_{k=1}^\infty$;

- последовательность масштабирующих параметров $\{A_k\}_{k=1}^\infty$:

  $$A_{k+1} \overset{\mathrm{def}}{=} A_k + a_k, \quad k = 1, 2, \ldots.$$

Для этих объектов мы поддерживаем следующие соотношения:

$$\left.\begin{array}{lrcl} \mathcal{R}_k^1: & A_k f(x_k) & \leq & \psi_k^* \;\equiv\; \min\limits_x \psi_k(x), \\[2ex] \mathcal{R}_k^2: & \psi_k(x) & \leq & A_k f(x) + \frac{2L_3+N}{6}\|x - x_0\|^3, \ \forall x \in E. \end{array}\right\}, \quad k \geq 1. \qquad (4.2.21)$$

Обеспечим сначала выполнение соотношений (4.2.21) при $k = 1$. Выберем

$$x_1 \;=\; T_{L_3}(x_0), \quad l_1(x) \;\equiv\; f(x_1),\ x \in E, \quad A_1 \;=\; 1. \qquad (4.2.22)$$

Тогда $\psi_1^* = f(x_1)$ и, следовательно, соотношение $\mathcal{R}_1^1$ выполнено. С другой стороны,

$$\begin{aligned} \psi_1(x) \quad &= \quad f(x_1) + \frac{N}{6}\|x - x_0\|^3 \\ &\overset{(4.2.13)}{\leq} \quad \min_y \left[f(y) + \frac{2L_3}{6}\|y - x_0\|^3\right] + \frac{N}{6}\|x - x_0\|^3, \end{aligned}$$

и это означает, что соотношение $\mathcal{R}_1^2$ тоже выполнено.

Предположим теперь, что соотношения (4.2.21) выполняются при некотором $k \geq 1$. Обозначим

$$v_k \;=\; \arg\min_x \; \psi_k(x).$$

Выберем некоторые числа $a_k > 0$ и $M \geq 2L_3$. Положим

$$\begin{array}{rcl} \alpha_k & = & \frac{a_k}{A_k+a_k}, \\[2ex] y_k & = & (1-\alpha_k)x_k + \alpha_k v_k, \\[2ex] x_{k+1} & = & T_M(y_k), \\[2ex] \psi_{k+1}(x) & = & \psi_k(x) + a_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle]. \end{array} \tag{4.2.23}$$

В силу соотношения $\mathcal{R}_k^2$ для любого $x \in E$ имеем

$$\begin{array}{rcl} \psi_{k+1}(x) & \leq & A_k f(x) + \frac{2L_3+N}{6}\|x-x_0\|^3 + a_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle] \\[2ex] & \leq & (A_k+a_k)f(x) + \frac{2L_3+N}{6}\|x-x_0\|^3, \end{array}$$

и это в точности соотношение $\mathcal{R}_{k+1}^2$. Покажем теперь, что при подходящем выборе a_k, N и M соотношение $\mathcal{R}_{k+1}^1$ также будет выполнено.

Действительно, в силу соотношения $\mathcal{R}_k^1$ и утверждения леммы 1.1.8 для $p = 3$ при любом $x \in E$ имеем:

$$\begin{array}{rcl} \psi_k(x) & \equiv & l_k(x) + \frac{N}{2}d_3(x) \;\geq\; \psi_k^* + \frac{N}{2}\cdot\frac{1}{6}\|x-v_k\|^3 \\[2ex] & \geq & A_k f(x_k) + \frac{N}{2}\cdot\frac{1}{6}\|x-v_k\|^3. \end{array} \tag{4.2.24}$$

Таким образом,

$$\begin{array}{rcl}
\psi_{k+1}^* & = & \min\limits_x \{\psi_k(x) + a_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle]\} \\
& \overset{(4.2.24)}{\geq} & \min\limits_x \left\{A_k f(x_k) + \frac{N}{12}\|x - v_k\|^3 + a_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle]\right\} \\
& \geq & \min\limits_x \{(A_k + a_k) f(x_{k+1}) + A_k\langle \nabla f(x_{k+1}), x_k - x_{k+1}\rangle \\
& & +a_k\langle \nabla f(x_{k+1}), x - x_{k+1}\rangle + \frac{N}{12}\|x - v_k\|^3]\} \\
& \overset{(4.2.23)}{=} & \min\limits_x \{A_{k+1} f(x_{k+1}) + \langle \nabla f(x_{k+1}), A_{k+1}y_k - a_k v_k - A_k x_{k+1}\rangle \\
& & +a_k\langle \nabla f(x_{k+1}), x - x_{k+1}\rangle + \frac{N}{12}\|x - v_k\|^3]\} \\
& = & \min\limits_x \{A_{k+1} f(x_{k+1}) + A_{k+1}\langle \nabla f(x_{k+1}), y_k - x_{k+1}\rangle \\
& & +a_k\langle \nabla f(x_{k+1}), x - v_k\rangle + \frac{N}{12}\|x - v_k\|^3]\}.
\end{array}$$

Далее, если мы выберем $M \geq 2L_3$, то в силу неравенства (4.2.14) получим

$$\langle \nabla f(x_{k+1}), y_k - x_{k+1}\rangle \ \geq \ \sqrt{\frac{2}{L_3+M}} \cdot \|\nabla f(x_{k+1})\|_*^{3/2}.$$

Поэтому наши параметры должны удовлетворять следующему неравенству:

$$A_{k+1}\sqrt{\frac{2}{L_3+M}} \cdot \|\nabla f(x_{k+1})\|_*^{3/2} + a_k\langle \nabla f(x_{k+1}), x - v_k\rangle + \frac{N}{12}\|x - v_k\|^3 \geq 0$$

при всех $x \in E$. Используя неравенство (1.1.6) с $p = 3$, $s = a_k \nabla f(x_{k+1})$ и $\sigma = \frac{1}{4}N$, мы приходим к следующему условию:

$$A_{k+1}\sqrt{\frac{2}{L_3+M}} \ \geq \ \frac{2}{3}\sqrt{\frac{4}{N}}\, a_k^{3/2}. \tag{4.2.25}$$

Для $k \geq 1$ выберем

$$\begin{array}{rcl}
A_k & = & \frac{k(k+1)(k+2)}{6}, \\
a_k & = & A_{k+1} - A_k \ = \ \frac{(k+1)(k+2)(k+3)}{6} - \frac{k(k+1)(k+2)}{6} \\
& = & \frac{(k+1)(k+2)}{2}.
\end{array} \tag{4.2.26}$$

Поскольку

$$a_k^{-3/2} A_{k+1} \ = \ \frac{2^{3/2}(k+1)(k+2)(k+3)}{6[(k+1)(k+2)]^{3/2}} \ = \ \frac{2^{1/2}(k+3)}{3[(k+1)(k+2)]^{1/2}} \ \geq \ \frac{1}{3}\sqrt{2},$$

неравенство (4.2.25) дает следующее условие для выбора параметров:

$$\frac{1}{\sqrt{L_3+M}} \;\ge\; \frac{2}{\sqrt{N}}.$$

Поэтому можно взять

$$M \;=\; 2L_3, \quad N \;=\; 4(L_3+M) \;=\; 12L_3. \tag{4.2.27}$$

В этом случае $2L_3 + N = 14L_3$.

Теперь мы можем собрать все наши наблюдения вместе.

Ускоренная версия кубической регляризации

Инициализация Выбираем $x_0 \in E$. Полагаем $M = 2L_3$ и $N = 12L_3$.

Вычисляем $x_1 = T_{L_3}(x_0)$ и полагаем $\psi_1(x) = f(x_1) + \frac{N}{6}\|x - x_0\|^3$.

Итерация k **(**$k \ge 1$**)**

1. Вычисляем $v_k = \arg\min\limits_{x\in E} \psi_k(x)$ и полагаем $y_k = \frac{k}{k+3}x_k + \frac{3}{k+3}v_k$.

2. Вычисляем $x_{k+1} = T_M(y_k)$ и пересчитываем

$$\psi_{k+1}(x) = \psi_k(x) + \tfrac{(k+1)(k+2)}{2} \cdot [f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle].$$

(4.2.28)

Предыдущие рассуждения доказывают следующую теорему.

Теорема 4.2.2 *Если последовательность точек* $\{x_k\}_{k=1}^{\infty}$ *сформирована методом (4.2.28) применительно к задаче (4.2.6), то для любого* $k \ge 1$ *выполняется неравенство*

$$f(x_k) - f(x^*) \;\le\; \frac{14\, L_3\, \|x_0 - x^*\|^3}{k(k+1)(k+2)}, \tag{4.2.29}$$

где x^* *- оптимальное решение задачи.*

Доказательство.

Действительно, мы показали что

$$A_k f(x_k) \;\overset{\mathcal{R}_k^1}{\le}\; \psi_k^* \;\overset{\mathcal{R}_k^2}{\le}\; A_k f(x^*) + \tfrac{2L_3+N}{6}\|x_0 - x^*\|^3.$$

Таким образом, неравенство (4.2.29) следует из соотношений (4.2.26) и (4.2.27). □

Заметим, что точка v_k может быть найдена в методе (4.2.28) в явном виде. Рассмотрим

$$s_k = \nabla l_k(x).$$

Этот вектор не зависит от x так как функция $l_k(x)$ линейна. Поэтому

$$v_k \; = \; x_0 - \sqrt{\tfrac{2}{N}} \cdot \tfrac{B^{-1}s_k}{\|s_k\|_*^{1/2}}.$$

2. Квадратичные оценочные функции. С помощью аналогичных рассуждений можно обосновать вариант метода (4.2.28), использующий квадратичные оценочные функции. Действительно, рассмотрим функцию

$$\psi_k(x) \; = \; q_k(x) + \tfrac{N}{6}\|x - x_0\|^3, \quad k = 1, 2, \ldots,$$

где $q_k(x)$ – некоторая *квадратичная* функция. Если выбрать

$$x_1 \; = \; T_{L_3}(x_0), \quad q_1(x) \; = \; f_2(x_0; x), \quad A_1 \; = \; 1, \tag{4.2.30}$$

то для любого $N \geq L_3$ можно гарантировать, что условия

$$A_k f(x_k) \overset{(4.2.13)}{\leq} \psi_k^*,$$

$$\psi_k(x) \overset{(4.2.4)}{\leq} f(x) + \tfrac{L_3+N}{6}\|x - x_0\|^3, \quad \forall x \in E,$$

будут выполнены при $k = 1$. Тогда, рассуждая так же, как и выше, можно показать, что в случае $N \geq 4(L_3 + M)$ правила (4.2.23) обеспечивают выполнение вышеприведенных условий уже при всех $k \geq 1$. Таким образом, новая версия (4.2.28) имеет немного лучшую константу в оценке скорости сходимости:

$$f(x_k) - f(x^*) \; \leq \; \frac{13\; L_3\; \|x_0 - x^*\|^3}{k(k+1)(k+2)}.$$

Однако правила пересчета точек v_k при этом усложняются.

4.2.4 Глобальная невырожденность для методов второго порядка

В численном анализе термин *невырожденный* обычно используется для обозначения эффективно решаемых классов задач. Для задач безусловной минимизации невырожденность целевой функции обычно характеризуется с помощью нижней границы на угол между градиентом в точке x и направлением на точку минимума:

$$\alpha(x) \overset{\text{def}}{=} \tfrac{\langle \nabla f(x), x - x^* \rangle}{\|\nabla f(x)\|_* \cdot \|x - x^*\|} \; \geq \; \mu(f) \; > \; 0, \quad x \in E. \tag{4.2.31}$$

Это условие имеет наглядную геометрическую интерпретацию. При этом существует широкий класс задач для которых оно работает. Это сильно выпуклые функции с липшицевым градиентом.

Лемма 4.2.2 *Выполнено неравенство* $\mu(f) \geq \frac{2\sqrt{\gamma_2(f)}}{1+\gamma_2(f)} > \sqrt{\gamma_2(f)}$.

Доказательство.

Действительно, в силу неравенства (2.1.24) в монографии [83] имеем

$$\begin{aligned} \langle \nabla f(x), x - x^* \rangle &\geq \tfrac{1}{\sigma_2+L_2}\|\nabla f(x)\|_*^2 + \tfrac{\sigma_2 L_2}{\sigma_2+L_2}\|x-x^*\|^2 \\ &\geq \tfrac{2\sqrt{\sigma_2 L_2}}{\sigma_2+L_2} \cdot \|\nabla f(x)\|_* \cdot \|x-x^*\|, \end{aligned}$$

что и доказывает требуемое неравенство. □

Заметим, что оценки эффективности методов первого порядка на классе гладких сильно выпуклых функций полностью описываются в терминах числа обусловленности γ_2. Действительно, с одной стороны доказано что нижняя оценка для нахождения ϵ-решения любой задачи из этого класса имеет порядок

$$O\left(\tfrac{1}{\sqrt{\gamma_2}} \ln \tfrac{L_2 D^2}{\epsilon}\right) \tag{4.2.32}$$

обращений к оракулу, где параметр D ограничивает расстояние между начальной и оптимальной точками. С другой стороны, простой метод (1.2.29) обладает нужной для этого скоростью сходимости.

Что можно сказать о сложности этого класса задач для методов второго порядка? Интересно, что для этих методов условие (4.2.31) практически бесполезно. Мы обсудим сложность этого класса в п. 4.2.5. А сейчас приведем условие невырожденности, которое заменяет условие (4.2.31) для методов второго порядка.

Предположим, что $\gamma_3(f) = \frac{\sigma_3(f)}{L_3(f)} > 0$. В этом случае

$$f(x) - f(x^*) \overset{(1.1.23)}{\leq} \tfrac{2}{3\sqrt{\sigma_3}} \cdot \|\nabla f(x)\|_*^{3/2}. \tag{4.2.33}$$

Таким образом, для метода (4.2.16) имеем

$$f(x_k) - f(x_{k+1}) \overset{(4.2.12)}{\geq} \tfrac{1}{3\sqrt{L_3}}\|\nabla f(x_{k+1})\|_*^{3/2} \overset{(4.2.33)}{\geq} \tfrac{1}{2}\sqrt{\gamma_3(f)} \cdot (f(x_{k+1}) - f(x^*)). \tag{4.2.34}$$

Поэтому для любого $k \geq 1$ получаем

$$f(x_k) - f(x^*) \overset{(4.2.34)}{\leq} \frac{f(x_1) - f^*}{\left(1+\frac{1}{2}\sqrt{\gamma_3(f)}\right)^{k-1}} \overset{(4.2.13)}{\leq} e^{-\frac{\sqrt{\gamma_3(f)}\cdot(k-1)}{2+\sqrt{\gamma_3(f)}}} \cdot \tfrac{L_3}{3}\|x_0 - x^*\|^3 \tag{4.2.35}$$

Таким образом, сложность минимизации функций с *положительным* числом обусловленности третьего порядка $\gamma_3(f)$ с помощью метода (4.2.16) имеет порядок

$$O\left(\tfrac{1}{\sqrt{\gamma_3(f)}} \ln \tfrac{L_3 D^3}{\epsilon}\right) \tag{4.2.36}$$

вызовов оракула. Структура этой оценки аналогична (4.2.32). Таким образом, мы можем говорить, что такие функции обладают *глобальной невырожденностью второго порядка*.

Покажем, что ускоренный вариант метода Ньютона (4.2.28) имеет лучшую оценку, чем оценка (4.2.36). Обозначим через $\mathcal{A}_k(x_0)$, $k \geq 1$, точку x_k, сформированную методом (4.2.28) начиная со стартовой точки x_0. Рассмотрим следующий процесс.

$$\begin{array}{l} \textbf{1.}\ \ \text{Положим } m = 5\left\lfloor \frac{1}{\gamma_3(f)} \right\rfloor^{1/3}, \text{ и выберем } y_0 = x_0. \\ \\ \textbf{2.}\ \ \text{Для } k \geq 0 \text{ повторяем } y_{k+1} = \mathcal{A}_m(y_k). \end{array} \tag{4.2.37}$$

Оценки эффективности этой схемы выводятся из следующей леммы.

Лемма 4.2.3 *При любом* $k \geq 0$ *выполнено неравенство* $\|y_{k+1} - x^*\|^3 \ \leq \ \frac{1}{e}\|y_k - x^*\|^3$.

Доказательство.
Действительно, поскольку $m \geq \left(\frac{42e}{\gamma_3(f)}\right)^{1/3}$, то

$$\tfrac{1}{3}\sigma_3\|y_{k+1} - x^*\|^3 \ \overset{(1.1.20)}{\leq} \ f(y_{k+1}) - f(x^*) \ \overset{(4.2.29)}{\leq} \ \tfrac{14L_3\|y_k - x^*\|^3}{m(m+1)(m+2)} \ \leq \ \tfrac{1}{3e}\sigma_3\|y_k - x^*\|^3. \quad \square$$

Таким образом,

$$f(T_{L_3}(y_k)) - f(x^*) \ \overset{(4.2.13)}{\leq} \ \tfrac{L_3}{3}\|y_k - x^*\|^3 \ \overset{(4.2.13)}{\leq} \ \tfrac{L_3}{3}\|y_0 - x^*\|^3 \cdot e^{-k}$$

и мы заключаем, что ϵ-решение нашей задачи может быть найдено методом (4.2.37) за

$$O\left(\tfrac{1}{[\gamma_3(f)]^{1/3}} \ln\left[\tfrac{L_3}{\epsilon}\|x_0 - x^*\|^3\right]\right) \tag{4.2.38}$$

итераций. Заметим, что теория сложности для данного класса задач еще не разработана.

4.2.5 Минимизация сильно выпуклых функций

Оценим сложность решения задачи (4.2.6) с параметрами

$$\sigma_2(f) \ > \ 0, \quad L_3(f) \ < \ \infty. \tag{4.2.39}$$

Основным достоинством этой постановки является возможность квадратичной сходимости метода Ньютона (4.2.16) в некоторой окрестности решения. Действительно, для $T = T_{L_3}(x)$ имеем

$$\begin{array}{rcl} f(x) - f(T) & \overset{(4.2.11)}{\geq} & \frac{1}{2}\langle \nabla^2 f(T)(T - x), T - x\rangle \ \overset{(1.1.34)}{\geq} \ \frac{\sigma_2}{2} \cdot r^2_{L_3}(x) \\ & \overset{(4.2.9)}{\geq} & \frac{\sigma_2}{2L_3} \cdot \|\nabla f(T)\|_* \ \overset{(1.1.23)}{\geq} \ \frac{\sigma_2}{2L_3} \cdot \left[2\sigma_2(f(T) - f(x^*))\right]^{1/2}. \end{array} \tag{4.2.40}$$

Поэтому

$$f(T) - f(x^*) \overset{(4.2.40)}{\leq} \frac{2L_3^2}{\sigma_2^3}(f(x) - f(T))^2 \leq \frac{2L_3^2}{\sigma_2^3}(f(x) - f(x^*))^2 \tag{4.2.41}$$

и, следовательно, область квадратичной сходимости метода (4.2.16) описывается следующим образом:

$$\mathcal{Q}_f = \left\{ x \in E : \ f(x) - f(x^*) \leq \frac{\sigma_2^3}{2L_3^2} \right\}. \tag{4.2.42}$$

Область квадратичной сходимости можно описать и с помощью нормы градиента. Действительно,

$$\frac{\sigma_2}{2} \cdot r_{L_3}^2(x) \overset{(1.1.34)}{\leq} \frac{1}{2}\langle \nabla^2 f(T)(T - x), T - x \rangle \overset{(4.2.11)}{\leq} f(x) - f(T) \leq \|\nabla f(x)\|_* \cdot r_{L_3}(x).$$

Таким образом,

$$\|\nabla f(x)\|_* \geq \frac{\sigma_2}{2} \cdot r_{L_3}(x) \overset{(4.2.9)}{\geq} \frac{\sigma_2}{2} \left[\frac{1}{L_3} \|\nabla f(T)\|_* \right]^{1/2}.$$

Следовательно,

$$\|\nabla f(T)\|_* \leq \frac{4L_3}{\sigma_2^2} \|\nabla f(x)\|_*^2 \tag{4.2.43}$$

и область квадратичной сходимости имеет вид

$$\mathcal{Q}_g = \left\{ x \in E : \ \|\nabla f(x)\|_* \leq \frac{\sigma_2^2}{4L_3} \right\}. \tag{4.2.44}$$

Таким образом, глобальная сложность задач (4.2.6), (4.2.39) в основном определяется числом итераций, необходимых для попадания из точки x_0 в область $\mathcal{Q}_f$ (или $\mathcal{Q}_g$). Для метода (4.2.16) это число допускает верхнюю оценку порядка

$$O\left(\sqrt{\frac{L_3(f) D}{\sigma_2(f)}} \right), \tag{4.2.45}$$

где величина D определяется в (4.2.17) (см. параграф 4.1). Покажем, что эту оценку можно улучшить с помощью ускоренного метода Ньютона (4.2.28).

Пусть нам известна верхняя оценка для расстояния до решения

$$\|x_0 - x^*\| \leq R \quad (\leq D).$$

Рассмотрим следующий процесс.

$$\begin{array}{l} \textbf{1.}\ \ \text{Полагаем } y_0 = T_{L_3}(x_0) \text{ и } m_0 = 5\left[\frac{L_3(f)R}{\sigma_2(f)}\right]^{1/3}. \\ \\ \textbf{2.}\ \ \text{Пока } \|\nabla f(T_{L_3}(y_k))\|_* \geq \frac{\sigma_2^2}{4L_3} \text{ повторяем } \{y_{k+1} = \mathcal{A}_{m_k}(y_k),\ m_{k+1} = \frac{1}{2^{1/3}} m_k\}. \end{array} \tag{4.2.46}$$

Теорема 4.2.3 *Метод (4.2.46) останавливается не более чем после*

$$\frac{1}{\ln 4} \ln \left[\frac{8}{3} \cdot \left(\frac{L_3(f)R}{\sigma_2(f)} \right)^3 \right] \tag{4.2.47}$$

этапов. Общее число ньютоновских шагов на всех этапах не превосходит $4m_0$.

Доказательство.
Обозначим $R_k = R \cdot \left(\frac{1}{2}\right)^k$. Ясно, что справедлива оценка

$$m_k \;\; \geq \;\; 5\left[\tfrac{L_3(f)R_k}{\sigma_2(f)}\right]^{1/3}, \quad k \geq 0. \tag{4.2.48}$$

Для $k \geq 0$ докажем по индукции, что

$$\| y_k - x^* \| \;\; \leq \;\; R_k. \tag{4.2.49}$$

Пусть это верно при некотором $k \geq 0$ (при $k = 0$ утверждение очевидно). Тогда

$$\begin{aligned} \tfrac{\sigma_2}{2}\| y_{k+1} - x^* \|^2 \;\; & \overset{(1.1.20)}{\leq} \;\; f(y_{k+1}) - f(x^*) \;\; \overset{(4.2.29)}{\leq} \;\; \tfrac{14 L_3 R_k^3}{m_k(m_k+1)(m_k+2)} \\ & \overset{(4.2.48)}{\leq} \;\; \tfrac{14}{125}\sigma_2 R_k^2 \;\leq\; \tfrac{1}{8}\sigma_2 R_k^2 \;=\; \tfrac{1}{2}\sigma_2 R_{k+1}^2. \end{aligned}$$

Таким образом, неравенство (4.2.49) выполняется при всех $k \geq 0$. С другой стороны,

$$\begin{aligned} f(y_{k+1}) - f(x^*) \;\; & \overset{(4.2.29)}{\leq} \;\; \tfrac{14 L_3 \| y_k - x^* \|^3}{m_k(m_k+1)(m_k+2)} \;\; \overset{(4.2.49)}{\leq} \;\; \tfrac{14 L_3 \| y_k - x^* \|^2 R_k}{m_k(m_k+1)(m_k+2)} \\ & \overset{(4.2.48)}{\leq} \;\; \tfrac{1}{8}\sigma_2 \| y_k - x^* \|^2 \;\; \overset{(1.1.20)}{\leq} \;\; \tfrac{1}{4}(f(y_k) - f(x^*)). \end{aligned}$$

Поэтому

$$\begin{aligned} \tfrac{\sigma_2}{2L_3}\| \nabla f(T_{L_3}(y_k)) \|_* \;\; & \overset{(4.2.40)}{\leq} \;\; f(y_k) - f(T_{L_3}(y_k)) \;\leq\; f(y_k) - f(x^*) \\ & \leq \;\; \left(\tfrac{1}{4}\right)^k (f(y_0) - f(x^*)) \;\; \overset{(4.2.13)}{\leq} \;\; \left(\tfrac{1}{4}\right)^k \tfrac{L_3}{3} R^3 \end{aligned}$$

и оценка (4.2.47) следует из определения области квадратичной сходимости (4.2.44). Наконец, общее число шагов метода Ньютона не превосходит

$$\sum_{k=0}^{\infty} m_k \;=\; m_0 \sum_{k=0}^{\infty} \tfrac{1}{2^{k/3}} \;=\; \tfrac{m_0}{2^{1/3}-1} \;<\; 4m_0. \qquad \square$$

4.2.6 Ложное ускорение

Заметим, что свойства класса гладких сильно выпуклых функций (4.2.39) оставляют некоторые возможности для ошибочных утверждений, связанных со скоростью сходимости методов минимизации на первой стадии процесса, которая заканчивается при попадании точек минимизирующей последовательности в область квадратичной сходимости метода Ньютона. Продемонстрируем одну из возможных ошибок.

Рассмотрим модификацию $\mathcal{M}'$ метода (4.2.28). Единственное изменение вносится в шаг 2. Теперь он выглядит так:

2'. Вычисляем $\hat{y}_k = T_M(y_k)$ и пересчитываем

$$\psi_{k+1}(x) = \psi_k(x) + \tfrac{(k+1)(k+2)}{2} \cdot [f(\hat{y}_k) + \langle \nabla f(\hat{y}_k), x - \hat{y}_k \rangle]. \tag{4.2.50}$$

Выбираем $\hat{x}_k :\ f(\hat{x}_k) = \min\{f(x_k), f(\hat{y}_k)\}$. Полагаем $x_{k+1} = T_M(\hat{x}_k)$.

Заметим, что для метода $\mathcal{M}'$ утверждение теоремы 4.2.2 остается верным. Более того, метод теперь стал монотонным. Поэтому, пользуясь тем же рассуждением, что и в цепочке неравенств (4.2.40), и полагая $M = 2L_3$ мы получаем неравенство

$$f(x_k) - f(x_{k+1}) \ \ \geq \ \ f(\hat{x}_k) - f(x_{k+1}) \ \geq \ \tfrac{\sqrt{2}\,\sigma_2^{3/2}}{3L_3} \cdot [f(x_{k+1}) - f(x^*)]^{1/2}. \tag{4.2.51}$$

Далее, зафиксируем число шагов N. Положим $\hat{k} = \frac{2}{3}N$. Тогда в силу (4.2.29) можно гарантировать, что

$$f(x_{\hat{k}}) - f(x^*) \ \ \leq \ \ \tfrac{3^3 \cdot 7 \cdot L_3 R^3}{2^2 N^3}. \tag{4.2.52}$$

С другой стороны,

$$f(x_{\hat{k}}) - f(x^*) \ \ \geq \ \ f(x_{\hat{k}}) - f(x_{N+1}) \ \ \overset{(4.2.51)}{\geq} \ \ \tfrac{1}{3}N \cdot \tfrac{\sqrt{2}\,\sigma_2^{3/2}}{3L_3} \cdot [f(x_{N+1}) - f(x^*)]^{1/2}. \tag{4.2.53}$$

Объединяя неравенства (4.2.52) и (4.2.53), получаем оценку

$$f(x_{N+1}) - f(x^*) \ \ \leq \ \ \tfrac{3^{10} \cdot 7^2 \cdot L_3^4 \cdot \mathbb{R}^6}{2^5 \cdot \sigma_2^3} \cdot N^{-8}. \tag{4.2.54}$$

По сравнению с (4.2.29), предложенная модификация выглядит необычайно эффективной. Однако это лишь иллюзия. На самом деле, в силу определения области (4.2.42), для того, чтобы найти точку в области квадратичной сходимости метода Ньютона, надо сделать правую часть неравенства (4.2.54) меньше, чем $\frac{\sigma_2^3}{2L_3^2}$. Для этого нужно

$$O\left(\left[\tfrac{L_3 R}{\sigma_2} \right]^{3/4} \right) \tag{4.2.55}$$

итераций метода $\mathcal{M}'$. Это намного больше, чем в требуется в оценке (4.2.45) для не самой быстрой основной схемы (4.2.16).

Другим тестом является оценка числа шагов, необходимых методу $\mathcal{M}'$ для уменьшения расстояния до точки минимума вдвое. Из оценки (4.2.54) видно, что для этого потребуется $O\left(\left[\frac{L_3 R}{\sigma_2} \right]^{1/2} \right)$ итераций. Это хуже, чем соответствующая оценка для метода (4.2.28).

4.2.7 Обсуждение

1. С помощью вышеприведенных оценок сложности можно понять какой класс задач является *простым* для методов второго порядка. Это функции со следующими параметрами:

$$\sigma_2(f) \ > \ 0, \quad \sigma_3(f) \ > \ 0, \quad L_3(f) \ < \ \infty. \tag{4.2.56}$$

На них методы второго порядка демонстрируют глобальную линейную скорость сходимости и локальную квадратичную. В соответствии с (4.2.38) и (4.2.42), нам потребуется

$$O\left(\left[\frac{L_3(f)}{\sigma_3(f)}\right]^{1/3} \ln \left[\frac{L_3(f)}{\sigma_2(f)} \| x_0 - x^* \| \right] \right) \tag{4.2.57}$$

итераций метода (4.2.28) для вхождения в область квадратичной сходимости.

Заметим, что класс функций (4.2.56) не является тривиальным. Он содержит, например, все функции вида

$$\xi_{\alpha,\beta}(x) = \alpha d_2(x) + \beta d_3(x), \quad \alpha, \beta > 0,$$

с параметрами

$$\sigma_2(\xi_{\alpha,\beta}) = \alpha, \quad \sigma_3(\xi_{\alpha,\beta}) = \tfrac{1}{2}\beta, \quad L_3(\xi_{\alpha,\beta}) = 2\beta.$$

Более того, любая выпуклая функция с липшицевым гессианом может быть искусственно *регуляризирована* с помощью сложения с функцией $\xi_{\alpha,\beta}$.

2. Для одного важного класса выпуклых задач, имеющих параметры

$$\sigma_2(f) \ > \ 0, \quad L_2(f) \ < \ \infty, \quad L_3(f) \ < \ \infty, \tag{4.2.58}$$

ситуация остается неясной. Стандартная теория оптимальных методов *первого порядка* (см. п. 1.2.3), ограничивает число шагов необходимых для вхождения в область квадратичной сходимости (4.2.42) следующим образом:

$$O\left(\left[\frac{L_2(f)}{\sigma_2(f)}\right]^{1/2} \ln \left[\frac{L_2(f) L_3^2(f)}{\sigma_2^3(f)} \| x_0 - x^* \|^2 \right] \right). \tag{4.2.59}$$

Заметим, что в этой оценке роль методов второго порядка весьма скромная: они используются только для обоснования оценок для завершающей стадии процесса. Конечно же, как было показано в п. 4.2.5, их можно было бы использовать и на первом этапе. Однако в это случае размер оптимального решения x^* вошел бы *полиномиально* в оценку для числа итераций. Таким образом, следующий вопрос до сих пор открыт:

> *Можно ли получить какие-либо преимущества от использования методов второго порядка на первой стадии процесса при минимизации функций из класса (4.2.58)?*

4.3 Модифицированный метод Гаусса-Ньютона

4.3.1 Введение

Мотивировка

Задача решения системы нелинейных уравнений является одной из наиболее важных задач в численном анализе. Стандартный подход к решению этой задачи состоит в замене ее начальной формулировки

$$\text{найти } x \in \mathbb{R}^n : \ f_i(x) = 0, \quad i = 1, \dots, m, \tag{4.3.1}$$

на задачу минимизации

$$\min_{x \in \mathbb{R}^n} \left[f(x) \stackrel{\mathrm{def}}{=} \phi(f_1(x), \dots, f_m(x)) \right], \tag{4.3.2}$$

где функция $\phi(u)$ неотрицательна и обращается в нуль только в начале координат. Стандартная рекомендация по выбору *функции близости* $\phi(u)$ заключается в следующем:

$$\phi(u) = \|u\|_{(2)}^2 \equiv \sum_{i=1}^{m} \left(u^{(i)}\right)^2, \tag{4.3.3}$$

где возведение в квадрат евклидовой нормы позволяет получить в задаче (4.3.2) гладкую целевую функцию. Конечно же задача (4.3.2), (4.3.3) может решаться стандартными методами второго порядка. Однако порядок используемых производных можно уменьшить если воспользоваться методом *Гаусса-Ньютона*, в котором направление поиска определяется с помощью решения следующей вспомогательной задачи:

$$\min_{h} \left\{ \phi\left(f_1(x) + \langle \nabla f_1(x), h \rangle, \dots, f_m(x) + \langle \nabla f_m(x), h \rangle\right) : \ x + h \in D(x) \right\},$$

где через $D(x)$ обозначена определенным образом выбранная окрестность текущей точки x. При некоторых естественных предположениях можно показать что эта схема имеет локальную квадратичную сходимость.

Несмотря на свою простоту и естественность, этот подход имеет несколько недостатков. Действительно, преобразование задачи (4.3.1) в задачу (4.3.2) сделано достаточно прямолинейно. Если, например, производная оператора нашей системы есть симметрическая матрица, то такое преобразование возводит в квадрат ее число обусловленности. Кроме увеличивающейся численной неустойчивости, это приводит к возведению в квадрат числа итераций, нужных для получения ϵ-решения исходной задачи.

В этом параграфе мы предлагаем другой подход. На первый взгляд он очень похож на стандартный: мы заменяем нашу исходную задачу на задачу минимизации (4.3.2), но с *негладкой* функцией близости. Например, можно взять $\phi(u) = \|u\|$ с произвольной нормой. Другое отличие состоит в том, что в качестве новой точки выбирается точка минимума вспомогательной функции, формируемой как сумма “линеаризированной” функции близости и прокс-функции. Можно показать, что при естественных предположениях в задаче (4.3.2) эта стратегия гарантирует монотонное убывание целевой функции. При этом в окрестности точки минимума возникает квадратичная сходимость. Более того, для некоторых естественных классов невыпуклых задач можно получить глобальные оценки эффективности. Заметим, что в большинстве публикаций о методах решения задачи (4.3.1) рассматривается ситуация $m \geq n$ (что соответствует постановкам, решаемых с помощью метода наименьших квадратов). В этом параграфе наиболее интересные результаты (см. п. 4.3.4) получаются для $m \leq n$, что, по-видимому, более естественно для разрешимых систем нелинейных уравнений.

Содержание

В п. 4.3.2 мы определяем один шаг модифицированного метода Гаусса-Ньютона и доказываем его основные свойства. В п. 4.3.3 рассматривается уже весь процесс и доказывается, что любая предельная точка сформированной последовательности удовлетворяет условиям оптимальности первого порядка. Если решение системы (4.3.1) удовлетворяет (прямым) условиям невырожденности, то сходимость процесса будет квадратичной. В п. 4.3.4 мы изучаем класс задач с равномерной *двойственной* невырожденностью (терминология будет объяснена чуть ниже). Задачи из этого класса могут иметь даже нетривиальное непрерывное множество решений (в этом случае соответствующие якобианы будут вырожденными). Тем не менее, для этого класса задач доказываются глобальные оценки эффективности и устанавливается локальная квадратичная сходимость. В последнем п. 4.3.5 обсуждаются полученные результаты. В п. 4.3.5A сравнивается эффективность модифицированного метода Гаусса-Ньютона с кубической регуляризацией метода Ньютона (см. параграф 4.1). В п. 4.3.5B обсуждается сложность вспомогательных задач, возникающих в новом методе. Показывается, что для их решения вполне достаточно средств стандартной Линейной Алгебры.

Обозначения

Для линейного оператора $A: \; E_1 \to E_2$, его *операторная норма* $\|A\|$ определяется как

$$\|A\| = \max_{x \in E_1} \{ \|Ax\| : \; \|x\| \leq 1 \}.$$

Для этого определения смысл пространств E_1 и E_2 всегда ясен из контекста. Для такого оператора мы рассматриваем также минимальное сингулярное число

$$\sigma_{\min}(A) = \min_{x \in E_1} \{ \|Ax\| : \; \|x\| = 1 \}.$$

Если оператор A обратим, то $\sigma_{\min}(A) = 1/\|A^{-1}\|$. Заметим, что для двух линейных операторов A_1 и A_2 выполнено неравенство

$$\sigma_{\min}(A_1 A_2) \geq \sigma_{\min}(A_1) \cdot \sigma_{\min}(A_2).$$

Если $\sigma_{\min}(A) > 0$, то говорится, что оператор A обладает *прямой невырожденностью*. Если $\sigma_{\min}(A^*) > 0$, то мы говорим, что A обладает *двойственной невырожденностью*.

Наконец, для нелинейной функции $F(x) : E_1 \to E_2$, мы обозначаем через $\nabla F(x)$ ее *якобиан*, который является линейным оператором из E_1 в E_2:

$$\nabla F(x)h = \lim_{\alpha \to 0} \tfrac{1}{\alpha}[F(x + \alpha h) - F(x)] \in E_2, \quad h \in E_1.$$

В специальном случае $f(x) : E_1 \to E_2 \equiv R$, обозначение $\nabla f(x)$ продолжает использоваться для *градиента* функции $f(x)$. В этом случае $\nabla f(x)$ рассматривается как элемент из E_1^*. Для недифференцируемой выпуклой функции $f(x)$ через $\partial f(x)$ обозначается ее субдифференциал.

4.3.2 Модифицированный метод Гаусса-Ньютона

Рассмотрим гладкую нелинейную функцию $F(x) : \ E_1 \to E_2$. Мы собираемся построить метод для приближенного решения следующего уравнения:

$$F(x) = 0, \quad x \in E_1. \tag{4.3.4}$$

Для измерения качества такого решения, введем (острую) *функцию близости* $\phi(u)$, $u \in E_2$, которая удовлетворяет следующим условиям.

- Она выпукла, неотрицательна и обращается в нуль только в начале координат. (Поэтому ее линии уровня ограничены.)
- Она удовлетворяет условию Липшица с единичной константой:

  $$|\phi(u) - \phi(v)| \leq \|u - v\|, \quad \forall u, v \in E_2.$$

- У нее острый минимум в нуле:

  $$\phi(u) \geq \gamma_\phi \|u\|, \quad \forall u \in E_2, \tag{4.3.5}$$

  при некотором $\gamma_\phi \in (0, 1]$.

Например, можно взять $\phi(u) = \|u\|$. Тогда $\gamma_\phi = 1$.

Мы воспользуемся функцией близости для преобразования задачи (4.3.4) в следующую задачу *безусловной минимизации*:

$$\min_{x \in E_1} \{ \ f(x) \equiv \phi(F(x)) \ \} \stackrel{\mathrm{def}}{=} f^*. \tag{4.3.6}$$

Ясно, что решение x^* уравнения (4.3.4) существует в том и только том случае, когда оптимальное значение f^* задачи (4.3.6) равно нулю. Итеративная процедура, предлагаемая ниже, может рассматриваться как специальный метод минимизации для задачи (4.3.6), существенно использующий структуру целевой функции. Целевая функция $f(x)$ может быть даже негладкой. Однако мы увидим, что ее значение можно уменьшить относительно любой точки x в E_1 за исключением множества стационарных точек задачи (4.3.6).

Зафиксируем некоторую точку $x \in E_1$. Рассмотрим следующую *локальную модель* нашей целевой функции:

$$\psi(x;y) = \phi\left(F(x) + \nabla F(x)(y-x)\right), \quad y \in E_1.$$

Заметим, что функция $\psi(x;y)$ выпукла по y. Таким образом, кажется естественным выбрать следующее приближенное решение задачи (4.3.6) из множества

$$\operatorname{Arg}\min_{y \in E_1} \psi(x;y).$$

Такие схемы очень хорошо изучены (см. работы [45], [46], [104], [101]). Например, если взять

$$\phi(u) = \left[\sum_{i=1}^{m} \left(u^{(i)}\right)^2\right]^{1/2}, \quad u \in \mathbb{R}^m,$$

то получится классический метод *Гаусса-Ньютона*. Однако ниже будет показано что имеет смысл каким-то образом регуляризовать этот выбор. В этом случае мы сможем даже говорить о глобальных оценках эффективности такого процесса.

Введем следующее условие гладкости. Пусть выпуклое замкнутое множество $\mathcal{F}$ в E_1 имеет непустую внутренность.

Предположение 4.3.1 Функция $F(x)$ является дифференцируемой на $\mathcal{F}$ и ее производная удовлетворяет условию Липшица:

$$\|\nabla F(x) - \nabla F(y)\| \le L\|x-y\| \quad \forall x,y \in \mathcal{F}, \tag{4.3.7}$$

с некоторой константой $L > 0$.

Немедленным следствием этого предположения является неравенство

$$\|F(y) - F(x) - F'(x)(y-x)\| \;\; \le \;\; \tfrac{1}{2}L\|y-x\|^2, \quad x,y \in \mathcal{F}. \tag{4.3.8}$$

В этом параграфе мы предполагаем что предположение 4.3.1 всегда выполнено.

Лемма 4.3.1 *При любых $x, y \in \mathcal{F}$ выполнено неравенство*

$$|f(y) - \psi(x;y)| \le \tfrac{1}{2}L\|y-x\|^2. \tag{4.3.9}$$

Доказательство.

Обозначим $\xi(x,y) = F(y) - F(x) - \nabla F(x)(y-x) \in E_2$. В силу неравенства (4.3.8)

$$\|\xi(x,y)\| \le \tfrac{1}{2}L\|x-y\|^2.$$

Поэтому, поскольку обе точки x и y лежат в области $\mathcal{F}$, получаем

$$\begin{aligned} |f(y) - \psi(x;y)| &= |\phi(F(y)) - \phi(F(x) + \nabla F(x)(y-x))| \\ &\leq \|\xi(x,y)\| \leq \tfrac{1}{2}L\|y-x\|^2. \end{aligned}$$

□

Заметим, что неравенство (4.3.9) дает нам верхнюю аппроксимацию для функции $f(x)$:

$$f(y) \leq \psi(x;y) + \tfrac{1}{2}L\|y-x\|^2, \quad \forall x, y \in \mathcal{F}.$$

Воспользуемся ею для построения метода минимизации. Пусть M – положительный параметр. Для задачи (4.3.6), определим один шаг *модифицированного метода Гаусса-Ньютона* из точки $x \in \mathcal{F}$ следующим образом:

$$\boxed{V_M(x) \;\in\; \operatorname{Arg}\min_{y \in E_1} \left[\psi(x;y) + \tfrac{1}{2}M\|y-x\|^2\right],} \tag{4.3.10}$$

где "Arg" означает что точка $V_M(x)$ выбрана из множества глобальных минимумов соответствующей задачи минимизации[1]. Заметим, что вспомогательная задача в (4.3.10) *выпукла* по y. Мы отложим вопросы, связанные со сложностью нахождения точки $V_M(x)$, до п. 4.3.5.

Докажем несколько вспомогательных результатов. Обозначим

$$\begin{aligned} r_M(x) &= \|V_M(x) - x\|, \\ f_M(x) &= \psi(x; V_M(x)) + \tfrac{1}{2}Mr_M^2(x), \\ \delta_M(x) &= f(x) - f_M(x). \end{aligned}$$

Заметим, что при фиксированном x функция $f_M(x)$ *вогнута* по M поскольку она определена как минимум линейных по M функций:

$$f_M(x) = \min_{y \in E_1} \left[\psi(x;y) + \tfrac{1}{2}M\|y-x\|^2\right].$$

Следовательно, значение $\frac{1}{2}r_M^2(x)$, которое является производной функции $f_M(x)$ по M, оказывается *убывающей* функцией от M.

Лемма 4.3.2 *При любом $x \in E_1$ выполнено неравенство*

$$\delta_M(x) \geq \tfrac{1}{2}Mr_M^2(x). \tag{4.3.11}$$

---

[1] Поскольку мы не предполагаем, что норма $\|x\|$, $x \in E_1$, сильно выпукла, эта задача может иметь нетривиальное множество решений.

Доказательство.

Зафиксируем произвольную точку $x \in E_1$. Рассмотрим функцию

$$\xi(t) \quad = \quad \min_{y \in E_1} \left[\phi(F(x) + \nabla F(x)(y-x)) + \tfrac{1}{2t}\|y-x\|^2 \right]. \tag{4.3.12}$$

Заметим, что множество $\{(y,t,\alpha) \in E_1 \times \mathbb{R}^2_+ : \ \|y-x\| \leq (\alpha t)^{1/2}\}$ выпукло. Следовательно, целевая функция задачи оптимизации в (4.3.12) выпукла по совокупности переменных (y,t). Итак, функция $\xi(t)$ является выпуклой по t и

$$g(t) \equiv -\tfrac{1}{2t^2} r^2_{1/t}(x) \in \partial \xi(t).$$

Таким образом,

$$f(x) = \xi(0) \geq \xi(t) + g(t) \cdot (-t) = \xi(t) + \tfrac{1}{2t} r^2_{1/t}(x).$$

Поскольку $\xi(\frac{1}{M}) = f_M(x)$, получаем неравенство (4.3.11). □

Сравним величину $\delta_M(x)$ с другой естественной мерой локального убывания модели $\psi(x;\cdot)$. Для $r > 0$ обозначим

$$\Delta_r(x) = f(x) - \min_{y \in E_1} \{\psi(x;y) : \ \|y-x\| \leq r\}.$$

Лемма 4.3.3 *Для любого* $x \in E_1$ *и* $r > 0$ *выполнено неравенство*

$$\delta_M(x) \geq Mr^2 \cdot \kappa\left(\tfrac{1}{Mr^2}\Delta_r(x)\right), \tag{4.3.13}$$

где

$$\kappa(t) = \begin{cases} t - \frac{1}{2}, & \textit{если } t \geq 1, \\ \\ \frac{1}{2}t^2, & \textit{если } t \in [0,1]. \end{cases}$$

Правая часть оценки (4.3.13) является убывающей функцией от M.

Доказательство.

Выберем $h_r \in \operatorname{Arg} \min\limits_{h \in E_1} \{\psi(x; x+h) : \ \|h\| \leq r\}$. Тогда

$$\begin{aligned} f_M(x) &\leq \min_\tau \{\phi(F(x) + \tau \nabla F(x) h_r) + \tfrac{1}{2} M \tau^2 r^2 : \ \tau \in [0,1]\} \\ &= \min_\tau \{\phi((1-\tau)F(x) + \tau(F(x) + \nabla F(x) h_r)) + \tfrac{1}{2} M \tau^2 r^2 : \ \tau \in [0,1]\} \\ &\leq \min_\tau \{(1-\tau)\phi(F(x)) + \tau \phi(F(x) + \nabla F(x) h_r)) + \tfrac{1}{2} M \tau^2 r^2 : \ \tau \in [0,1]\} \\ &= \min_\tau \{f(x) - \tau \Delta_r(x) + \tfrac{1}{2} M \tau^2 r^2 : \ \tau \in [0,1]\}. \end{aligned}$$

Таким образом,

$$\delta_M(x) \geq \max_{\tau\in[0,1]}\{\tau\Delta_r(x) - \tfrac{1}{2}M\tau^2r^2\} = Mr^2 \cdot \kappa\left(\tfrac{1}{Mr^2}\Delta_r(x)\right).$$

Осталось заметить, что правая часть этого неравенства убывает по M. □

Обозначим

$$\mathcal{L}(\tau) = \{y \in E_1 : \ f(y) \leq \tau\}.$$

Лемма 4.3.4 *Пусть* $\mathcal{L}(f(x)) \subseteq \operatorname{int}\mathcal{F}$ *и* $M \geq L$. *Тогда* $V_M(x) \in \mathcal{L}(f(x))$.

Доказательство.
Предположим, что $V_M(x) \notin \mathcal{L}(f(x))$. Рассмотрим точки

$$y(\alpha) = x + \alpha\cdot(V_M(x) - x), \quad \alpha \in [0,1].$$

Поскольку $y(0) = x \in \operatorname{int}\mathcal{F}$, можно определить такое значение $\bar{\alpha} \in (0,1)$, что точка $y(\bar{\alpha})$ лежит на границе множества $\mathcal{F}$. Заметим, что

$$f(y(\bar{\alpha})) \geq f(x) \geq f_M(x)$$

и $r_M(x) > 0$. Ввиду нашего предположения, имеем $\bar{\alpha} \in (0,1)$. Обозначим

$$d = F(y(\bar{\alpha})) - F(x) - \bar{\alpha}\nabla F(x)(V_M(x) - x) \in E_2.$$

В силу неравенства (4.3.8), $\|d\| \leq \frac{L}{2}\,\bar{\alpha}^2 r_M^2(x)$. Таким образом,

$$\begin{aligned} f(x) \leq f(y(\bar{\alpha})) &= \phi(F(x) + \bar{\alpha}\nabla F(x)(y(1) - x) + d) \\ &\leq \phi((F(x) + \bar{\alpha}\nabla F(x)(V_M(x) - x)) + \|d\| \\ &\leq (1-\bar{\alpha})f(x) + \bar{\alpha}\phi((F(x) + \nabla F(x)(V_M(x) - x)) + \tfrac{1}{2}M\bar{\alpha}^2 r_M^2(x) \\ &\leq (1-\bar{\alpha})f(x) + \bar{\alpha}f_M(x) - \tfrac{1}{2}M\bar{\alpha}(1-\bar{\alpha})r_M^2(x). \end{aligned}$$

Следовательно, $f(x) \leq f_M(x) - \frac{1}{2}M(1-\bar{\alpha})r_M^2(x)$, что противоречит неравенству (4.3.11). □

Лемма 4.3.5 *Пусть обе точки* x *и* $V_M(x)$ *принадлежат множеству* $\mathcal{F}$. *Тогда*

$$f_M(x) \leq \min_{y\in\mathcal{F}}\left[f(y) + \tfrac{1}{2}(L+M)\|y-x\|^2\right]. \tag{4.3.14}$$

Доказательство.

Для $y \in \mathcal{F}$ обозначим $d(x, y) = F(y) - F(x) - \nabla F(x)(y - x) \in E_2$. В силу утверждения 3.2.12 в монографии [104],

$$\|d(x, y)\| \le \tfrac{1}{2} L \|x - y\|^2.$$

Поэтому, поскольку обе точки x и $V_M(x)$ лежат в $\mathcal{F}$, имеем

$$\begin{aligned} f_M(x) &= \min_{y \in \mathcal{F}} \left[\phi(F(x) + \nabla F(x)(y - x)) + \tfrac{1}{2} M \|y - x\|^2 \right] \\ &= \min_{y \in \mathcal{F}} \left[\phi(F(y) - d(x, y)) + \tfrac{1}{2} M \|y - x\|^2 \right] \\ &\le \min_{y \in \mathcal{F}} \left[f(y) + \tfrac{1}{2}(L + M) \|y - x\|^2 \right]. \qquad \Box \end{aligned}$$

Следствие 4.3.1 *Пусть точка* x^* *является решением задачи (4.3.6) и* $\mathcal{L}(f(x)) \subseteq \mathcal{F}$. *Тогда выполнено неравенство*

$$f_M(x) \le f^* + \tfrac{1}{2}(L + M) \|x - x^*\|^2. \tag{4.3.15}$$

Доказательство.

Достаточно подставить точку $y = x^*$ в правую часть неравенства (4.3.14). □

4.3.3 Процесс минимизации

Теперь мы сможем проанализировать сходимость следующего процесса. Зафиксируем некоторое $L_0 \in (0, L]$.

Модифицированный метод Гаусса-Ньютона

Инициализация Выбираем $x_0 \in \mathbb{R}^n$.

Итерация k $(k \ge 0)$ (4.3.16)

1. Находим такое $M_k \in [L_0, 2L]$, что

$$f(V_{M_k}(x_k)) \le f_{M_k}(x_k).$$

2. Полагаем $x_{k+1} = V_{M_k}(x_k)$.

Поскольку $f_M(x) \le f(x)$, этот процесс монотонен:

$$f(x_{k+1}) \le f(x_k). \tag{4.3.17}$$

Если константа L известна, то в п. 1 этой схемы можно взять $M_k \equiv L$. В противном случае, можно применить простую процедуру настройки (см., например, п. 4.1.5). Приведем теперь результаты о сходимости процесса.

Пусть для вышеприведенного метода точка $x_0 \in \operatorname{int} \mathcal{F}$ является стартовой. Нам потребуется следующее предположение.

Предположение 4.3.2 Множество $\mathcal{F}$ является достаточно большим: $\mathcal{L}(f(x_0)) \subseteq \mathcal{F}$.

В этом пункте мы всегда считаем, что предположение 4.3.2 выполнено. В силу свойства (4.3.17), из него следует, что $\mathcal{L}(f(x_k)) \subseteq \mathcal{F}$ для любого $k \ge 0$.

Теорема 4.3.1 *Для любого $k \ge 0$ и $r > 0$ выполнены неравенства*

$$\begin{array}{rcl} f(x_k) - f^* & \ge & \frac{1}{2} L_0 \sum\limits_{i=k}^{\infty} r^2_{M_i}(x_i) \\ & \ge & \frac{1}{2} L_0 \sum\limits_{i=k}^{\infty} r^2_{2L}(x_i), \\ f(x_k) - f^* & \ge & r^2 \sum\limits_{i=k}^{\infty} M_i \kappa \left(\frac{1}{M_i r^2} \Delta_r(x) \right) \\ & \ge & 2Lr^2 \sum\limits_{i=k}^{\infty} \kappa \left(\frac{1}{2Lr^2} \Delta_r(x) \right). \end{array} \tag{4.3.18}$$

Доказательство.
Действительно, в силу правил шага 1 в методе (4.3.16),

$$f_{M_i}(x_i) \ge f(x_{i+1}), \quad M_i \ge L_0, \quad r_{M_i}(x_i) \ge r_{2L}(x_i).$$

Таким образом, неравенство (4.3.11) обеспечивает выполнение первого из неравенств (4.3.18). Для того, чтобы доказать второе из этих неравенств, воспользуемся неравенством (4.3.13) и границей $M_i \le 2L$, включенной в условие одномерного поиска метода (4.3.16). □

Следствие 4.3.2 *Пусть последовательность $\{x_k\}_{k=0}^{\infty}$ сформирована методом (4.3.16). Тогда выполнены соотношения*

$$\lim_{k\to\infty} \| x_k - x_{k+1} \| = 0, \quad \lim_{k\to\infty} \Delta_r(x_k) = 0.$$

Следовательно, множество предельных точек X^* связано. Для любого $\bar{x}$ из X^* имеем $\Delta_r(\bar{x}) = 0$.

Обоснуем теперь локальную скорость сходимости метода (4.3.16).

Теорема 4.3.2 *Пусть точка* $x^* \in \mathcal{L}(f(x_0))$, $F(x^*) = 0$, *является невырожденным решением задачи (4.3.4):*

$$\sigma \equiv \sigma_{\min}(\nabla F(x^*)) > 0.$$

И пусть величина γ_ϕ *определяется из неравенства (4.3.5). Если* $x_k \in \mathcal{L}(f(x_0))$ *и*

$$\|x_k - x^*\| \;\; \leq \;\; \tfrac{2}{L} \cdot \tfrac{\sigma\gamma_\phi}{3+5\gamma_\phi},$$

то $x_{k+1} \in \mathcal{L}(f(x_0))$ *и*

$$\|x_{k+1} - x^*\| \;\; \leq \;\; \tfrac{3(1+\gamma_\phi)L\,\|x_k-x^*\|^2}{2\gamma_\phi(\sigma-L\|x_k-x^*\|)} \;\; \leq \;\; \|x_k - x^*\|. \qquad (4.3.19)$$

Доказательство.
Поскольку $f(x^*) = 0$, в силу неравенства (4.3.15) и предложения 3.2.12 [104] имеем

$$\begin{aligned}
\tfrac{3L}{2}\|x_k - x^*\|^2 \geq \;\; & f_{M_k}(x_k) \geq \psi(x_k; x_{k+1}) \\
\geq \;\; & \gamma_\phi \|F(x_k) + \nabla F(x_k)(x_{k+1} - x_k)\| \\
= \;\; & \gamma_\phi \|\nabla F(x^*)(x_{k+1} - x^*) + (F(x_k) - F(x^*) - \nabla F(x^*)(x_k - x^*)) \\
& + (\nabla F(x_k) - \nabla F(x^*))(x_{k+1} - x_k)\| \\
\geq \;\; & \gamma_\phi \left[\|\nabla F(x^*)(x_{k+1} - x^*)\| - \tfrac{L}{2}\|x_k - x^*\|^2 - L\|x_k - x^*\| \cdot \|x_{k+1} - x_k\| \right] \\
\geq \;\; & \gamma_\phi \left[(\sigma - L\|x_k - x^*\|) \cdot \|x_{k+1} - x^*\| - \tfrac{3L}{2}\|x_k - x^*\|^2 \right]. \qquad \square
\end{aligned}$$

4.3.4 Глобальная скорость сходимости

Для получения оценок глобальной эффективности метода (4.3.16) нам потребуется дополнительное предположение.

Предположение 4.3.3 Оператор $\nabla F(x) : E_1 \to E_2$ обладает свойством равномерной *двойственной* невырожденности:

$$\sigma_{\min}(\nabla F(x)^*) \geq \sigma > 0 \quad \forall x \in \mathcal{L}(f(x_0)).$$

Заметим, что из этого предположения следует $\dim E_2 \leq \dim E_1$. Важность предположения 4.3.3 понятна из следующего стандартного факта.

Лемма 4.3.6 *Пусть оператор* $A : E_1 \to E_2$ *обладает двойственной невырожденностью:* $\sigma_{\min}(A^*) > 0$. *Тогда для любого* $b \in E_2$*существует такая точка* $x(b) \in E_1$, *что*

$$Ax(b) = b, \quad \|x(b)\| \leq \tfrac{\|b\|}{\sigma_{\min}(A^*)}.$$

(Этот факт тривиально выводится из сингулярного разложения матрицы A.)

Докажем важное следствие леммы 4.3.6.

Лемма 4.3.7 *Пусть оператор $\nabla F(x)$ является двойственно невырожденным:*

$$\sigma_{\min}(\nabla F(x)^*) > 0.$$

Тогда для любого $M > 0$ выполнено неравенство

$$r_M(x) \leq \tfrac{\|F(x)\|}{\sigma_{\min}(\nabla F(x)^*)}. \tag{4.3.20}$$

Доказательство.

Действительно, в силу леммы 4.3.6 существует такое h^*, что $F(x) + \nabla F(x) h^* = 0$ и

$$\|h^*\| \leq \tfrac{\|F(x)\|}{\sigma_{\min}(\nabla F(x)^*)}.$$

Таким образом,

$$\begin{array}{rcl} \frac{M}{2} r_M^2(x) & \leq & \psi(x; V_M(x)) + \frac{M}{2} r_M^2(x) \\ \\ & = & \min\limits_{h \in E_1} \left[\psi(x; x+h) + \frac{M}{2}\|h\|^2 \right] \\ \\ & \leq & \frac{M}{2}\|h^*\|^2 \leq \frac{M\|F(x)\|^2}{2\sigma_{\min}^2(\nabla F(x)^*)}. \qquad \Box \end{array}$$

Теперь мы можем получить глобальную оценку скорости сходимости метода (4.3.16).

Теорема 4.3.3 *Пусть предположения 4.3.1, 4.3.2 и 4.3.3 выполнены.*

1) Пусть последовательность $\{x_k\}_{k=0}^{\infty}$ сформирована методом (4.3.16). Если $f(x_k) \geq \frac{\sigma^2}{2L}\gamma_\phi^2$, то

$$f(x_{k+1}) \leq f(x_k) - \tfrac{\sigma^2}{4L}\gamma_\phi^2. \tag{4.3.21}$$

В противном случае,

$$f(x_{k+1}) \leq \tfrac{L}{\sigma^2\gamma_\phi^2} f^2(x_k) \leq \tfrac{1}{2} f(x_k). \tag{4.3.22}$$

2) Пусть последовательность $\{x_k\}_{k=0}^{\infty}$ сформирована методом (4.3.16) с $M_k \equiv L$. Если $f(x_k) \geq \frac{\sigma^2}{L}\gamma_\phi^2$, то

$$f(x_{k+1}) \leq f(x_k) - \tfrac{\sigma^2}{2L}\gamma_\phi^2. \tag{4.3.23}$$

В противном случае,

$$f(x_{k+1}) \leq \tfrac{L}{2\sigma^2\gamma_\phi^2} f^2(x_k) \leq \tfrac{1}{2} f(x_k). \tag{4.3.24}$$

Доказательство.

Докажем первое утверждение теоремы. Поскольку оператор $\nabla F(x_k)$ двойственно невырожден, в силу леммы 4.3.6 существует решение h_k^* системы линейных уравнений $F(x_k) + \nabla F(x_k) h = 0$ с ограниченной нормой:

$$\|h_k^*\| \leq \tfrac{1}{\sigma}\|F(x_k)\| \leq \tfrac{1}{\sigma\gamma_\phi} f(x_k).$$

Таким образом, в силу правил выбора шага в (4.3.16) и верхней оценки на значения M_k, имеем

$$\begin{aligned} f(x_{k+1}) &\le \min_{h \in E_1} \left[\phi(F(x_k) + \nabla F(x_k) h) + \tfrac{1}{2} M_k \|h\|^2 \right] \\ &\le \min_{t \in [0,1]} [\phi(F(x_k) + t \nabla F(x_k) h_k^*) + L \|t h_k^*\|^2] \\ &\le \min_{t \in [0,1]} \left[\phi((1-t)F(x_k)) + \tfrac{L}{\sigma^2 \gamma_\phi^2} t^2 f^2(x_k) \right] \\ &\le \min_{t \in [0,1]} \left[(1-t) f(x_k) + \tfrac{L}{\sigma^2 \gamma_\phi^2} t^2 f^2(x_k) \right] \end{aligned}$$

Итак, если $f(x_k) \le \frac{\sigma^2}{2L} \gamma_\phi^2$, то минимум в последней задаче одномерной минимизации достигается при $t = 1$, что приводит к неравенствам (4.3.22). В противном случае, минимум достигается при $t = \frac{\sigma^2 \gamma_\phi^2}{2L f(x_k)}$ и мы получаем оценку (4.3.21).

Второе утверждение теоремы доказывается аналогично. □

Теорема 4.3.3 помогает установить некоторые свойства задачи (4.3.6).

Теорема 4.3.4 *Пусть выполняются предположения 4.3.1, 4.3.2 и 4.3.3. Тогда существует такое решение x^* задачи (4.3.6), что $f(x^*) = 0$ и*

$$\|x^* - x_0\| \le \tfrac{2}{\sigma} \|F(x_0)\|. \tag{4.3.25}$$

Доказательство.
Выберем $\phi(u) = \|u\|$. Тогда $\gamma_\phi = 1$. Применим теперь метод (4.3.16) с шаговыми множителями $M_k \equiv L$ к соответствующей задаче (4.3.6) с целевой функцией $f(x) = \|F(x)\|$.

Предположим сначала, что $f(x_0) > \frac{\sigma^2}{L}$. В соответствии со вторым утверждением теоремы 4.3.3, пока значение $f(x_k) \ge \frac{\sigma^2}{L}$, будет выполняться неравенство

$$f(x_k) - f(x_{k+1}) \ge \tfrac{\sigma^2}{2L}. \tag{4.3.26}$$

Обозначим через N длину первой стадии процесса:

$$f(x_N) \ge \tfrac{\sigma^2}{L} \ge f(x_{N+1}).$$

Суммируя неравенства (4.3.26) для $k = 0, \dots, N$, получаем

$$N + 1 \le \tfrac{2L}{\sigma^2} (f(x_0) - f(x_{N+1})). \tag{4.3.27}$$

С другой стороны, в силу неравенства (4.3.11) имеем

$$f(x_k) - f(x_{k+1}) \ge \tfrac{L}{2} \|x_k - x_{k+1}\|^2. \tag{4.3.28}$$

Суммируя эти неравенства для $k = 0, \dots, N$, получаем

$$\begin{aligned} f(x_0) - f(x_{N+1}) &\geq \tfrac{L}{2} \sum_{k=0}^{N} \|x_k - x_{k+1}\|^2 \\ &\geq \tfrac{L}{2(N+1)} \left(\sum_{k=0}^{N} \|x_k - x_{k+1}\| \right)^2 \\ &\geq \tfrac{L}{2(N+1)} \|x_0 - x_{N+1}\|^2. \end{aligned}$$

Теперь, пользуясь неравенством (4.3.27), приходим к следующей оценке:

$$\begin{aligned} \|x_0 - x_{N+1}\| &\leq \left[\tfrac{2(N+1)}{L} (f(x_0) - f(x_{N+1})) \right]^{1/2} \\ &\leq \tfrac{2}{\sigma} (f(x_0) - f(x_{N+1})). \end{aligned} \tag{4.3.29}$$

Далее, в силу теоремы 4.3.3, на второй стадии процесса мы можем гарантировать, что

$$f(x_{k+1}) \leq \tfrac{L}{2\sigma^2} f^2(x_k) \leq \tfrac{1}{2} f(x_k), \quad k \geq N+1. \tag{4.3.30}$$

Таким образом, $f(x_{N+k+1}) \leq (\frac{1}{2})^k f(x_{N+1})$ для $k \geq 0$. Поэтому в силу неравенства (4.3.20) имеем

$$\|x_{N+k+2} - x_{N+k+1}\| \leq \tfrac{1}{\sigma} (\tfrac{1}{2})^k f(x_{N+1}), \quad k \geq 0.$$

Следовательно, последовательность $\{x_k\}_{k=0}^{\infty}$ сходится к точке x^* с $F(x^*) = 0$ и

$$\|x^* - x_{N+1}\| \leq \tfrac{2}{\sigma} f(x_{N+1}).$$

Принимая во внимание это неравенство и оценку (4.3.29), получаем нравенство (4.3.25).

Если $f(x_0) \leq \frac{\sigma^2}{L}$, то выщеприведенное рассуждение можно использовать с самого начала:

$$\begin{aligned} \sum_{k=0}^{\infty} \|x_{k+1} - x_k\| &\leq \tfrac{1}{\sigma} \sum_{k=0}^{\infty} f(x_k) \\ &\leq \tfrac{1}{\sigma} f(x_0) \sum_{k=0}^{\infty} (\tfrac{1}{2})^k \\ &= \tfrac{2}{\sigma} f(x_0). \qquad \Box \end{aligned}$$

Применяя в точности такие же рассуждения, как и при доказательстве теоремы 4.3.4, можно обосновать следующее утверждение.

Теорема 4.3.5 *Пусть выполнены предположения 4.3.1, 4.3.2 и 4.3.3. Предположим, что последовательность $\{x_k\}_{k=0}^{\infty}$ сформирована методом (4.3.16) применительно к задаче (4.3.6). Тогда эта последовательность сходится к единственной точке x^* с $F(x^*) = 0$.*

В завершение этого раздела приведем следующее замечание. Мы видели, что предположения 4.3.1, 4.3.2 и 4.3.3 гарантируют существование решения задачи (4.3.4). Обозначим

$$D = \min_x \{ \|x - x_0\| : \ x \in \mathcal{L}(f(x_0)), \ F(x) = 0 \}.$$

В силу следствия 4.3.1 и верхних границ на M_k в методе (4.3.16) можно всегда гарантировать, что

$$f(x_1) \leq \tfrac{3}{2} L D^2. \tag{4.3.31}$$

Таким образом, в силу теоремы 4.3.3 число итераций N метода (4.3.16), необходимых для вхождения в область квадратичной сходимости, может быть ограничено следующим образом:

$$N \leq 1 + \tfrac{4L}{\sigma^2 \gamma_\phi^2} f(x_1) \leq 1 + 6 \left(\tfrac{LD}{\sigma \gamma_\phi} \right)^2. \tag{4.3.32}$$

Мы будем ссылаться на эту оценку как на верхнюю оценку сложности класса задач, определяемого предположениями 4.3.1, 4.3.2 и 4.3.3. Эта оценка выводится из скорости сходимости модифицированного метода Гаусса-Ньютона (4.3.16).

4.3.5 Обсуждение

Сравнительный анализ метода (4.3.16)

Нам не удалось найти в литературе других методов решения системы нелинейных уравнений, для которых были бы известны глобальные оценки эффективности. Таким образом, нам приходится сравнивать метод (4.3.16) с единственным общим методом, для которого хоть что-то известно. Это кубическая регуляризация метода Ньютона (см. параграф 4.1). Заметим, что хотя области применения этих методов различны, они имеют нечто общее. Действительно, любая задача решения системы нелинейных уравнений может быть превращена в задачу минимизации с помощью какой-нибудь меры близости. С другой стороны, любая задача безусловной минимизации может быть сведена к решению системы нелинейных уравнений, возникающих в условии оптимальности первого порядка.

Рассмотрим следующую задачу безусловной минимизации.

$$\min_{x \in E_1} \varphi(x), \tag{4.3.33}$$

где у дважды дифференцированной сильно выпуклой функции $\varphi(x)$ гессиан удовлетворяет условию Липшица. Таким образом, предполагается, что существуют положительные константы σ и L, обеспечивающие выполнение условий

$$\begin{aligned} \langle \varphi''(x) h, h \rangle &\geq \sigma \|h\|^2, \\ \|\varphi''(x+h) - \varphi''(x)\| &\leq L \|h\|, \end{aligned} \tag{4.3.34}$$

при всех x и h из E_1. Обозначим $D = \|x_0 - x^*\|$. В параграфе 4.1, мы показали, что сложность задачи (4.3.33) для кубической регуляризации метода Ньютона (4.1.16) зависит от следующей характеристики:

$$\zeta = \tfrac{LD}{\sigma}$$

Если $\zeta < 1$, то задача (4.3.33) простая. В противоположном случае, число итераций модифицированного метода Ньютона, необходимых для вхождения в область квадратичной сходимости, ограничено величиной

$$N_1 = 6.25\sqrt{\zeta}, \tag{4.3.35}$$

(см. (4.1.51)).

Заметим, что задача (4.3.33) может быть записана и в форме (4.3.4):

$$\text{найти } x: \; F(x) \overset{\mathrm{def}}{=} \varphi'(x) = 0. \tag{4.3.36}$$

При этом $\nabla F(x) = \varphi''(x)$. Таким образом, в силу условий (4.3.34) наша задача (4.3.36) удовлетворяет предположениям 4.3.1, 4.3.2 и 4.3.3. Выберем $f(x) = \|F(x)\|$. Тогда в силу оценки (4.3.32) число итераций модифицированного метода Гаусса-Ньютона (4.3.16), необходимых для вхождения в область квадратичной сходимости, ограничено величиной

$$N_2 = 1 + 6\zeta^2. \tag{4.3.37}$$

Ясно, что оценка (4.3.35) намного лучше оценки (4.3.37). Однако это наблюдение только подтверждает стандартное правило, что специализированные методы должны быть более эффективны, чем общие. Вопрос состоит только в том насколько велико это преимущество. В настоящий момент нам к сожалению неизвестны никакие нижние оценки сложности для класса задач описываемых предположениями 4.3.1, 4.3.2 и 4.3.3. Поэтому имеются шансы что оценка (4.3.37) может быть улучшена другими методами.

На самом деле, по сравнению с кубической регуляризацией метода Ньютона (4.1.16), метод (4.3.16) имеет одно важное преимущество. Вспомогательная задача в КРН решается за полиномиальное время только для евклидовой нормы. Для модифицированого метода Гаусса-Ньютона нормы в пространствах E_1 и E_2 могут быть любыми. В п. 4.3.5 мы покажем, что в любом случае мы поучаем эффективно решаемую выпуклую вспомогательную задачу. В результате, можно надеяться что за счет выбора правильных норм отношение $\frac{L}{\sigma}$ можно делать небольшим.

Вычислительные аспекты

Изучим теперь вычислительную сложность вспомогательной задачи (4.3.10). Для простоты предположим, что мы выбрали $f(x) = \|F(x)\|$. В результате мы имеем дело со следующей задачей:

$$\text{найти } f_M(x) = \min_{h \in E_1} \left[\|F(x) + \nabla F(x) h\| + \tfrac{1}{2} M \|h\|^2 \right]. \tag{4.3.38}$$

Заметим, что иногда эта задача становится проще, если перейти к ее двойственной форме:

$$\begin{aligned}
\min_{h\in E_1}\left[\|F(x)+\nabla F(x)h\|+\tfrac{1}{2}M\|h\|^2\right] &= \min_{h\in E_1}\max_{\substack{s\in E_2^*\\ \|s\|\le 1}}\left[\langle s, F(x)+\nabla F(x)h\rangle+\tfrac{1}{2}M\|h\|^2\right]\\
&= \max_{\substack{s\in E_2^*\\ \|s\|\le 1}}\min_{h\in E_1}\left[\langle s, F(x)+\nabla F(x)h\rangle+\tfrac{1}{2}M\|h\|^2\right]\\
&= \max_{s\in E_2^*}\left[\langle s, F(x)\rangle-\tfrac{1}{2M}\|\nabla F(x)^*s\|^2 :\ \|s\|\le 1\right].
\end{aligned}$$

Таким образом, задача двойственная к (4.3.38) – это просто задача квадратичной максимизации с простым ограничением. Поскольку эта задача выпукла, мы можем применять к ней стандартные эффективные методы выпуклой минимизации.

Покажем, что в случае евклидовой нормы в задаче (4.3.38) можно обойтись средствами стандартной линейной алгебры.

Лемма 4.3.8 *Введем в пространствах E_1 и E_2 некоторые евклидовы нормы:*

$$\|x\| = \langle Q_1 x, x\rangle^{1/2},\ x\in E_1,\quad \|u\| = \langle Q_2 u, u\rangle^{1/2},\ u\in E_2.$$

Тогда задача (4.3.38) представляется в следующей двойственной форме:

$$f_M(x) = \min_{\lambda\in R}\left[\tfrac{1}{2}\lambda+\tfrac{1}{2}\langle(\lambda Q_2+\tfrac{1}{M}\nabla F(x)Q_1^{-1}\nabla F(x)^*)^{-1}F(x), F(x)\rangle :\ \lambda\ge 0\right]. \qquad (4.3.39)$$

Пусть λ^* будет оптимальным решением этой задачи. Тогда решением задачи (4.3.38) является

$$h^* = -\tfrac{1}{M}Q_1^{-1}\nabla F(x)^*\left(\lambda^* Q_2+\tfrac{1}{M}\nabla F(x)Q_1^{-1}\nabla F(x)^*\right)^{-1}F(x). \qquad (4.3.40)$$

Доказательство.
Действительно,

$$\begin{aligned}
f_M(x) &= \min_{h\in E_1}\max_{s\in E_2^*}\left[\langle s, F(x)+\nabla F(x)h\rangle+\tfrac{M}{2}\langle Q_1 h, h\rangle :\ \langle s, Q_2 s\rangle\le 1\right]\\
&= \max_{s\in E_2^*}\left[\langle s, F(x)\rangle-\tfrac{1}{2M}\langle Q_1^{-1}\nabla F(x)^*s, \nabla F(x)^*s\rangle :\ \langle s, Q_2 s\rangle\le 1\right]\\
&= \max_{s\in E_2^*}\min_{\lambda\ge 0\in R}\left[\langle s, F(x)\rangle-\tfrac{1}{2M}\langle Q_1^{-1}\nabla F(x)^*s, \nabla F(x)^*s\rangle+\tfrac{1}{2}\lambda(1-\langle s, Q_2 s\rangle)\right]\\
&= \min_{\lambda\ge 0\in R}\left[\tfrac{1}{2}\lambda+\tfrac{1}{2}\langle(\lambda Q_2+\tfrac{1}{M}\nabla F(x)Q_1^{-1}\nabla F(x)^*)^{-1}F(x), F(x)\rangle\right]. \qquad \square
\end{aligned}$$

Заметим, что задача одномерной минимизации (4.3.39) может эффективно решаться после подходящей факторизации обращаемых матриц.

Наконец, заметим, что в евклидовом случае наш шаг (4.3.40) может интерпретироваться как вариант подхода Левенберга-Маркварта со специальным правилом выбора параметра λ^*. В неевклидовой ситуации соответсвующая стратегия близка к методам с доверительной областью. Однако, основным преимуществом нашего подхода является его недвусмысленная геометрическая интерпретация, которая делает возможным глобальный и локальный анализ эффективности метода (4.3.16).

Глава 5

Техника сглаживания

5.1 Сглаживание для явной модели целевой функции

5.1.1 Введение

Мотивировка

Исторически первым методом решения негладких выпуклых задач был субградиентный метод (для подробных исторических комментариев см. монографии [117] и [24]). Очень скоро было показано, что его оценка эффективности составляет

$$O\left(\frac{1}{\epsilon^2}\right), \tag{5.1.1}$$

итераций, где ϵ – требуемая точность приближенного решения задачи по невязке по функции.

До сих пор разные варианты этого метода остаются привлекательными как для теоретиков, так и для практиков. Это не удивительно, так как медленная сходимость этих методов компенсируется очень низкой ценой одной итерации. Более того, как было показано в работе [79], оценка эффективности простейшего субградиентного метода *не может быть улучшена* равномерно по размерности пространства переменных. Конечно же это утверждение справедливо только для черноящичных моделей целевой функции. Однако доказательство этого факта конструктивно: можно показать что простейшая функция

$$\min_x \left\{ \max_{1\le i\le n} \sigma_i x^{(i)} : \sum_{i=1}^n (x^{(i)})^2 \le 1 \right\},$$

где веса σ_i с единичными модулями подбираются сопротивляющимся оракулом, является трудной для любого метода оптимизации. Возможно именно этот факт является причиной всеобщего убеждения в том что реальная сложность нахождения ϵ-приближения к минимуму кусочно-линейной функции действительно дается выражением (5.1.1).

На самом деле это не так. На практике у нас никогда не бывает чистой черноящичной модели. Мы всегда хоть что-нибудь знаем о структуре используемых объектов. И правильное использование этих знаний может и должно помочь нам в решении этой задачи.

В этом параграфе мы обсудим такую возможность. А именно, мы опишем систематическую процедуру аппроксимации исходной негладкой функции функцией с липшицевым градиентом. Затем, мы будем минимизировать новую функцию быстрым градиентным методом (1.2.18). Известно, что этот метод имеет оценку эффективности $O\left(\sqrt{\frac{L}{\epsilon}}\right)$ итераций, где L – константа Липшица для градиента целевой функции. Мы покажем что для сглаживающей функции, приближающей исходную с точностью ϵ, константа L может быть выбрана на уровне $\frac{1}{\epsilon}$. Таким образом, мы приходим к градиентному методу минимизации негладких функций с оценкой эффективности $O\left(\frac{1}{\epsilon}\right)$.

Содержание

В п. 5.1.2 мы описываем простую технику создания гладких аппроксимаций для негладких функций. Она основана на понятии *сопряженной* задачи, которое является обобщением понятие *двойственной* задачи. Для одной и той же прямой задачи, сопряженных задач может быть много и их размерность может отличаться от размерности исходной задачи. Обычно увеличение размерности сопряженной задачи приводит к упрощению ее структуры. В п. 5.1.3 мы применяем полученные результаты к различным конкретным постановкам: матричным играм, вариационным неравенствам с линейным оператором, задачам минимизации кусочно-линейных функций. Для всех них мы выписываем верхние оценки для числа итераций быстрого градиентного метода, необходимых для получения ϵ-решения прямой и сопряженной задачи. В п. 5.1.4 мы обсудим вычислительные аспекты и возможные модификации предложенных методов. И наконец, в п. 5.1.5 мы приведем результаты численных экспериментов. Мы сравним их с теоретическими оценками для методов отсечений и для полиномиальных алгоритмов внутренней точки. Как мы увидим, новые методы могут составить серьезную конкуренцию традиционным алгоритмам.

Обозначения

Напомним, что *норма* линейного оператора $A : E_1 \to E_2^*$ определяется следующим образом:

$$\|A\|_{1,2} = \max_{x,u}\{\langle Ax, u\rangle_2 : \ \|x\|_1 = 1, \ \|u\|_2 = 1\}.$$

Ясно, что

$$\|A\|_{1,2} = \|A^*\|_{2,1} = \max_{x}\{\|Ax\|_2^* : \ \|x\|_1 = 1\} = \max_{u}\{\|A^*u\|_1^* : \ \|u\|_2 = 1\}.$$

Поэтому для любых $x \in E_1$ и $u \in E_2$ имеем

$$\|Ax\|_2^* \leq \|A\|_{1,2} \cdot \|x\|_1, \quad \|A^*u\|_1^* \leq \|A\|_{1,2} \cdot \|u\|_2. \tag{5.1.2}$$

5.1.2 Гладкие аппроксимации негладких функций

В этом разделе мы рассматриваем задачу

$$\text{найти } f^* = \min_x \{ f(x) : \ x \in Q_1 \}, \tag{5.1.3}$$

где выпуклое замкнутое множество Q_1 в конечномерном вещественном пространстве E_1 является ограниченным, а выпуклая функция $f(x)$ непрерывна на Q_1. Мы не предполагаем, что f дифференцируема.

Очень часто нам известна *структура* целевой функции в задаче (5.1.3). Предположим, что эта структура описывается следующей *моделью*:

$$f(x) = \hat{f}(x) + \max_u \{ \langle Ax, u \rangle_2 - \hat{\phi}(u) : \ u \in Q_2 \}, \tag{5.1.4}$$

где выпуклая функция $\hat{f}(x)$ непрерывна на Q_1, выпуклое замкнутое множество Q_2 из конечномерного вещественного пространства E_2 ограничено, выпуклая функция $\hat{\phi}(u)$ непрерывна на Q_2, а линейный оператор A отображает E_1 в E_2^*. В этом случае задача (5.1.3) может быть переписана в *сопряженной форме*:

$$\begin{gathered} \max_u \{ \phi(u) : \ u \in Q_2 \}, \\ \phi(u) = -\hat{\phi}(u) + \min_x \{ \langle Ax, u \rangle_2 + \hat{f}(x) : \ x \in Q_1 \}. \end{gathered} \tag{5.1.5}$$

Заметим, что эта задача не всегда является полноценной заменой задачи (5.1.3), так как в ней мы неявно предполагали, что функция $\hat{\phi}(u)$ и множество Q_2 настолько простые, что решение внутренней оптимизационной задачи в представлении (5.1.4) может быть найдено в явном виде. Понятно, что для постановки (5.1.5) это предположение может иногда нарушаться.

Заметим, что одна и та же выпуклая функция $f(x)$ может иметь несколько различных представлений (5.1.4). Если мы возьмем, например,

$$Q_2 \equiv E_2 = E_1^*, \quad \hat{\phi}(u) \equiv f_*(u) = \max_x \{ \langle u, x \rangle_1 - f(x) : \ x \in E_1 \},$$

то $\hat{f}(x) \equiv 0$, и в качестве A можно взять единичный оператор. Однако, в этом случае функция $\hat{\phi}(u)$ может оказаться слишком сложной. Интуитивно ясно, что чем больше размерность пространства E_2, тем проще структура сопряженных объектов, функции $\hat{\phi}(u)$ и множества Q_2. Покажем это на простом примере.

Пример 5.1.1 Рассмотрим функцию $f(x) = \max\limits_{1 \le j \le m} |\langle a_j, x \rangle_1 - b^{(j)}|$. Можно взять $A = I$,

$E_2 = E_1^* = \mathbb{R}^n$ и

$$\begin{aligned} \hat{\phi}(u) &= \max_x \left\{ \langle u, x\rangle_1 - \max_{1\le j\le m} |\langle a_j, x\rangle_1 - b^{(j)}| \right\} \\ &= \max_x \min_{s\in\mathbb{R}^m} \left\{ \langle u, x\rangle_1 - \sum_{j=1}^m s^{(j)}[\langle a_j, x\rangle_1 - b^{(j)}] : \sum_{j=1}^m |s^{(j)}| \le 1 \right\} \\ &= \min_{s\in\mathbb{R}^m} \left\{ \sum_{j=1}^m s^{(j)} b^{(j)} : u = \sum_{j=1}^m s^{(j)} a_j, \sum_{j=1}^m |s^{(j)}| \le 1 \right\}. \end{aligned}$$

Ясно, что структура такой функции может быть очень сложной.

Рассмотрим другую возможность. Заметим, что

$$\begin{aligned} f(x) &= \max_{1\le j\le m} |\langle a_j, x\rangle_1 - b^{(j)}| \\ &= \max_{u\in\mathbb{R}^m} \left\{ \sum_{j=1}^m u^{(j)}[\langle a_j, x\rangle_1 - b^{(j)}] : \sum_{j=1}^m |u^{(j)}| \le 1 \right\}. \end{aligned}$$

В этом случае $E_2 = \mathbb{R}^m$, $\hat{\phi}(u) = \langle b, u\rangle_2$ и $Q_2 = \{u \in \mathbb{R}^m : \sum\limits_{j=1}^m |u^{(j)}| \le 1\}$.

Наконец, можно представить $f(x)$ следующим образом:

$$f(x) = \max_{u=(u_1,u_2)\in\mathbb{R}^{2m}} \left\{ \sum_{j=1}^m (u_1^{(j)} - u_2^{(j)}) \cdot [\langle a_j, x\rangle_1 - b^{(j)}] : \sum_{j=1}^m (u_1^{(j)} + u_2^{(j)}) = 1, \; u \ge 0 \right\}.$$

В этом случае $E_2 = \mathbb{R}^{2m}$, функция $\hat{\phi}(u)$ линейна, а множество Q_2 - симплекс. В п. 5.1.3 мы увидим, что такое представление является наилучшим. □

Покажем как знание структуры (5.1.4) может нам помочь в решении задач (5.1.3) и (5.1.5). Мы будем пользоваться этой структурой для построения гладкой аппроксимации целевой функции в задаче (5.1.3).

Рассмотрим *прокс-функцию* $d_2(u)$ множества Q_2. Эта функция непрерывна и сильно выпукла на Q_2 с параметром выпуклости $\sigma_2 = 1$. Обозначим через

$$u_0 = \arg\min_u \{d_2(u) : u \in Q_2\}$$

ее *прокс-центр*. Не ограничивая общности считаем, что $d_2(u_0) = 0$. Таким образом, для любого $u \in Q_2$ имеем

$$d_2(u) \ge \tfrac{1}{2}\|u - u_0\|_2^2. \tag{5.1.6}$$

Выберем положительное μ в качестве *параметра сглаживания*. Рассмотрим следующую функцию:

$$f_\mu(x) = \max_u \{\langle Ax, u\rangle_2 - \hat{\phi}(u) - \mu d_2(u) : u \in Q_2\}. \tag{5.1.7}$$

Обозначим через $u_\mu(x)$ оптимальное решение этой задачи. Поскольку функция $d_2(u)$ сильно выпукла, это решение единственно.

Теорема 5.1.1 *Функция $f_\mu(x)$ определена и непрерывно дифференцируема при всех $x \in E_1$. Более того, она выпукла и ее градиент, вычисляемый по формуле*

$$\nabla f_\mu(x) = A^* u_\mu(x), \tag{5.1.8}$$

непрерывен по Липшицу с константой

$$L_\mu = \tfrac{1}{\mu}\|A\|_{1,2}^2.$$

Доказательство.

Действительно, функция $f_\mu(x)$ является выпуклой как максимум линейных по x функций. Она дифференцируема так как точка $u_\mu(x)$ однозначно определена. Покажем, что ее градиент липшицев. Зафиксируем две точки x_1 и x_2. Для упрощения обозначений, не ограничивая общности можно считать, что функции $\hat{\phi}(\cdot)$ и $d_2(\cdot)$ дифференцируемы. Из условия оптимальности первого порядка имеем

$$\langle Ax_1 - \nabla\hat{\phi}(u_\mu(x_1)) - \mu\nabla d_2(u_\mu(x_1)), u_\mu(x_2) - u_\mu(x_1)\rangle_2 \le 0,$$

$$\langle Ax_2 - \nabla\hat{\phi}(u_\mu(x_2)) - \mu\nabla d_2(u_\mu(x_2)), u_\mu(x_1) - u_\mu(x_2)\rangle_2 \le 0.$$

Складывая эти неравенства и пользуясь выпуклостью функции $\hat{\phi}(\cdot)$ и сильной выпуклостью функции $d_2(\cdot)$ получаем:

$$\begin{array}{rl} & \langle A(x_1 - x_2), u_\mu(x_1) - u_\mu(x_2)\rangle_2 \\ \ge & \langle \nabla\hat{\phi}(u_\mu(x_1)) - \nabla\hat{\phi}(u_\mu(x_2)) + \mu(\nabla d_2(u_\mu(x_1)) - \nabla d_2(u_\mu(x_2))), u_\mu(x_1) - u_\mu(x_2)\rangle_2 \\ \ge & \mu\langle \nabla d_2(u_\mu(x_1)) - \nabla d_2(u_\mu(x_2)), u_\mu(x_1) - u_\mu(x_2)\rangle_2 \\ \ge & \mu\|u_\mu(x_1) - u_\mu(x_2)\|_2^2. \end{array}$$

Таким образом, в силу (5.1.2), имеем

$$\begin{array}{rl} (\|A^*u_\mu(x_1) - A^*u_\mu(x_2))\|_1^*)^2 & \le \|A\|_{1,2}^2 \cdot \|u_\mu(x_1) - u_\mu(x_2)\|_2^2 \\ & \le \frac{1}{\mu}\|A\|_{1,2}^2 \langle A^*(u_\mu(x_1) - u_\mu(x_2)), x_1 - x_2\rangle_1 \\ & \le \frac{1}{\mu}\|A\|_{1,2}^2 \cdot \|A^*u_\mu(x_1) - A^*u_\mu(x_2)\|_1^* \cdot \|x_1 - x_2\|_1. \end{array}$$

□

Обозначим $D_2 = \max\limits_u \{d_2(u) : \ u \in Q_2\}$ и $f_0(x) = \max\limits_u \{\langle Ax, u\rangle_2 - \hat{\phi}(u) : \ u \in Q_2\}$. Тогда для любого $x \in E_1$ выполнено неравенство

$$f_\mu(x) \le f_0(x) \le f_\mu(x) + \mu D_2. \tag{5.1.9}$$

Таким образом, при $\mu > 0$ функция $f_\mu(x)$ является равномерной гладкой аппроксимацией функции $f_0(x)$.

5.1.3 Примеры приложений

Покажем, как можно использовать результаты предыдущего раздела. Предположим, что функция $\hat{f}(\cdot)$ в (5.1.4) является дифференцируемой и ее градиент удовлетворяет условию Липшица с некоторой константой $M \ge 0$. Тогда техника сглаживания, применяемая к задаче (5.1.3), дает следующую целевую функцию:

$$\bar{f}_\mu(x) = \hat{f}(x) + f_\mu(x) \quad \to \quad \min : \ x \in Q_1. \tag{5.1.10}$$

В силу теоремы 5.1.1, ее градиент удовлетворяет условию Липшица с константой

$$L_\mu = M + \tfrac{1}{\mu}\|A\|^2_{1,2}.$$

Выберем теперь некоторую прокс-функцию $d_1(x)$ для множества Q_1 с параметром выпуклости $\sigma_1 = 1$. Напомним что мы предполагаем ограниченность множества Q_1:

$$\max_x \{d_1(x) : \ x \in Q_1\} \le D_1.$$

Теорема 5.1.2 *Пусть метод (1.2.18) применяется для решения задачи (5.1.10) со следующим значением параметра сглаживания:*

$$\mu = \mu(N) = \tfrac{2\|A\|_{1,2}}{N+1} \cdot \sqrt{\tfrac{D_1}{D_2}}.$$

Тогда после N итераций можно сформировать приближенные решения задач (5.1.3) и (5.1.5), а именно,

$$\hat{x} = y_N \in Q_1, \quad \hat{u} = \sum_{i=0}^{N} \tfrac{2(i+1)}{(N+1)(N+2)}\, u_\mu(x_i) \in Q_2, \tag{5.1.11}$$

которые удовлетворяют следующему неравенству:

$$0 \le f(\hat{x}) - \phi(\hat{u}) \le \tfrac{4\|A\|_{1,2}}{N+1} \cdot \sqrt{D_1 D_2} + \tfrac{4MD_1}{(N+1)^2}. \tag{5.1.12}$$

Таким образом, сложность нахождения ϵ-решения для задач (5.1.3), (5.1.5) с помощью техники сглаживания не превосходит

$$4\|A\|_{1,2}\sqrt{D_1 D_2} \cdot \tfrac{1}{\epsilon} \ + \ 2\sqrt{\tfrac{MD_1}{\epsilon}} \tag{5.1.13}$$

итераций.

Доказательство.

Зафиксируем произвольное $\mu > 0$. В силу теоремы 1.2.3, после N итераций метода (1.2.18) можно получить такую точку $\hat{x} = y_N$, что

$$\bar{f}_\mu(\hat{x}) \le \tfrac{L_\mu D_1}{(N+1)^2} + \min_x \left\{ \sum_{i=0}^{N} \tfrac{2(i+1)}{(N+1)(N+2)} [\bar{f}_\mu(x_i) + \langle \nabla \bar{f}_\mu(x_i), x - x_i \rangle_1] : x \in Q_1 \right\}. \tag{5.1.14}$$

Заметим, что

$$\begin{aligned} f_\mu(x) &= \max_u \{ \langle Ax, u \rangle_2 - \hat{\phi}(u) - \mu d_2(u) : \ u \in Q_2 \} \\ &= \langle Ax, u_\mu(x) \rangle_2 - \hat{\phi}(u_\mu(x)) - \mu d_2(u_\mu(x)), \\ \langle \nabla f_\mu(x), x \rangle_1 &= \langle A^* u_\mu(x), x \rangle_1. \end{aligned}$$

Таким образом,

$$f_\mu(x_i) - \langle \nabla f_\mu(x_i), x_i \rangle_1 = -\hat{\phi}(u_\mu(x_i)) - \mu d_2(u_\mu(x_i)), \quad i = 0, \ldots, N. \tag{5.1.15}$$

Следовательно, в силу (5.1.8) и (5.1.15) имеем

$$\begin{aligned} &\sum_{i=0}^{N} (i+1)[\bar{f}_\mu(x_i) + \langle \nabla \bar{f}_\mu(x_i), x - x_i \rangle_1] \\ &\le \sum_{i=0}^{N} (i+1)[f_\mu(x_i) - \langle \nabla f_\mu(x_i), x_i \rangle_1] + \tfrac{1}{2}(N+1)(N+2)(\hat{f}(x) + \langle A^* \hat{u}, x \rangle_1) \\ &\le -\sum_{i=0}^{N} (i+1) \hat{\phi}(u_\mu(x_i)) + \tfrac{1}{2}(N+1)(N+2)(\hat{f}(x) + \langle A^* \hat{u}, x \rangle_1) \\ &\le \tfrac{1}{2}(N+1)(N+2)[-\hat{\phi}(\hat{u}) + \hat{f}(x) + \langle Ax, \hat{u} \rangle_2]. \end{aligned}$$

Поэтому, пользуясь неравенствами (5.1.14), (5.1.9) и представлением (5.1.5), получаем следующую оценку:

$$\tfrac{L_\mu D_1}{(N+1)^2} \ge \bar{f}_\mu(\hat{x}) - \phi(\hat{u}) \ge f(\hat{x}) - \phi(\hat{u}) - \mu D_2.$$

Это приводит к неравенству

$$0 \le f(\hat{x}) - \phi(\hat{u}) \le \mu D_2 + \tfrac{4\|A\|_{1,2}^2 D_1}{\mu (N+1)^2} + \tfrac{4 M D_1}{(N+1)^2}. \tag{5.1.16}$$

Минимизируя правую часть этого неравенства по μ, получаем неравенство (5.1.12). □

Заметим, что оценка эффективности (5.1.13) гораздо лучше стандартной оценки $O(\frac{1}{\epsilon^2})$. В соответствии с вышеприведенной теоремой, в случае $M = 0$ оптимальная зависимость параметров μ, L_μ и N от ϵ будет такой:

$$\mu = \tfrac{\epsilon}{2D_2}, \quad L_\mu = D_2 \cdot \tfrac{\|A\|_{1,2}^2}{\epsilon}, \quad N + 1 = 4\|A\|_{1,2} \sqrt{D_1 D_2} \cdot \tfrac{1}{\epsilon}. \tag{5.1.17}$$

Рассмотрим теперь несколько примеров.

Минимаксные стратегии для матричных игр

Обозначим через Δ_n стандартный симплекс в $\mathbb{R}^n$:

$$\Delta_n = \{x \in \mathbb{R}^n : \ x \geq 0, \ \sum_{i=1}^{n} x^{(i)} = 1\}.$$

Пусть $A : \mathbb{R}^n \to \mathbb{R}^m$, $E_1 = \mathbb{R}^n$ и $E_2 = \mathbb{R}^m$. Рассмотрим следующую седловую задачу:

$$\min_{x\in\Delta_n} \max_{u\in\Delta_m} \{\langle Ax, u\rangle_2 + \langle c, x\rangle_1 + \langle b, u\rangle_2\}. \tag{5.1.18}$$

С точки зрения игроков эта задача может быть сведена к задаче негладкой оптимизации:

$$\begin{aligned} &\min_{x\in\Delta_n} f(x), \quad f(x) = \langle c, x\rangle_1 + \max_{1\leq j\leq m} [\langle a_j, x\rangle_1 + b^{(j)}], \\ &\max_{u\in\Delta_m} \phi(u), \quad \phi(u) = \langle b, u\rangle_2 + \min_{1\leq i\leq n} [\langle \hat{a}_i, u\rangle_2 + c^{(i)}], \end{aligned} \tag{5.1.19}$$

где a_j – столбцы, а $\hat{a}_i$ – строки матрицы A. Для того чтобы решить эту пару задач с помощью техники сглаживания, нам потребуется прокс-функция для симплекса. Рассмотрим две возможности.

1. Евклидово расстояние. Выберем

$$\|x\|_1 = \left[\sum_{i=1}^{n} (x^{(i)})^2\right]^{1/2}, \quad d_1(x) = \tfrac{1}{2}\sum_{i=1}^{n} (x^{(i)} - \tfrac{1}{n})^2,$$

$$\|u\|_2 = \left[\sum_{j=1}^{m} (u^{(j)})^2\right]^{1/2}, \quad d_2(x) = \tfrac{1}{2}\sum_{j=1}^{m} (u^{(j)} - \tfrac{1}{m})^2.$$

Тогда $\sigma_1 = \sigma_2 = 1$, $D_1 = 1 - \frac{1}{n} < 1$, $D_2 = 1 - \frac{1}{m} < 1$ и

$$\|A\|_{1,2} = \max_u \{\|Ax\|_2^* : \ \|x\|_1 = 1\} = \lambda_{\max}^{1/2}(A^T A).$$

Итак, в этом случае оценка (5.1.12) для результата (5.1.11) выглядит следующим образом:

$$0 \leq f(\hat{x}) - \phi(\hat{u}) \leq \tfrac{4\lambda_{\max}^{1/2}(A^T A)}{N+1}. \tag{5.1.20}$$

2. Энтропийное расстояние. Выберем

$$\|x\|_1 = \sum_{i=1}^{n} |x^{(i)}|, \quad d_1(x) = \ln n + \sum_{i=1}^{n} x^{(i)} \ln x^{(i)},$$

$$\|u\|_2 = \sum_{j=1}^{m} |u^{(j)}|, \quad d_2(u) = \ln m + \sum_{j=1}^{m} u^{(j)} \ln u^{(j)}.$$

Лемма 5.1.1 *При таком выборе прокс-функции получаем*

$$\sigma_1 = \sigma_2 = 1, \quad D_1 = \ln n, \quad D_2 = \ln m.$$

Доказательство.

Заметим, что функция $d_1(x)$ дважды непрерывно дифференцируема и $\langle d_1''(x)h,h\rangle = \sum\limits_{i=1}^{n} \frac{(h^{(i)})^2}{x^{(i)}}$. Остается воспользоваться следующим вариантом неравенства Коши–Буняковского:

$$\left(\sum_{i=1}^{n} |h^{(i)}|\right)^2 \leq \left(\sum_{i=1}^{n} x^{(i)}\right) \cdot \left(\sum_{i=1}^{n} \frac{(h^{(i)})^2}{x^{(i)}}\right),$$

которое справедливо при всех положительных векторах x. Для функции $d_2(u)$ рассуждение аналогично. □

В данном случае мы должны пользоваться следующей нормой оператора A:

$$\|A\|_{1,2} = \max_x\{\max_{1\leq j\leq m} |\langle a_j,x\rangle| : \ \|x\|_1 = 1\} = \max_{i,j} |A^{(i,j)}|.$$

Таким образом, при использовании энтропийных расстояний, оценка (5.1.12) превращается в следующую:

$$0 \leq f(\hat{x}) - \phi(\hat{u}) \leq \tfrac{4\sqrt{\ln n \ln m}}{N+1} \cdot \max_{i,j} |A^{(i,j)}|. \tag{5.1.21}$$

Заметим, что обычно оценка (5.1.21) намного лучше чем ее евклидов вариант (5.1.20).

Запишем теперь явное выражение для гладкой аппроксимации целевой функции в первой задаче в (5.1.19), используя энтропийные расстояния. По определению,

$$\bar{f}_\mu(x) = \langle c,x\rangle_1 + \max_{u\in\Delta_m}\left\{\sum_{j=1}^{m} u^{(j)}[\langle a_j,x\rangle + b^{(j)}] - \mu\sum_{j=1}^{m} u^{(j)}\ln u^{(j)} - \mu\ln m\right\}.$$

Воспользуемся следующим результатом.

Лемма 5.1.2 *Решение задачи*

$$\text{найти } \phi_*(s) = \max_{u\in\Delta_m}\left\{\sum_{j=1}^{m} u^{(j)}s^{(j)} - \mu\sum_{j=1}^{m} u^{(j)}\ln u^{(j)}\right\} \tag{5.1.22}$$

дается вектором $u_\mu(s)\in\Delta_m$ с координатами

$$u_\mu^{(j)}(s) = \frac{e^{s^{(j)}/\mu}}{\sum\limits_{l=1}^{m} e^{s^{(l)}/\mu}}, \quad j = 1,\ldots,m. \tag{5.1.23}$$

Таким образом $\phi_*(s) = \mu\ln\left(\sum\limits_{l=1}^{m} e^{s^{(l)}/\mu}\right)$.

Доказательство.

Действительно, условия оптимальности для задачи (5.1.22) выглядят следующим образом:

$$\begin{aligned} s^{(j)} - \mu(1+\ln u^{(j)}) &= \lambda, \ j = 1,\ldots,m, \\ \sum_{j=1}^{m} u^{(j)} &= 1. \end{aligned}$$

Ясно, что они выполняются в точке (5.1.23) с $\lambda = \mu \ln \left(\sum\limits_{l=1}^{m} e^{s^{(l)}/\mu} \right) - \mu$. □

Пользуясь леммой 5.1.2, мы заключаем, что в нашем случае задача (5.1.10) выглядит следующим образом:

$$\bar{f}_\mu(x) = \langle c, x\rangle_1 + \mu \ln \left(\tfrac{1}{m} \sum_{j=1}^{m} e^{[\langle a_j, x\rangle + b^{(j)}]/\mu} \right) \quad \to \quad \min: \ x \in \Delta_n.$$

Заметим, что трудоемкость оракула этой задачи практически такая же, как и у исходной задачи (5.1.19).

Непрерывная задача размещения

Рассмотрим следующую *задачу размещения*. Имеется p городов с населением m_j, которые расположены в точках $c_j \in \mathbb{R}^n$, $j = 1, \ldots, p$. Нам нужно построить сервисный центр в некоторой точке $x \in \mathbb{R}^n \equiv E_1$ так, чтобы общая цена $f(x)$ транспортных расходов на обслуживание была минимальной. При этом центр не должен находиться слишком далеко от начала координат.

Математически эта задача записывается следующим образом:

$$\text{найти } f^* = \min_x \left\{ f(x) = \sum_{j=1}^{p} m_j \|x - c_j\|_1 : \ \|x\|_1 \le \bar{r} \right\}. \tag{5.1.24}$$

В соответствии с интерпретацией, естественно выбрать

$$\|x\|_1 = \left[\sum_{i=1}^{n} (x^{(i)})^2 \right]^{1/2}, \quad d_1(x) = \tfrac{1}{2}\|x\|_1^2.$$

Тогда $\sigma_1 = 1$ и $D_1 = \frac{1}{2}\bar{r}^2$.

В нашем случае структура сопряженного пространства E_2 достаточно проста:

$$E_2 = (E_1^*)^p, \quad Q_2 = \{u = (u_1, \ldots, u_p) \in E_2 : \ \|u_j\|_1^* \le 1, \ j = 1, \ldots, p\}.$$

Можно выбрать

$$\|u\|_2 = \left[\sum_{j=1}^{p} m_j (\|u_j\|_1^*)^2 \right]^{1/2}, \quad d_2(u) = \tfrac{1}{2}\|u\|_2^2.$$

Тогда $\sigma_2 = 1$ и $D_2 = \frac{1}{2}P$, где $P \overset{\text{def}}{=} \sum\limits_{j=1}^{p} m_j$. Таким образом, значение P имеет интерпретацию суммарного населения всех городов.

Остается вычислить норму оператора A:

$$\begin{aligned} \|A\|_{1,2} &= \max_{x,u} \left\{ \sum_{j=1}^{p} m_j \langle u_j, x\rangle_1 : \ \sum_{j=1}^{p} m_j (\|u_j\|_1^*)^2 = 1, \ \|x\|_1 = 1 \right\} \\ &= \max_{r_j} \left\{ \sum_{j=1}^{p} m_j r_j : \ \sum_{j=1}^{p} m_j r_j^2 = 1 \right\} = P^{1/2}. \end{aligned}$$

Подставляя вычисленные значения в оценку (5.1.12), получаем следующую скорость сходимости:

$$f(\hat{x}) - f^* \leq \tfrac{2P\bar{r}}{N+1}. \tag{5.1.25}$$

Заметим, что значение $\tilde{f}(x) = \frac{1}{P} f(x)$ соответствует средним индивидуальным затратам при постройке центра в точке x. Таким образом,

$$\tilde{f}(\hat{x}) - \tilde{f}^* \leq \tfrac{2\bar{r}}{N+1}.$$

Интересно, что правая часть этого неравенства не зависит от размерности. В то же время ясно, что разумная точность решения этой задачи не может быть большой. Учитывая низкую трудоемкость каждой итерации в методе (1.2.18), эффективность предложенного метода в данном случае оказывается очень высокой.

В завершение этого пункта выпишем явное выражение для сглаженной целевой функции:

$$\begin{aligned} f_\mu(x) &= \max_u \left\{ \sum_{j=1}^{p} m_j \langle u_j, x - c_j \rangle_1 - \mu d_2(u) : \; u \in Q_2 \right\} \\ &= \max_u \left\{ \sum_{j=1}^{p} m_j \left(\langle u_j, x - c_j \rangle_1 - \tfrac{1}{2}\mu (\|u_j\|_1^*)^2 \right) : \; \|u_j\|_1^* \leq 1, \; j = 1, \ldots, p \right\} \\ &= \sum_{j=1}^{p} m_j \psi_\mu(\|x - c_j\|_1), \end{aligned}$$

где функция $\psi_\mu(\tau)$, $\tau \geq 0$, определяется так:

$$\psi_\mu(\tau) = \max_{\gamma \in [0,1]} \{\gamma\tau - \tfrac{1}{2}\mu\gamma^2\} = \begin{cases} \frac{\tau^2}{2\mu}, & \text{если } 0 \leq \tau \leq \mu, \\ \tau - \frac{\mu}{2}, & \text{если } \mu \leq \tau. \end{cases} \tag{5.1.26}$$

Вариационные неравенства с линейным оператором

Рассмотрим линейный оператор $B(w) = Bw + c$: $E \to E^*$, который является *монотонным*:

$$\langle Bh, h \rangle \geq 0 \quad \forall h \in E_1.$$

Пусть выпуклое замкнутое множество Q в E будет ограниченным. Нам нужно решить следующее *вариационное неравенство*:

$$\text{найти } w^* \in Q : \quad \langle B(w^*), w - w^* \rangle \geq 0 \quad \forall w \in Q. \tag{5.1.27}$$

Заметим, что задача (5.1.27) может быть переписана в виде задачи минимизации. Действительно, положим

$$\psi(w) = \max_v \{ \langle B(v), w - v \rangle : \; v \in Q \}.$$

Ясно, что $\psi(w)$ – выпуклая функция. Как известно, задача

$$\min_{w}\{\psi(w):\ w \in Q\} \tag{5.1.28}$$

эквивалентна задаче (5.1.27). Для полноты картины снабдим этот факт простым доказательством.

Лемма 5.1.3 *Точка w^* является решением задачи (5.1.28) тогда и только тогда когда она решает вариационное неравенство (5.1.27). Для таких w^* имеем $\psi(w^*) = 0$.*

Доказательство.
Действительно, в любой точке $w \in Q$ функция ψ неотрицательна. Если w^* – решение задачи (5.1.27), то для любого $v \in Q$ выполнено неравенство

$$\langle B(v), v - w^*\rangle \geq \langle B(w^*), v - w^*\rangle \geq 0.$$

Поэтому $\psi(w^*) = 0$ и $w^* \in \operatorname{Arg}\min\limits_{w \in Q} \psi(w)$.

Рассмотрим теперь любую точку $w^* \in Q$ с $\psi(w^*) = 0$. Тогда для всех $v \in Q$ имеем

$$\langle B(v), v - w^*\rangle \geq 0.$$

Предположим, что существует такая точка $v_1 \in Q$, что $\langle B(w^*), v_1 - w^*\rangle < 0$. Рассмотрим точки

$$v_\alpha = w^* + \alpha(v_1 - w^*), \quad \alpha \in [0, 1].$$

Тогда

$$\begin{aligned} 0 &\leq \langle B(v_\alpha), v_\alpha - w^*\rangle \\ &= \alpha\langle B(v_\alpha), v_1 - w^*\rangle \\ &= \alpha\langle B(w^*), v_1 - w^*\rangle + \alpha^2\langle B \cdot (v_1 - w^*), v_1 - w^*\rangle. \end{aligned}$$

Поэтому при достаточно маленьком α мы приходим к противоречию. □

Ясно, что существует две возможности для представления задач (5.1.27), (5.1.28) в виде (5.1.3), (5.1.4).

1. Прямая форма. Выберем $E_1 = E_2 = E$, $Q_1 = Q_2 = Q$, $d_1(x) = d_2(x) = d(x)$, $A = B$ и

$$\hat{f}(x) = \langle b, x\rangle_1, \quad \hat{\phi}(u) = \langle b, u\rangle_1 + \langle Bu, u\rangle_1.$$

Заметим, что квадратичная функция $\hat{\phi}(u)$ является выпуклой. Для вычисления значения функции $f_\mu(x)$ необходимо решить следующую задачу:

$$\max_{u}\{\langle Bx, u\rangle_1 - \mu d(u) - \langle b, u\rangle_1 + \langle Bu, u\rangle_1 :\ u \in Q\}. \tag{5.1.29}$$

Поскольку в нашем случае $M = 0$, теорема 5.1.2 дает следующую оценку сложности задачи (5.1.27):

$$\tfrac{4D_1\|B\|_{1,2}}{\epsilon}. \tag{5.1.30}$$

Однако заметим, что в силу наличия нетривиальной квадратичной функции в целевой функции задачи (5.1.29), ее вычисление может быть достаточно сложным. Эта неприятность отсутствует в двойственном представлении задачи.

2. Двойственная форма. Рассмотрим двойственный вариант задачи (5.1.28):

$$\begin{aligned} \min_{w\in Q}\max_{v\in Q}\langle B(v), w-v\rangle &= \max_{v\in Q}\min_{w\in Q}\langle B(v), w-v\rangle \\ &= -\min_{v\in Q}\max_{w\in Q}\langle B(v), v-w\rangle. \end{aligned}$$

Таким образом, можно взять $E_1 = E_2 = E$, $Q_1 = Q_2 = Q$, $d_1(x) = d_2(x) = d(x)$, $A = B$ и

$$\hat{f}(x) = \langle b, x\rangle_1 + \langle Bx, x\rangle_1, \quad \hat{\phi}(u) = \langle b, u\rangle_1.$$

Теперь вычисление функции $f_\mu(x)$ стало гораздо проще:

$$f_\mu(x) = \max_u\{\langle Bx, u\rangle_1 - \mu d(u) - \langle b, u\rangle_1 : \ u \in Q\}.$$

Интересно, что стоимость этого упрощения вполне приемлема. Действительно, теперь $M = \|B\|_{1,2}$. Поэтому оценка сложности (5.1.30) вырастет до следующего уровня:

$$\tfrac{4D_1\|B\|_{1,2}}{\epsilon} + \sqrt{\tfrac{D_1\|B\|_{1,2}}{\epsilon}}.$$

Заметим, что важным частным случаем неотрицательно определенного оператора B является кососимметрический оператор. Для него $B + B^* = 0$ и, следовательно, сложность прямого и двойственного вариантов совпадают.

Минимизация кусочно-линейных функций

1. Максимум абсолютных значений. Рассмотрим следующую задачу:

$$\min_x\left\{ f(x) = \max_{1\le j\le m}|\langle a_j, x\rangle_1 - b^{(j)}| : \ x \in Q_1\right\}. \tag{5.1.31}$$

Для простоты выберем

$$\|x\|_1 = \left[\sum_{i=1}^n (x^{(i)})^2\right]^{1/2}, \quad d_1(x) = \tfrac{1}{2}\|x\|^2.$$

Обозначим через A матрицу со строками a_j, $j = 1, \ldots, m$. Удобно взять

$$E_2 = \mathbb{R}^{2m}, \quad \|u\|_2 = \sum_{j=1}^{2m}|u^{(j)}|, \quad d_2(u) = \ln(2m) + \sum_{j=1}^{2m} u^{(j)}\ln u^{(j)}.$$

Тогда

$$f(x) = \max_u \{\langle \hat{A}x, u\rangle_2 - \langle \hat{b}, u\rangle_2 : \ u \in \Delta_{2m}\},$$

где $\hat{A} = \begin{pmatrix} A \\ -A \end{pmatrix}$ и $\hat{b} = \begin{pmatrix} b \\ -b \end{pmatrix}$. Таким образом, $\sigma_1 = \sigma_2 = 1$, $D_2 = \ln(2m)$, и

$$D_1 = \tfrac{1}{2}\bar{r}^2, \quad \bar{r} = \max_x \{\|x\|_1 : \ x \in Q_1\}.$$

Остается вычислить норму оператора $\hat{A}$:

$$\begin{aligned} \|\hat{A}\|_{1,2} &= \max_{x,u}\{\langle \hat{A}x, u\rangle_2 : \ \|x\|_1 = 1, \ \|u\|_2 = 1\} \\ &= \max_x \{\max_{1\le j\le m} |\langle a_j, x\rangle_1| : \ \|x\|_1 = 1\} \\ &= \max_{1\le j\le m} \|a_j\|_1^*. \end{aligned}$$

Подставляя вычисленные значения в оценку (5.1.13), мы видим, что задача (5.1.31) может быть решена за

$$2\sqrt{2}\ \bar{r} \max_{1\le j\le m} \|a_j\|_1^* \sqrt{\ln(2m)} \cdot \tfrac{1}{\epsilon}$$

итераций метода (1.2.18). Стандартные градиентные методы могут гарантировать в этой ситуации только оценку порядка

$$O\left(\left[\bar{r} \max_{1\le j\le m} \|a_j\|_1^* \cdot \tfrac{1}{\epsilon}\right]^2\right)$$

итераций.

Наконец, сглаженный вариант целевой функции в задаче (5.1.31) выглядит следующим образом:

$$\bar{f}_\mu(x) = \mu \ln\left(\tfrac{1}{m} \sum_{j=1}^m \xi\left(\tfrac{1}{\mu}[\langle a_j, x\rangle + b^{(j)}]\right)\right)$$

с $\xi(\tau) = \frac{1}{2}[e^\tau + e^{-\tau}]$.

2. Сумма абсолютных значений. Рассмотрим теперь задачу

$$\min_x \left\{ f(x) = \sum_{j=1}^m |\langle a_j, x\rangle_1 - b^{(j)}| : \ x \in Q_1 \right\}. \tag{5.1.32}$$

Наиболее простое представление функции $f(x)$ выглядит следующим образом. Обозначим через A матрицу со строками a_j. Выберем

$$E_2 = \mathbb{R}^m, \quad Q_2 = \{u \in \mathbb{R}^m : \ |u^{(j)}| \le 1, \ j = 1, \dots, m\},$$

$$d_2(u) = \tfrac{1}{2}\|u\|_2^2 = \tfrac{1}{2} \sum_{j=1}^m \|a_j\|_1^* \cdot (u^{(j)})^2.$$

Тогда сглаженная версия целевой функции выглядит так:

$$\begin{aligned} f_\mu(x) &= \max_u \{ \langle Ax - b, u\rangle_2 - \mu d_2(u) : \; u \in Q_2 \} \\ &= \sum_{j=1}^{m} \|a_j\|_1^* \cdot \psi_\mu\left(\tfrac{|\langle a_j, x\rangle_1 - b^{(j)}|}{\|a_j\|_1^*} \right), \end{aligned}$$

где функция $\psi_\mu(\tau)$ определена в формуле (5.1.26). Заметим, что

$$\begin{aligned} \|A\|_{1,2} &= \max_{x,u} \left\{ \sum_{j=1}^{m} u^{(j)} \langle a_j, x\rangle_1 : \; \|x\|_1 \le 1, \; \|u\|_2 \le 1 \right\} \\ &\le \max_u \left\{ \sum_{j=1}^{m} \|a_j\|_1^* \cdot |u^{(j)}| : \; \sum_{j=1}^{m} \|a_j\|_1^* \cdot (u^{(j)})^2 \le 1 \right\} \\ &= D^{1/2} \equiv \left[\sum_{j=1}^{m} \|a_j\|_1^* \right]^{1/2}. \end{aligned}$$

С другой стороны, $D_2 = \frac{1}{2} D$ и $\sigma_2 = 1$. Таким образом, с помощью теоремы 5.1.2 получаем следующую оценку сложности:

$$\tfrac{2}{\epsilon} \cdot \sqrt{2D_1} \cdot \sum_{j=1}^{m} \|a_j\|_1^*.$$

5.1.4 Вычислительные аспекты

Вычислительная сложность

Обсудим вычислительную сложность метода (1.2.18), применяемого для минимизации функции $\bar{f}_\mu(x)$. Основной объем вычислений приходится на шаги 1-3 этого метода.

Шаг 1. Вызов оракула. Для этого необходимо решить следующую задачу максимизации:

$$\max_u \{ \langle Ax, u\rangle_2 - \hat{\phi}(u) - \mu d_2(u) : \; u \in Q_2 \}.$$

Из происхождения этой задачи ясно, что при $\mu = 0$ она легко решается. Поэтому можно ожидать, что при правильно выбранной прокс-функции ее решение будет не слишком сложным и при $\mu > 0$. В п. 5.1.3 мы привели три примера для которых это так.

Шаг 3. Вычисление точки z_k. Для этого нужно решить задачу

$$\min_x \{ d_1(x) + \langle s, x\rangle_1 : \; x \in Q_1 \}$$

при некотором фиксированном $s \in E_1^*$. Если множество Q_1 и прокс-функция $d_1(x)$ достаточно простые, это вычисление может быть сделано в явном виде (см. п. 5.1.3). Для некоторых множеств приходится решать одно вспомогательное уравнение с одной переменной. Такая же вспомогательная задача возникает в разных методах оптимизации (см. например, п. 2.1).

Шаг 2. Вычисление точки $T_Q(x)$. Сложность этой операции также зависит от сложности множества Q_1 и нормы $\| \cdot \|_1$. Случай евклидовой нормы хорошо понятен. Поэтому ниже мы рассматриваем нахождение этой точки для произвольной нормы.

Приведем следующее полезное утверждение.

Лемма 5.1.4 *Для любых $g \in E^*$ и $h \in E$ имеем*

$$\langle g, h \rangle_1 + \tfrac{1}{2} L \|h\|^2 = \max_s \left\{ \langle s, h \rangle_1 - \tfrac{1}{2L} (\|s - g\|^*)^2 : \ s \in E^* \right\}.$$

Доказательство.
Действительно,

$$\begin{aligned} \langle g, h \rangle_1 + \tfrac{1}{2} L \|h\|^2 &= \max_{r \geq 0} \{ \langle g, h \rangle_1 + r \|h\|_1 - \tfrac{1}{2L} r^2 \} \\ &= \max_{r,s} \{ \langle g, h \rangle_1 + \langle rs, h \rangle_1 - \tfrac{1}{2L} r^2 : \ r \geq 0, \ \|s\|^* = 1 \} \\ &= \max_s \{ \langle g + s, h \rangle_1 - \tfrac{1}{2L} (\|s\|^*)^2 : \ s \in E^* \} \\ &= \max_s \left\{ \langle s, h \rangle_1 - \tfrac{1}{2L} (\|s - g\|^*)^2 : \ s \in E^* \right\}. \qquad \Box \end{aligned}$$

Оценим сложность вычисления точки $T_Q(x)$ в ситуации рассмотренной в п.5.1.3. Нам нужно решить задачу

$$\text{найти } \psi^* = \min_x \{ \langle \bar{g}, x - \bar{x} \rangle + \tfrac{1}{2} L \|x - \bar{x}\|^2 : \ x \in \Delta_n \}, \tag{5.1.33}$$

где $\|x\| = \sum\limits_{i=1}^{n} |x^{(i)}|$ и $\bar{x} \in \Delta_n$. Таким образом, не ограничивая общности можно считать, что

$$\min_{1 \leq i \leq n} \bar{g}^{(i)} = 0. \tag{5.1.34}$$

В силу леммы5.1.4 эта задача может быть переписана так:

$$\begin{aligned} \psi^* &= \min_{x \in \Delta_n} \max_s \left\{ \langle s, x - \bar{x} \rangle - \tfrac{1}{2L} (\|s - \bar{g}\|^*)^2 \right\} \\ &= \min_{x \geq 0} \max_{s,\lambda} \left\{ \langle s, x - \bar{x} \rangle - \tfrac{1}{2L} (\|s - \bar{g}\|^*)^2 + \lambda(1 - \langle e_n, x \rangle) \right\} \\ &= \max_{s,\lambda} \left\{ -\langle s, \bar{x} \rangle - \tfrac{1}{2L} (\|s - \bar{g}\|^*)^2 + \lambda) : \ s \geq \lambda e_n \right\}. \end{aligned}$$

Заметим, что в нашем случае $\|s\|^* = \max\limits_{1 \leq i \leq n} |s^{(i)}|$. Таким образом,

$$-\psi^* = \min_{s,\lambda,\tau} \left\{ \langle s, \bar{x} \rangle + \tfrac{\tau^2}{2L} - \lambda : \ s^{(i)} \geq \lambda, \ |s^{(i)} - \bar{g}^{(i)}| \leq \tau, \ i = 1, \ldots, n \right\}. \tag{5.1.35}$$

В последней задаче легко находятся оптимальные значения для $s^{(i)}$:

$$s_*^{(i)} = \max\{\lambda, \bar{g}^{(i)} - \tau\}, \quad i = 1, \dots, n.$$

Более того, допустимое множество этой задачи непусто тогда и только тогда, когда

$$\lambda \le \bar{g}^{(i)} + \tau, \quad i = 1, \dots, n.$$

В силу (5.1.34) это означает, что $\lambda \le \tau$. Таким образом,

$$\begin{aligned} -\psi^* &= \min_{\tau \ge \lambda} \left\{ \sum_{i=1}^{n} \bar{x}^{(i)} \max\{\lambda, \bar{g}^{(i)} - \tau\} + \tfrac{\tau^2}{2L} - \lambda \right\} \\ &= \min_{\tau \ge \lambda} \left\{ \sum_{i=1}^{n} \bar{x}^{(i)} (\bar{g}^{(i)} - \tau - \lambda)_+ + \tfrac{\tau^2}{2L} \right\}, \end{aligned}$$

где $(a)_+ = \max\{a, 0\}$. Поскольку целевая функция последней задачи убывает по λ, мы заключаем, что $\lambda^* = \tau$.

Наконец, мы приходим к следующему представлению:

$$-\psi^* \;=\; \min_{\tau \ge 0} \left\{ \sum_{i=1}^{n} \bar{x}^{(i)} (\bar{g}^{(i)} - 2\tau)_+ + \tfrac{\tau^2}{2L} \right\}.$$

Ясно, что его решение может быть найдено с помощью упорядочивания компонент $\bar{g}^{(i)}$ и вычисления производных целевой функции в точках

$$\tau_i = \tfrac{1}{2} \bar{g}^{(i)}, \quad i = 1, \dots, n.$$

Общая трудоемкость этого вычисления имеет порядок $O(n \ln n)$ арифметических операций. При этом прямое решение x^* задачи (5.1.33) может быть легко восстановлено. .

Вычислительная устойчивость

В нашем подходе необходимо сглаживать негладкие функции. В соответствии с правилами выбора (5.1.17) значение параметра сглаживания μ должно быть порядка ϵ. Это может привести к численной неустойчивости при вычислении значений функции $\bar{f}_\mu(x)$ и ее градиента. Среди примеров п. 5.1.3 только гладкий вариант целевой функции (5.1.26) относительно безопасен. Все остальные требуют аккуратного обращения.

Довольно часто нам нужна устойчивая техника вычисления значений и производных функции

$$\eta(u) = \mu \ln \left(\sum_{j=1}^{m} e^{u^{(j)}/\mu} \right) \tag{5.1.36}$$

с очень маленькими значениями параметра μ. Это может быть сделано следующим образом. Обозначим

$$\bar{u} = \max_{1 \le j \le m} u^{(j)}, \quad v^{(j)} = u^{(j)} - \bar{u}, \; j = 1, \dots, m.$$

Тогда

$$\eta(u) = \bar{u} + \eta(v)$$

Заметим, что все компоненты вектора v неотрицательны, и одна из них ноль. Таким образом, значение $\eta(v)$ может быть вычислена с маленькой ошибкой. Тот же подход может быть использован и для вычисления градиента этой функции поскольку $\nabla\eta(u) = \nabla\eta(v)$.

Вычислительные эксперименты, представленные в разделе 5.1.5, подтверждают, что техника сглаживания, реализованная правильным образом, работает устойчиво даже для высоких точностей.

Модифицированный метод

Как мы видели, на каждой итерации метода (1.2.18) необходимо решить две вспомогательные задачи разных типов. Оказывается, что задача вычисления точки y_k часто бывает сложнее, чем задача вычисления z_k. Покажем как можно модифицировать метод (1.2.18), чтобы обе задачи формулировались только в терминах прокс-функции $d(x)$.

Для простоты предположим, что $d(x)$ дифференцируема. Обозначим через

$$\rho(z, x) = d(x) - d(z) - \langle \nabla d(z), x - z \rangle, \quad z, x \in Q,$$

брегмановское расстояние между z и x. Ясно, что

$$\rho(z, x) \geq \tfrac{1}{2}\|x - z\|^2.$$

Определим отображение

$$\mathcal{B}_Q(z, g) = \arg\min_x \{\langle g, x - z \rangle + \rho(z, x) : \ x \in Q\}.$$

В последующем мы пользуемся обозначениями п. 1.2.2.

Лемма 5.1.5 *Пусть последовательность* $\{\alpha_k\}_{k=0}^{\infty}$ *удовлетворяет условию (1.2.13). Предположим, что соотношение* $(\mathcal{R}_k)$ *выполнено при некотором* $k \geq 0$. *Выберем* $\gamma_k = \frac{1}{L}\alpha_{k+1}$. *Положим*

$$\begin{aligned} x_{k+1} &= \tau_k z_k + (1 - \tau_k) y_k, \\ \hat{x}_{k+1} &= \mathcal{B}_Q(z_k, \gamma_k \nabla f(x_{k+1})), \\ y_{k+1} &= \tau_k \hat{x}_{k+1} + (1 - \tau_k) y_k. \end{aligned} \tag{5.1.37}$$

Тогда соотношение $(\mathcal{R}_{k+1})$ *также выполнено.*

Доказательство.

Обозначим $l_k(x) \equiv \beta_k + \langle l_k, x - z_k \rangle = \sum\limits_{i=0}^{k} \alpha_i [f(x_i) + \langle \nabla f(x_i), x - x_i \rangle]$. Тогда

$$\langle L d'(z_k) + l_k, x - z_k \rangle \geq 0 \quad \forall x \in Q.$$

Поэтому, поскольку $\psi_k = Ld(z_k) + \beta_k$, в силу неравенства (1.2.15) имеем

$$\begin{array}{rl} & Ld(x) + l_k(x) + \alpha_{k+1}\langle \nabla f(x_{k+1}), x - x_{k+1}\rangle \\ = & L\rho(z_k, x) + L(d(z_k) + \langle d'(z_k), x - z_k\rangle) \\ & +\beta_k + \langle l_k, x - z_k\rangle + \alpha_{k+1}\langle \nabla f(x_{k+1}), x - x_{k+1}\rangle \\ \geq & L\rho(z_k, x) + \psi_k + \alpha_{k+1}\langle \nabla f(x_{k+1}), x - x_{k+1}\rangle \\ \geq & L\rho(z_k, x) + A_{k+1} f(x_{k+1}) + \alpha_{k+1}\langle \nabla f(x_{k+1}), x - z_k\rangle. \end{array}$$

Таким образом, пользуясь неравенством (1.2.13), получаем:

$$\begin{array}{rcl} \psi_{k+1} & \geq & \min\limits_x \{L\rho(z_k, x) + A_{k+1} f(x_{k+1}) + \alpha_{k+1}\langle \nabla f(x_{k+1}), x - z_k\rangle : \ x \in Q\} \\ & = & L\rho(z_k, \hat{x}_{k+1}) + A_{k+1} f(x_{k+1}) + \alpha_{k+1}\langle \nabla f(x_{k+1}), \hat{x}_{k+1} - z_k\rangle \\ & \geq & \frac{1}{2}L\|\hat{x}_{k+1} - z_k\|^2 + A_{k+1} f(x_{k+1}) + \alpha_{k+1}\langle \nabla f(x_{k+1}), \hat{x}_{k+1} - z_k\rangle \\ & \geq & A_{k+1}\left(\frac{1}{2}L\tau_k^2\|\hat{x}_{k+1} - z_k\|^2 + f(x_{k+1}) + \tau_k\langle \nabla f(x_{k+1}), \hat{x}_{k+1} - z_k\rangle\right). \end{array}$$

Остается воспользоваться равенством $y_{k+1} - x_{k+1} = \tau_k(\hat{x}_{k+1} - z_k)$. □

Ясно, что при любом $\alpha_0 \in (0, 1]$ можно взять

$$y_0 = z_0 = \arg\min_x \left\{ \tfrac{L}{\sigma} d(x) + \alpha_0[f(x_0) + \langle \nabla f(x_0), x - x_0\rangle] : \ x \in Q \right\}.$$

В частности, можно воспользоваться последовательностью, предложенной в лемме 1.2.2. В этом случае мы приходим к следующей алгоритмической схеме.

1. Выбираем $y_0 = \arg\min\limits_x \left\{ Ld(x) + \frac{1}{2}[f(x_0) + \langle \nabla f(x_0), x - x_0\rangle] : \ x \in Q \right\}$.

2. Для $k \geq 0$ **повторяем:**

a) Находим $z_k = \arg\min\limits_x \left\{ Ld(x) + \sum\limits_{i=0}^{k} \frac{i+1}{2}[f(x_i) + \langle \nabla f(x_i), x - x_i\rangle] : \ x \in Q \right\}$.

b) Полагаем $\tau_k = \frac{2}{k+3}$ и $x_{k+1} = \tau_k z_k + (1 - \tau_k) y_k$.

c) Находим $\hat{x}_{k+1} = \mathcal{B}_Q(z_k, \frac{1}{L}\tau_k \nabla f(x_{k+1}))$.

d) Полагаем $y_{k+1} = \tau_k \hat{x}_{k+1} + (1 - \tau_k) y_k$.

(5.1.38)

Для этого метода утверждение теоремы 1.2.3 остается верным. В качестве примера приведем вид отображения $\mathcal{B}_Q(z, g)$ для энтропийного расстояния:

$$\mathcal{B}_Q^{(i)}(z, g) = z^{(i)} e^{-g^{(i)}} \cdot \left[\sum_{j=1}^{n} z^{(j)} e^{-g^{(j)}} \right]^{-1}, \quad j = 1, \ldots, n. \tag{5.1.39}$$

Ясно, что это вычисление более привлекательно по сравнению со стратегией рассмотренной в п. 5.1.4.

5.1.5 Вычислительные эксперименты

Мы завершаем этот раздел результатами численных экспериментов для случайно сгенерированных матричных игр

$$\min_{x \in \Delta_n} \max_{u \in \Delta_m} \langle Ax, u \rangle_2.$$

Элементы случайной матрицы A равномерно распределены в интервале $[-1, 1]$.

Перед нами стояли две цели. Во-первых, необходимо было проверить устойчивость техники сглаживания к реализации на компьютерах с плавающей арифметикой. Во-вторых, интересно было проверить точность теоретических предсказаний, а именно, что сложность нахождения ϵ-решения действительно растет пропорционально произведению $\frac{1}{\epsilon}$ на логарифмические множители, зависящие от n и m.

Для наших экспериментов мы реализовали метод (1.2.18) *в точности* как он и был описан. Параметры алгоритма были выбраны по рекомендациям (5.1.17). Заметим, что для маленьких ϵ эти значения становятся достаточно большими. Например, если взять

$$\|A\|_{1,2} = 1, \quad n = 10^4, \quad m = 10^3, \quad \epsilon = 10^{-3},$$

то параметры метода (1.2.18) окажутся такими:

$$\mu = 0.72 \cdot 10^{-4}, \quad L_\mu = 23858, 54, \quad N = 31906.$$

Таким образом, наши сомнения относительно численной устойчивости техники сглаживания кажутся вполне обоснованными.

Ниже мы приводим три таблицы с результатами счета. Они соответствуют различным уровням точности ϵ, а именно 10^{-2}, 10^{-3} и 10^{-4}. Для последнего значения ϵ мы опустили задачи максимального размера поскольку общая картина уже становится ясной. На каждом шаге метода мы вычисляем два матрично-векторных произведения с матрицей A. Для проверки критерия остановки вычисляются значения прямой и двойственной функции в точках $\hat{x}$ и $\hat{u}$ и проверяется условие

$$f(\hat{x}) - \phi(\hat{u}) \leq \epsilon.$$

Эта проверка осуществляется один раз за каждые сто (или тысячу) итераций. Поэтому она не слишком увеличивает общее время вычислений. Для экспериментов мы использовали персональный компьютер с частотой 2.6GHz. В приводимых таблицах мы даем общее число итераций, общее время в секундах, и процентное отношение реального времени вычислений к теоретичекому предсказанию.

Мы видим что во всех трех таблицах сложность задач действительно растет пропорционально $\frac{1}{\epsilon}$. Более того, теоретические предсказания сложности очень точны. Затрачиваемое время кажется достаточно большим. Но это, естественно, в первую очередь связано с максимальной заполненностью матрицы A. В реальных задачах уровень заполненности гораздо ниже.

Результаты для $\epsilon = 0.01$. **Таблица 1**

$m \backslash n$	100	300	1000	3000	10000
100	808 $0'', 44\%$	1011 $0'', 49\%$	1112 $3'', 49\%$	1314 $12'', 54\%$	1415 $44'', 54\%$
300	910 $0'', 44\%$	1112 $2'', 49\%$	1415 $10'', 56\%$	1617 $35'', 60\%$	1819 $135'', 63\%$
1000	1112 $2'', 49\%$	1213 $8'', 48\%$	1415 $32'', 51\%$	1718 $115'', 58\%$	2020 $451'', 63\%$

Результаты для $\epsilon = 0.001$. **Таблица 2**

$m \backslash n$	100	300	1000	3000	10000
100	6970 $2'', 38\%$	8586 $8'', 42\%$	9394 $29'', 42\%$	10000 $91'', 41\%$	10908 $349'', 42\%$
300	7778 $8'', 38\%$	10101 $27'', 44\%$	12424 $97'', 49\%$	14242 $313'', 53\%$	15656 $1162'', 54\%$
1000	8788 $30'', 39\%$	11010 $105'', 44\%$	13030 $339'', 47\%$	15757 $1083'', 53\%$	18282 $4085'', 57\%$

Результаты, представленные в Таблицах 1 и 2, обнадеживают. Этот уровень точности для субградиентных методов с оценками $O(\frac{1}{\epsilon^2})$ уже достаточно высок. Конечно же эти задачи могут решаться и методами отсечений с линейной скоростью сходимости. Однако для таких методов обычно невязка уменьшается вдвое за число итераций, пропорциональных размерности пространства n. В этом отношении результаты последней колонки в Таблице 2 выглядят очень хорошо: мы получаем три верных знака после n или $2n$ итераций. В то же время, сложность одной итерации в методах отсечений составляет по крайней мере $\frac{1}{3}n^3$ операций. Таким образом, даже если она реализуется в меньшем размере m, трудоемкость вычислений для достижения результата, показанного в правом нижнем углу Таблицы 3, должна была бы составить $180 \cdot 3 \cdot 2 = 1080$ итераций (поскольку здесь $n = 10m$).

Результаты для $\epsilon = 0.0001$. Таблица 3

$m \backslash n$	100	300	1000	3000
100	67068 $25'', 36\%$	72073 $80'', 35\%$	74075 $287'', 33\%$	80081 $945'', 33\%$
300	85086 $89'', 42\%$	92093 $243'', 40\%$	101102 $914'', 40\%$	112113 $3302'', 41\%$
1000	97098 $331'', 43\%$	100101 $760'', 40\%$	116117 $2936'', 42\%$	139140 $11028'', 47\%$

Уровень точности, представленный в Таблице 3, просто недостижим для субградиентных методов. Он достаточно высок даже для методов отсечений. Еще раз напомним, что вычислительные затраты нашего метода, необходимые для получения результата в ячейке (3,3) эквиваленты всего $116 \cdot 3 \cdot 2 = 696$ итераций метода отсечений в размерности $n = 1000$. Этот ресурс вряд ли достаточен для получения решения с четырьмя верными знаками.

5.2 Условие обратного зазора в негладкой выпуклой минимизации

5.2.1 Введение

Мотивировка

В этом разделе мы приводим дальнейшее развитие результатов параграфа 5.1, в котором было показано, что некоторые задачи негладкой минимизации с известной структурой могут быть решены градиентными методами за $O(\frac{1}{\epsilon})$ итераций, где ϵ – требуемая точность решения. Эта оценка гораздо лучше нижней теоретической оценки сложности $O(\frac{1}{\epsilon^2})$, которая верна для черноящичных задач (см. монографию [79]). Это ускорение стало возможным благодаря устранению концепции черного ящика. Наоборот, теперь мы предполагаем, что задача имеет явную минимаксную структуру. Численный метод, предложенный в параграфе 5.1, имеет один недостаток. Число итераций в нем должно быть выбрано заранее в соответствием с теоретическим анализом для наихудшего случая.

В этом параграфе мы рассмотрим несколько прямо-двойственных методов, работающих на том же классе задач, что и в параграфе 5.1. Однако эти схемы является более гибкими. Они выводятся с помощью специальной симметричной прямо-двойственной техники, основанной на *условии обратного зазора.*

Содержание

В п. 5.2.2 описывается *модель* нашей оптимизационной задачи. В п. 5.2.3 вводится условие обратного зазора. В следующих двух пунктах анализируются два рекуррентных способа выполнения этого условия. В п. 5.2.6 выводится оценка скорости сходимости порядка $O(\frac{1}{k})$, где $k =$ это счетчик итераций. Этот результат применим к негладким функциям, описываемым нашей моделью. Однако, при дополнительных предположениях (сильная выпуклость целевой функции), скорость возрастает до $O(\frac{1}{k^2})$. Это ускорение обосновывается в последнем п. 5.2.7. Заметим, что оба результата лучше соответствующих черноящичных нижних оценок сложности на целый порядок.

5.2.2 Описание структуры задач

Мы рассматриваем задачи минимизации с той же структурой, что и в параграфе 5.1:

$$\text{Find } f^* = \min_{x \in Q_1} f(x), \tag{5.2.1}$$

где выпуклое замкнутое множество Q_1 из конечномерного векторного пространства E_1 является ограниченным, а выпуклая функция $f(x)$ непрерывна на Q_1. Эта функция может быть недифференцируемой. Пусть структура целевой функции задается следующей *моделью*:

$$f(x) = \hat{f}(x) + \max_{u \in Q_2} \{\langle Ax, u \rangle_2 - \hat{\phi}(u)\}, \tag{5.2.2}$$

где выпуклая функция $\hat{f}(x)$ непрерывна на Q_1, множество Q_2 является выпуклым замкнутым и ограниченным в конечномерном вещественном пространстве E_2, выпуклая функция $\hat{\phi}(u)$ непрерывна на Q_2 и линейный оператор A отображает E_1 в E_2^*. В этом случае задача (5.2.1) может быть записана в *сопряженной форме*:

$$\begin{gathered} \max_{u \in Q_2} \phi(u), \\ \phi(u) = -\hat{\phi}(u) + \min_{x \in Q_1} \{\langle Ax, u \rangle_2 + \hat{f}(x)\}. \end{gathered} \tag{5.2.3}$$

Мы предполагаем, что это представление абсолютно аналогично представлению (5.2.1) в следующем смысле. Методы, описываемые в этом разделе, реализуемы только если задачи оптимизации, присутствующие в определениях функций $f(x)$ и $\phi(u)$, могут быть решены в явном виде. Таким образом, предполагается, что структура объектов $\hat{f}$, $\hat{\phi}$, Q_1 и Q_2 достаточно проста. Также предполагается, что градиенты функций $\hat{f}$ и $\hat{\phi}$ липшицевы с константами $L_1(\hat{f})$ и $L_2(\hat{\phi})$ соответственно. Опишем наши возможности по сглаживанию целевых функций в задачах (5.2.1) и (5.2.3).

Зафиксируем *прокс-функцию* $d_2(u)$ для множества Q_2. Функция $d_2(u)$ непрерывна и сильно выпукла на Q_2 с параметром $\sigma_2 = 1$. Обозначим через

$$u_0 = \arg \min_{u \in Q_2} d_2(u)$$

ее *прокс-центр*. Будем считать, что $d_2(u_0) = 0$. Таким образом, в силу (1.1.30) для любого $u \in Q_2$ имеем

$$d_2(u) \geq \tfrac{1}{2} \|u - u_0\|_2^2. \tag{5.2.4}$$

Для параметра сглаживания μ_2 введем функцию

$$f_{\mu_2}(x) = \hat{f}(x) + \max_{u \in Q_2} \{\langle Ax, u\rangle_2 - \hat{\phi}(u) - \mu_2 d_2(u)\}. \tag{5.2.5}$$

Обозначим через $u_{\mu_2}(x)$ оптимальное решение этой задачи (оно единственно). Поэтому функция f_{μ_2} дифференцируема и ее градиент вычисляется по формуле

$$\nabla f_{\mu_2}(x) = \nabla \hat{f}(x) + A^* u_{\mu_2}(x). \tag{5.2.6}$$

Более того, этот градиент Липшицев с константой

$$L_1(f_{\mu_2}) = L_1(\hat{f}) + \tfrac{1}{\mu_2}\|A\|_{1,2}^2. \tag{5.2.7}$$

(см., например, теорему 5.1.1).

Аналогично, введем прокс-функцию $d_1(x)$ для множества Q_1, с параметром $\sigma_1 = 1$ и прокс-центром x_0 с $d_1(x_0) = 0$. В силу (1.1.30), для любого $x \in Q_1$ имеем

$$d_1(x) \geq \tfrac{1}{2}\|x - x_0\|_1^2. \tag{5.2.8}$$

Для параметра сглаживания μ_1 введем функцию

$$\phi_{\mu_1}(u) = -\hat{\phi}(u) + \min_{x \in Q_1} \{\langle Ax, u\rangle_2 + \hat{f}(x) + \mu_1 d_1(x)\}. \tag{5.2.9}$$

Поскольку второй член в этом определении есть минимум линейных функций, функция $\phi_{\mu_1}(u)$ вогнута. Обозначим через $x_{\mu_1}(u)$ единственное оптимальное решение этой задачи. Тогда градиент этой функции, вычисляемый по формуле

$$\nabla \phi_{\mu_1}(u) = -\nabla \hat{\phi}(u) + A x_{\mu_1}(u), \tag{5.2.10}$$

удовлетворяет условию Липшица с константой

$$L_2(\phi_{\mu_1}) = L_2(\hat{\phi}) + \tfrac{1}{\sigma_1 \mu_1}\|A\|_{1,2}^2. \tag{5.2.11}$$

5.2.3 Условие обратного зазора

Заметим, что при любых $x \in Q_1$ и $u \in Q_2$ выполняется неравенство

$$\phi(u) \leq f(x). \tag{5.2.12}$$

Более того, наши предположения гарантируют в задачах (5.2.1), (5.2.3) нулевой зазор двойственности. Однако, $f_{\mu_2}(x) \leq f(x)$ и $\phi(u) \leq \phi_{\mu_1}(u)$. Такая ситуация открывает принципиальную возможность обеспечить выполнение условия *обратного зазора*:

$$f_{\mu_2}(\bar{x}) \ \leq \ \phi_{\mu_1}(\bar{u}) \tag{5.2.13}$$

для некоторых точек $\bar{x} \in Q_1$ и $\bar{u} \in Q_2$. Ясно, что из условия (5.2.13) нетрудно получить оценку качества соответствующей прямо-двойственной пары $(\bar{x}, \bar{u})$.

Лемма 5.2.1 *Пусть точки $\bar{x} \in Q_1$ и $\bar{u} \in Q_2$ удовлетворяют условию (5.2.13). Тогда справедлива оценка*

$$\begin{array}{rcl} 0 & \leq & \max\{f(\bar{x}) - f^*, f^* - \phi(\bar{u})\} \\ & \leq & f(\bar{x}) - \phi(\bar{u}) \ \leq \ \mu_1 D_1 + \mu_2 D_2, \end{array} \tag{5.2.14}$$

где $D_1 = \max\limits_{x \in Q_1} d_1(x)$, *и* $D_2 = \max\limits_{u \in Q_2} d_2(u)$.

Доказательство.

Действительно, для любых точек $\bar{x} \in Q_1$, $\bar{u} \in Q_2$ имеем

$$f(\bar{x}) - \mu_2 D_2 \leq f_{\mu_2}(\bar{x}) \leq \phi_{\mu_1}(\bar{u}) \leq \phi(\bar{u}) + \mu_1 D_1.$$

Осталось воспользоваться неравенством (5.2.12). □

Нашей целью является построение процесса рекуррентного пересчета пары $(\bar{x}, \bar{u})$, которая удовлетворят условию (5.2.13) при μ_1 и μ_2 стремящихся к нулю. Это можно сделать двумя различными способами, соответствуютщих двум разным типам вспомогательных задач, которые мы готовы решать на каждой итерации. Прежде чем приступать к описанию наших возможностей, докажем одно полезное неравенство.

Лемма 5.2.2 *При любых x и $\bar{y}$ из Q_1 выполнено неравенство*

$$f_{\mu_2}(\bar{y}) + \langle \nabla f_{\mu_2}(\bar{y}), x - \bar{y} \rangle_1 \leq \hat{f}(x) + \langle Ax, u_{\mu_2}(\bar{y}) \rangle_2 - \hat{\phi}(u_{\mu_2}(\bar{y})). \tag{5.2.15}$$

Доказательство.

Выберем произвольные точки $x, \bar{y} \in Q_1$. Обозначим $\bar{u} = u_{\mu_2}(\bar{y})$. Тогда

$$\begin{array}{rcl} & & f_{\mu_2}(\bar{y}) + \langle \nabla f_{\mu_2}(\bar{y}), x - \bar{y} \rangle_1 \\ (\text{в силу (5.2.5), (5.2.6)}) & = & \hat{f}(\bar{y}) + \langle A\bar{y}, \bar{u} \rangle_2 - \hat{\phi}(\bar{u}) - \mu_2 d_2(\bar{u}) + \langle \nabla \hat{f}(\bar{y}) + A^* \bar{u}, x - \bar{y} \rangle_1 \\ (\text{в силу выпуклости } \hat{f}) & \leq & \hat{f}(x) + \langle Ax, \bar{u} \rangle_2 - \hat{\phi}(\bar{u}). \end{array}$$

□

5.2.4 Метод с градиентным отображением

Обоснуем сначала возможность обеспечить условие обратного зазора (5.2.13) в стартовой прямо-двойственной точке. Для $x \in Q_1$ определим *прямое градиентное отображение*:

$$T_{\mu_2}(x) = \arg\min_{y \in Q_1} \left\{ \langle \nabla f_{\mu_2}(x), y - x \rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2}) \| y - x \|_1^2 \right\}. \tag{5.2.16}$$

Лемма 5.2.3 *Выберем произвольное $\mu_2 > 0$. Для прокс-центра x_0 определим*

$$\bar{x} = T_{\mu_2}(x_0), \quad \bar{u} = u_{\mu_2}(x_0). \tag{5.2.17}$$

Тогда условие обратного зазора (5.2.13) выполнено при любом

$$\mu_1 \geq L_1(f_{\mu_2}). \tag{5.2.18}$$

Доказательство.

Обозначим $\bar{x} = T_{\mu_2}(x_0)$, $L_1 = L_1(f_{\mu_2})$, и $\bar{u} = u_{\mu_2}(x_0)$. Поскольку градиент ∇f_{μ_2} липшицев, в силу неравенства (1.1.15) имеем

$$\begin{array}{rcl} f_{\mu_2}(\bar{x}) & \leq & f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), \bar{x} - x_0 \rangle_1 + \frac{1}{2} L_1 \| \bar{x} - x_0 \|_1^2 \\ \text{(из (5.2.16))} & = & \min\limits_{x \in Q_1} \left\{ f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), x - x_0 \rangle_1 + \frac{1}{2} L_1 \| x - x_0 \|_1^2 \right\} \\ \text{(из (5.2.15), (5.2.18))} & \leq & \min\limits_{x \in Q_1} \left\{ \hat{f}(x) + \langle Ax, \bar{u} \rangle_2 - \hat{\phi}(\bar{u}) + \frac{1}{2} \mu_1 \| x - x_0 \|_1^2 \right\} \\ \text{(из (5.2.8))} & \leq & -\hat{\phi}(\bar{u}) + \min\limits_{x \in Q_1} \left\{ \hat{f}(x) + \langle Ax, \bar{u} \rangle_2 + \mu_1 d_1(x) \right\} \;=\; \phi_{\mu_1}(\bar{u}). \Box \end{array}$$

Таким образом, условие (5.2.13) может быть выполнено для некоторой стартовой прямо-двойственной пары. Покажем, как можно так пересчитывать точки $\bar{x}$ и $\bar{u}$, чтобы условие (5.2.13) продолжало выполняться при меньших значениях μ_1 и μ_2. Заметим, что в силу симметричности ситуации на первом шаге мы можем попробовать сначала уменьшить μ_1, оставляя μ_2 без изменения. Затем на втором шаге мы уменьшим μ_2, не трогая μ_1, и так далее. Основным достоинством такой переключающейся стратегии состоит в том что обосновывать нужно только первый шаг. Второй и все последующие шаги будут обоснованы автоматически.

Теорема 5.2.1 *Пусть точки* $\bar{x} \in Q_1$ *и* $\bar{u} \in Q_2$ *удовлетворяют условию обратного зазора (5.2.13) при некоторых положительных* μ_1 *и* μ_2. *Зафиксируем* $\tau \in (0,1)$ *и выберем* $\mu_1^+ = (1-\tau)\mu_1$,

$$\begin{array}{rcl} \hat{x} & = & (1-\tau)\bar{x} + \tau x_{\mu_1}(\bar{u}), \\ \bar{u}_+ & = & (1-\tau)\bar{u} + \tau u_{\mu_2}(\hat{x}), \\ \bar{x}_+ & = & T_{\mu_2}(\hat{x}). \end{array} \tag{5.2.19}$$

Тогда пара $(\bar{x}_+, \bar{u}_+)$ *удовлетворяет условию (5.2.13) с параметрами сглаживания* μ_1^+ *и* μ_2 *при всех не слишком больших* τ:

$$\frac{\tau^2}{1-\tau} \leq \frac{\mu_1}{L_1(f_{\mu_2})}. \tag{5.2.20}$$

Доказательство.

Обозначим $\hat{u} = u_{\mu_2}(\hat{x})$ и $x_1 = x_{\mu_1}(\bar{u})$. Поскольку функция $\hat{\phi}$ выпукла, в силу второй строки в правилах (5.2.19) имеем $\hat{\phi}(\bar{u}_+) \le (1-\tau)\hat{\phi}(\bar{u}) + \tau\hat{\phi}(\hat{u})$. Таким образом,

$$\begin{aligned}
\phi_{\mu_1^+}(\bar{u}_+) &= \min_{x \in Q_1} \left\{ (1-\tau)\mu_1 d_1(x) + \langle Ax, (1-\tau)\bar{u} + \tau\hat{u}\rangle_2 + \hat{f}(x) \right\} - \hat{\phi}(\bar{u}_+) \\
&\ge \min_{x \in Q_1} \left\{ (1-\tau) \left[\mu_1 d_1(x) + \langle Ax, \bar{u}\rangle_2 + \hat{f}(x) - \hat{\phi}(\bar{u}) \right]_1 \right. \\
&\quad \left. + \tau[\hat{f}(x) + \langle Ax, \hat{u}\rangle_2 - \hat{\phi}(\hat{u})]_2 \right\}.
\end{aligned}$$

Заметим, что в силу условия (5.2.13) и первой строки в правилах (5.2.19) выполнено неравенство

$$\phi_{\mu_1}(\bar{u}) \ge f_{\mu_2}(\bar{x}) \ge f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), \bar{x} - \hat{x}\rangle_1 = f_{\mu_2}(\hat{x}) + \tau\langle \nabla f_{\mu_2}(\hat{x}), \bar{x} - x_1\rangle_1.$$

Следовательно, в силу свойства (1.1.30) и определения (5.2.9) выражение в первых скобках оценивается следующим образом:

$$\begin{aligned}
[\,\cdot\,]_1 &\ge \phi_{\mu_1}(\bar{u}) + \tfrac{1}{2}\mu_1 \|x - x_1\|_1^2 \\
(\text{из } (5.2.13)) &\ge f_{\mu_2}(\bar{x}) + \tfrac{1}{2}\mu_1 \|x - x_1\|_1^2 \\
(f \text{ выпукла}) &\ge f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), \bar{x} - \hat{x}\rangle_1 + \tfrac{1}{2}\mu_1 \|x - x_1\|_1^2 \\
(\text{строка } 1,\ (5.2.19)) &= f_{\mu_2}(\hat{x}) + \tau\langle \nabla f_{\mu_2}(\hat{x}), \bar{x} - x_1\rangle_1 + \tfrac{1}{2}\mu_1 \|x - x_1\|_1^2.
\end{aligned}$$

В силу (5.2.15), для выражения во вторых скобках имеем

$$\begin{aligned}
[\,\cdot\,]_2 &\ge f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), x - \hat{x}\rangle_1 \\
(\text{строка } 1,\ (5.2.19)) &= f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), x - x_1 + (1-\tau)(x_1 - \bar{x})\rangle_1.
\end{aligned}$$

Таким образом, собирая все вместе, можно завершить доказательство так:

$$\begin{aligned}
\phi_{\mu_1^+}(\bar{u}_+) &\ge \min_{x \in Q_1} \left\{ f_{\mu_2}(\hat{x}) + \tau\langle \nabla f_{\mu_2}(\hat{x}), x - x_1\rangle_1 + \tfrac{1}{2}(1-\tau)\mu_1 \|x - x_1\|_1^2 \right\} \\
(\text{из } (5.2.20)) &\ge \min_{x \in Q_1} \left\{ f_{\mu_2}(\hat{x}) + \tau\langle \nabla f_{\mu_2}(\hat{x}), x - x_1\rangle_1 + \tfrac{1}{2}\tau^2 L_1(f_{\mu_2}) \|x - x_1\|_1^2 \right\} \\
(y = \bar{x} + \tau(x - \bar{x})) &= \min_{y \in \bar{x} + \tau(Q_1 - \bar{x})} \left\{ f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), y - \hat{x}\rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2}) \|y - \hat{x}\|_1^2 \right\} \\
(Q_1 \text{ выпукло}) &\ge \min_{y \in Q_1} \left\{ f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), y - \hat{x}\rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2}) \|y - \hat{x}\|_1^2 \right\} \\
(\text{строка } 3,\ (5.2.19)) &= f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), \bar{x}_+ - \hat{x}\rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2}) \|\bar{x}_+ - \hat{x}\|_1^2 \\
(\text{из } (1.1.15)) &\ge f_{\mu_2}(\bar{x}_+).
\end{aligned}$$

□

5.2.5 Метод с брегмановской проекцией

Для простоты предположим, что функция $d_1(x)$ дифференцируема. Тогда для любого $x \in Q_1$ имеем

$$\langle \nabla d_1(x_0), x - x_0 \rangle_1 \geq 0. \tag{5.2.21}$$

Для точек x и z из множества Q_1 обозначим через

$$\rho_1(z, x) = d_1(x) - d_1(z) - \langle \nabla d_1(z), x - z \rangle_1$$

брегмановское расстояние между z и x. Если точка z зафиксирована, то $\rho(z, x)$ сильно выпукла по x. Более того, в силу (1.1.20)

$$\rho_1(z, x) \geq \tfrac{1}{2} \| x - z \|_1^2. \tag{5.2.22}$$

Определим *брегмановскую проекцию* линейного функционала $g \in E_1^*$ на множество Q_1 следующим образом:

$$\mathcal{B}_1(z, g) = \arg \min_{x \in Q_1} \{ \langle g, x - z \rangle_1 + \rho_1(z, x) \}. \tag{5.2.23}$$

Покажем что брегмановская проекция тоже может использоваться для нахождения исходной прямо-двойственной пары, удовлетворяющей условию (5.2.13).

Лемма 5.2.4 *Выберем произвольное $\mu_2 > 0$. Обозначим $\gamma = \frac{1}{L_1(f_{\mu_2})}$ и положим*

$$\bar{x} = \mathcal{B}_1(x_0, \gamma \nabla f_{\mu_2}(x_0)), \quad \bar{u} = u_{\mu_2}(x_0). \tag{5.2.24}$$

Тогда условие обратного зазора выполнено для любого $\mu_1 \geq \gamma^{-1}$.

Доказательство.

Действительно, в силу неравенства (1.1.15) имеем

$$\begin{array}{rcl}
f_{\mu_2}(\bar{x}) & \leq & f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), \bar{x} - x_0 \rangle_1 + \frac{1}{2} L_1(f_{\mu_2}) \| \bar{x} - x_0 \|_1^2 \\[2ex]
& = & f_{\mu_2}(x_0) + \frac{1}{\gamma} \left[\gamma \langle \nabla f_{\mu_2}(x_0), \bar{x} - x_0 \rangle_1 + \frac{1}{2} \| \bar{x} - x_0 \|_1^2 \right] \\[2ex]
\text{(из (5.2.22))} & \leq & f_{\mu_2}(x_0) + \frac{1}{\gamma} \left[\langle \gamma \nabla f_{\mu_2}(x_0), \bar{x} - x_0 \rangle_1 + \rho_1(x_0, \bar{x}) \right] \\[2ex]
\text{(из (5.2.23), (5.2.24))} & = & f_{\mu_2}(x_0) + \frac{1}{\gamma} \min\limits_{x \in Q_1} \left\{ \langle \gamma \nabla f_{\mu_2}(x_0), x - x_0 \rangle_1 + \rho_1(x_0, x) \right\} \\[2ex]
& = & \min\limits_{x \in Q_1} \left\{ f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), x - x_0 \rangle_1 + \frac{1}{\gamma} \rho_1(x_0, x) \right\} \\[2ex]
\text{(из (5.2.21))} & \leq & \min\limits_{x \in Q_1} \left\{ f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), x - x_0 \rangle_1 + \frac{1}{\gamma} d_1(x) \right\} \\[2ex]
\text{(в силу (5.2.15))} & \leq & \min\limits_{x \in Q_1} \left\{ \hat{f}(x) + \langle Ax, u_{\mu_2}(x_0) \rangle - \hat{\phi}(u_{\mu_2}(x_0)) + \frac{1}{\gamma} d_1(x) \right\} \\[2ex]
\text{(из (5.2.9))} & = & \phi_{\gamma^{-1}}(u_{\mu_2}(x_0)) \leq \phi_{\mu_1}(u_{\mu_2}(x_0)). \qquad \square
\end{array}$$

Как и в п. 5.2.4, дадим обоснование только для первого (прямого) шага в переключающейся прямо-двойственной стратегии, поддерживающей выполнение условия (5.2.13) при стремлении параметров сглаживания μ_1, μ_2 к нулю.

Теорема 5.2.2 *Пусть точки $\bar{x} \in Q_1$ и $\bar{u} \in Q_2$ удовлетворяют условию обратного зазора (5.2.13) при некоторых μ_1 и μ_2. Выберем параметр $\tau \in (0,1)$ в соответствии с условием (5.2.20) и положим*

$$\begin{aligned} \hat{x} &= (1-\tau)\bar{x} + \tau x_{\mu_1}(\bar{u}), \\ \bar{u}_+ &= (1-\tau)\bar{u} + \tau u_{\mu_2}(\hat{x}), \\ \tilde{x} &= \mathcal{B}_1(x_{\mu_1}(\bar{u}), \tfrac{\tau}{(1-\tau)\mu_1} \nabla f_{\mu_2}(\hat{x})), \\ \bar{x}_+ &= (1-\tau)\bar{x} + \tau\tilde{x}, \\ \mu_1^+ &= (1-\tau)\mu_1. \end{aligned} \tag{5.2.25}$$

Тогда пара $(\bar{x}_+, \bar{u}_+)$ удовлетворят условию (5.2.13) с параметрами сглаживания μ_1^+ и μ_2.

Доказательство.

Обозначим $\hat{u} = u_{\mu_2}(\hat{x})$ и $x_1 = x_{\mu_1}(\bar{u})$. Ввиду правил (5.2.25), выпуклости функции $\hat{\phi}$ и неравенства (5.2.15) имеем

$$\begin{aligned} &(1-\tau)\mu_1 d_1(x) + \langle Ax, (1-\tau)\bar{u} + \tau\hat{u}\rangle_2 + \hat{f}(x) - \hat{\phi}(\bar{u}_+) \\ &\geq (1-\tau)\left[\mu_1 d_1(x) + \langle Ax, \bar{u}\rangle_2 + \hat{f}(x) - \hat{\phi}(\bar{u})\right] + \tau[\hat{f}(x) + \langle Ax, \hat{u}\rangle_2 - \hat{\phi}(\hat{u})] \\ &\geq (1-\tau)\left[\mu_1 d_1(x) + \langle Ax, \bar{u}\rangle_2 + \hat{f}(x) - \hat{\phi}(\bar{u})\right]_1 + \tau[f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), x - \hat{x}\rangle_1]_2. \end{aligned}$$

Условия оптимальности первого порядка для точки x_1 выглядит следующим образом:

$$\langle \mu_1 \nabla d_1(x_1) + A^*\bar{u} + \nabla\hat{f}(x_1), x - x_1\rangle_1 \geq 0, \quad x \in Q_1. \tag{5.2.26}$$

Поэтому, пользуясь выпуклостью функций $\hat{f}$ и f_{μ_2}, мы можем оценить член $[\,\cdot\,]_1$:

$$\begin{array}{rcl}
[\,\cdot\,]_1 & = & \mu_1\left(\rho(x_1,x)+d_1(x_1)+\langle\nabla d_1(x_1),x-x_1\rangle_1\right)+\langle Ax,\bar{u}\rangle_2+\hat{f}(x)-\hat{\phi}(\bar{u}) \\[2ex]
(\text{из }(5.2.26)) & \geq & \mu_1\rho(x_1,x)+\mu_1 d_1(x_1)+\langle Ax_1,\bar{u}\rangle_2+\hat{f}(x)-\langle\nabla\hat{f}(x_1),x-x_1\rangle_1-\hat{\phi}(\bar{u}) \\[2ex]
 & \geq & \mu_1\rho(x_1,x)+\mu_1 d_1(x_1)+\langle Ax_1,\bar{u}\rangle_2+\hat{f}(x_1)-\hat{\phi}(\bar{u}) \\[2ex]
(\text{из }(5.2.9)) & = & \mu_1\rho(x_1,x)+\phi_{\mu_1}(\bar{u}) \\[2ex]
(\text{из }(5.2.13)) & \geq & \mu_1\rho(x_1,x)+f_{\mu_2}(\bar{x}) \\[2ex]
 & \geq & \mu_1\rho(x_1,x)+f_{\mu_2}(\hat{x})+\langle\nabla f_{\mu_2}(\hat{x}),\bar{x}-\hat{x}\rangle_1.
\end{array}$$

Таким образом, можно продолжить:

$$\begin{array}{rcl}
\phi_{\mu_1^+}(\bar{u}_+) & = & \min\limits_{x\in Q_1}\left\{(1-\tau)\mu_1 d_1(x)+\langle Ax,(1-\tau)\bar{u}+\tau\hat{u}\rangle_2+\hat{f}(x)\right\}-\hat{\phi}(\bar{u}_+) \\[2ex]
 & \geq & \min\limits_{x\in Q_1}\{(1-\tau)\mu_1\rho(x_1,x)+f_{\mu_2}(\hat{x})+\langle\nabla f_{\mu_2}(\hat{x}),(1-\tau)\bar{x}+\tau x-\hat{x}\rangle_1\} \\[2ex]
(\text{строка 1, }(5.2.25)) & = & \min\limits_{x\in Q_1}\{(1-\tau)\mu_1\rho(x_1,x)+f_{\mu_2}(\hat{x})+\tau\langle\nabla f_{\mu_2}(\hat{x}),x-x_1\rangle_1\} \\[2ex]
(\text{строка 3, }(5.2.25)) & = & (1-\tau)\mu_1\rho(x_1,\tilde{x})+f_{\mu_2}(\hat{x})+\tau\langle\nabla f_{\mu_2}(\hat{x}),\tilde{x}-x_1\rangle_1 \\[2ex]
(\text{из }(5.2.22)) & \geq & \frac{1}{2}(1-\tau)\mu_1\|\tilde{x}-x_1\|_1^2+f_{\mu_2}(\hat{x})+\tau\langle\nabla f_{\mu_2}(\hat{x}),\tilde{x}-x_1\rangle_1 \\[2ex]
(\text{из }(5.2.20)) & \geq & \frac{1}{2}\tau^2 L_1(f_{\mu_2})\|\tilde{x}-x_1\|_1^2+f_{\mu_2}(\hat{x})+\tau\langle\nabla f_{\mu_2}(\hat{x}),\tilde{x}-x_1\rangle_1 \\[2ex]
(\text{строка 4, }(5.2.25)) & = & \frac{1}{2}L_1(f_{\mu_2})\|\bar{x}_+-\hat{x}\|_1^2+f_{\mu_2}(\hat{x})+\langle\nabla f_{\mu_2}(\hat{x}),\bar{x}_+-\hat{x}\rangle_1 \\[2ex]
(\text{из }(1.1.15)) & \geq & f_{\mu_2}(\bar{x}_+).
\end{array}$$

□

5.2.6 Анализ скорости сходимости

В п. 5.2.4 и 5.2.5 было показано, что параметры сглаживания μ_1 и μ_2 можно уменьшать с помощью некоторой переключающейся стратегии. Таким образом, для превращение результатов теорем 5.2.1, 5.2.2 в алгоритмическую схему, нам нужно только указать стратегию пересчета этих параметров, совместимую с условием (5.2.20). В этом разделе мы для простоты считаем что $L_1(\hat{f}) = L_2(\hat{\phi}) = 0$.

Будет удобно представлять параметры сглаживания в следующем виде:

$$\mu_1 = \lambda_1\cdot\|A\|_{1,2}\cdot\sqrt{\tfrac{D_2}{D_1}},\quad \mu_2 = \lambda_2\cdot\|A\|_{1,2}\cdot\sqrt{\tfrac{D_1}{D_2}}. \tag{5.2.27}$$

Тогда оценка (5.2.14) для зазора двойственности становится симметричной:

$$f(\bar{x}) - \phi(\bar{u}) \le (\lambda_1 + \lambda_2) \cdot \|A\|_{1,2} \cdot \sqrt{D_1 D_2}. \tag{5.2.28}$$

Поскольку в силу формулы (5.2.7) $L_1(f_{\mu_2}) = \frac{1}{\sigma_2 \mu_2} \|A\|_{1,2}^2$, условие (5.2.20) больше не зависит от конкретной задачи:

$$\tfrac{\tau^2}{1-\tau} \le \mu_1 \mu_2 \cdot \tfrac{1}{\|A\|_{1,2}^2} = \lambda_1 \lambda_2. \tag{5.2.29}$$

Выпишем теперь переключающуюся алгоритмическую схему в явном виде. При этом удобно будет ввести непрерывный счетчик итераций. В таком случае, на четных итерациях мы применяем прямой пересчет (5.2.19) (или (5.2.25)), а на нечетных итерациях – двойственный. Поскольку на четных итерациях не меняется λ_2, а на нечетных итерациях не меняется λ_1, удобно поместить их значения в одну последовательность $\{\alpha_k\}_{k=-1}^{\infty}$ по следующему принципу:

$$\begin{array}{rcl} k = & 2l & : \; \lambda_{1,k} = \alpha_{k-1}, \quad \lambda_{2,k} = \alpha_k, \\ k = & 2l+1 & : \; \lambda_{1,k} = \alpha_k, \quad\;\;\; \lambda_{2,k} = \alpha_{k-1}. \end{array} \tag{5.2.30}$$

Тогда параметры τ_k определяют скорость убывания последовательности $\{\alpha_k\}_{k=-1}^{\infty}$.

Лемма 5.2.5 *При всех* $k \ge 0$ *имеем* $\alpha_{k+1} = (1-\tau_k)\alpha_{k-1}$.

Доказательство.
Действительно, в соответствии со структурой (5.2.30), если $k = 2l$, то

$$\alpha_{k+1} = \lambda_{1,k+1} = (1-\tau_k)\lambda_{1,k} = (1-\tau_k)\alpha_{k-1}.$$

А если $k = 2l+1$, то $\alpha_{k+1} = \lambda_{2,k+1} = (1-\tau_k)\lambda_{2,k} = (1-\tau_k)\alpha_{k-1}$. □

Следствие 5.2.1 *В терминах последовательности* $\{\alpha_k\}_{k=-1}^{\infty}$, *условие (5.2.29) выглядит следующим образом:*

$$(\alpha_{k+1} - \alpha_{k-1})^2 \le \alpha_{k+1}\alpha_k\alpha_{k-1}^2, \quad k \ge 0. \tag{5.2.31}$$

Доказательство.
В силу представления (5.2.30) у нас всегда будет $\lambda_{1,k}\lambda_{2,k} = \alpha_k \alpha_{k-1}$. Поскольку $\tau_k = 1 - \frac{\alpha_{k+1}}{\alpha_{k-1}}$, получаем неравенство (5.2.31). □

Ясно, что условие (5.2.31) выполняется если взять

$$\alpha_k = \tfrac{2}{k+2}, \quad k \ge -1. \tag{5.2.32}$$

Тогда

$$\tau_k = 1 - \tfrac{\alpha_{k+1}}{\alpha_{k-1}} = \tfrac{2}{k+3}, \quad k \ge 0. \tag{5.2.33}$$

Теперь мы можем выписать конкретную алгоритмическую схему. Сделаем это для метода, использующего градиентное отображение (5.2.19). В этой схеме мы пользуемся последовательностями $\{\mu_{1,k}\}_{k=-1}^{\infty}$ и $\{\mu_{2,k}\}_{k=-1}^{\infty}$, сформированными в соответствии с правилами (5.2.27), (5.2.30) и (5.2.32).

1. **Инициализация**
   Выбираем $\bar{x}_0$ и $\bar{u}_0$ по правилу (5.2.17) с $\mu_1 = \mu_{1,0}$ и $\mu_2 = \mu_{2,0}$.

2. **Итерация ($k \geq 0$)**

   a) Полагаем $\tau_k = \frac{2}{k+3}$.

   b) Если k четное, формируем $(\bar{x}_{k+1}, \bar{u}_{k+1})$ из $(\bar{x}_k, \bar{u}_k)$ пользуясь правилом (5.2.19).

   c) Если k нечетное, формируем $(\bar{x}_{k+1}, \bar{u}_{k+1})$ из $(\bar{x}_k, \bar{u}_k)$ пользуясь симметричным двойственным вариантом правила (5.2.19).

(5.2.34)

Теорема 5.2.3 *Пусть последовательности* $\{\bar{x}_k\}_{k=0}^{\infty}$ *и* $\{\bar{u}_k\}_{k=0}^{\infty}$ *формируются методом (5.2.34). Тогда любая пара точек удовлетворяет условию обратного зазора. Таким образом*

$$f(\bar{x}_k) - \phi(\bar{u}_k) \leq \frac{4\|A\|_{1,2}}{k+1}\sqrt{D_1 D_2}. \tag{5.2.35}$$

Доказательство.
В соответствии с нашим выбором параметров,

$$\mu_{1,0}\mu_{2,0} = \lambda_{1,0}\lambda_{2,0} \cdot \frac{\|A\|_{1,2}^2}{\sigma_1\sigma_2} = \frac{2\mu_{2,0}}{\sigma_1} L_1(f_{\mu_{2,0}}) > \frac{\mu_{2,0}}{\sigma_1} L_1(f_{\mu_{2,0}}).$$

Поэтому, в силу леммы 5.2.3 начальная пара $(\bar{x}_0, \bar{u}_0)$ удовлетворяет условию обратного зазора. Мы уже проверили, что последовательность $\{\tau_k\}_{k=0}^{\infty}$ пересчитываемая по правилу (5.2.33), удовлетворяет условиям теоремы 5.2.1. Таким образом, условия обратного зазора будут выполнены для последовательности, сформированной методом (5.2.34). Осталось воспользоваться неравенством (5.2.28). □

Ясно, что точно такое же утверждение может быть доказано для метода, использующего схему (5.2.25).

5.2.7 Минимизация сильно выпуклых функций

Рассмотрим теперь модель (5.2.2), удовлетворяющую следующему предположению.

Предположение 5.2.1 В представлении (5.2.2) функция $\hat{f}(x)$ сильно выпукла с параметром $\hat{\sigma} > 0$.

Тогда верен следующий вариант теоремы Данскина.

Лемма 5.2.6 *Если предположение 5.2.1 выполнено, то функция $\phi(u)$ определенная в (5.2.3) вогнута и дифференцируема. Более того, ее градиент представимый в виде*

$$\nabla\phi(u) = -\nabla\hat{\phi}(u) + Ax_0(u) \tag{5.2.36}$$

с точкой $x_0(u)$, определеной в (5.2.9), удовлетворяет условию Липшица с константой

$$L_2(\phi) = \tfrac{1}{\hat{\sigma}}\|A\|^2_{1,2} + L_2(\hat{\phi}). \tag{5.2.37}$$

Доказательство.
Обозначим $\tilde{\phi}(u) = \min\limits_{x\in Q_1}\{\langle Ax, u\rangle_2 + \hat{f}(x)\}$. Эта функция вогнута как минимум линейных функций. Поскольку функция $\hat{f}$ сильно выпукла, решение последней задачи минимизации единственно. Таким образом, функция $\tilde{\phi}(u)$ дифференцируема и $\nabla\tilde{\phi}(u) = Ax_0(u)$.

Рассмотрим две точки u_1 и u_2. Из условий оптимальности первого порядка для задачи (5.2.3) имеем

$$\langle A^*u_1 + \nabla\hat{f}(x_0(u_1)), x_0(u_2) - x_0(u_1)\rangle_1 \geq 0,$$

$$\langle A^*u_2 + \nabla\hat{f}(x_0(u_2)), x_0(u_1) - x_0(u_2)\rangle_1 \geq 0.$$

Складывая эти неравенства и пользуясь сильной выпуклостью функции $\hat{f}(\cdot)$ получаем

$$\begin{aligned}
\langle Ax_0(u_2) - Ax_0(u_1), u_1 - u_2\rangle_2 \quad &\geq \quad \langle\nabla\hat{f}(x_0(u_1)) - \nabla\hat{f}(x_0(u_2)), x_0(u_1) - x_0(u_2)\rangle_1 \\
\text{(из (1.1.21))} \quad &\geq \quad \hat{\sigma}\|x_0(u_1) - x_0(u_2)\|_1^2 \\
\text{(из (5.1.2))} \quad &\geq \quad \tfrac{\hat{\sigma}}{\|A\|^2_{1,2}}\left(\|\nabla\tilde{\phi}(u_1) - \nabla\tilde{\phi}(u_2)\|_2^*\right)^2.
\end{aligned}$$

Таким образом, $\|\nabla\tilde{\phi}(u_1) - \nabla\tilde{\phi}(u_2)\|_2^* \leq \frac{1}{\hat{\sigma}}\|A\|^2_{1,2}\cdot\|u_1 - u_2\|_2$, что и дает выражения (5.2.37).
□

Лемма 5.2.7 *Для любых точек $u, \hat{u} \in Q_2$ справедливо неравенство*

$$\phi(\hat{u}) + \langle\nabla\phi(\hat{u}), u - \hat{u}\rangle_2 \geq -\hat{\phi}(u) + \langle Ax_0(\hat{u}), u\rangle_2 + \hat{f}(x_0(\hat{u})). \tag{5.2.38}$$

Доказательство.
Выберем произвольные точки $u, \hat{u} \in Q_2$. Обозначим $\hat{x} = x_0(\hat{u})$. Тогда

$$\begin{aligned}
\phi(\hat{u}) + \langle\nabla\phi(\hat{u}), u - \hat{u}\rangle_2 \quad &= \quad -\hat{\phi}(\hat{u}) + \langle A\hat{x}, \hat{u}\rangle_2 + \hat{f}(\hat{x}) + \langle -\nabla\hat{\phi}(\hat{u}) + A\hat{x}, u - \hat{u}\rangle_2 \\
(\hat{\phi}\ \text{выпукла}) \quad &\geq \quad -\hat{\phi}(u) + \langle A\hat{x}, u\rangle_2 + \hat{f}(\hat{x}).
\end{aligned}$$

□

В этом пункте мы получим метод минимизации из следующего варианта условия обратного зазора:

$$f_{\mu_2}(\bar{x}) \leq \phi(\bar{u}) \tag{5.2.39}$$

при некоторых $\bar{x} \in Q_1$ и $\bar{u}$ in Q_2.

Это условие является частным случаем условия (5.2.13) с параметром $\mu_1 = 0$. Однако мы не можем напрямую воспользоваться результатами предыдущих разделов поскольку наши предположения изменились. В частности, теперь мы не предполагаем что множество Q_1 ограничено.

Лемма 5.2.8 *Предположим, что точки $\bar{x}$ из Q_1 и $\bar{u}$ из Q_2 удовлетворяют условию (5.2.39). Тогда*

$$0 \leq f(\bar{x}) - \phi(\bar{u}) \leq \mu_2 D_2. \tag{5.2.40}$$

Доказательство.
Действительно, для любого $x \in Q_1$ имеем $f_{\mu_2}(x) \geq f(x) - \mu_2 D_2$. □

Определим сопряженное градиентное отображение следующим образом:

$$V(u) = \arg\max_{v \in Q_2} \left\{ \langle \nabla\phi(u), v - u \rangle_2 - \tfrac{1}{2} L_2(\phi) \|v - u\|_2^2 \right\}. \tag{5.2.41}$$

Лемма 5.2.9 *Условие обратного зазора (5.2.39) выполняется для $\mu_2 = L_2(\phi)$ и точек*

$$\bar{x} = x_0(u_0), \quad \bar{u} = V(u_0). \tag{5.2.42}$$

Доказательство.
Действительно, в силу леммы 5.2.6 и неравенства (1.1.15) получаем:

$$\begin{array}{rcl}
\phi(V(u_0)) & \geq & \phi(u_0) + \langle \nabla\phi(u_0), V(u_0) - u_0 \rangle_2 - \frac{1}{2} L_2(\phi) \|V(u_0) - u_0\|_2^2 \\[2ex]
(\text{из } (5.2.41)) & = & \max\limits_{u \in Q_2} \left\{ \phi(u_0) + \langle \nabla\phi(u_0), u - u_0 \rangle_2 - \frac{1}{2} L_2(\phi) \|u - u_0\|_2^2 \right\} \\[2ex]
(\text{из } (5.2.3) \text{ и } (5.2.36)) & = & \max\limits_{u \in Q_2} \Big\{ -\hat{\phi}(u_0) + \langle A x_0(u_0), u_0 \rangle_2 + \hat{f}(x_0(u_0)) \\[2ex]
& & + \langle A x_0(u_0) - \nabla\hat{\phi}(u_0), u - u_0 \rangle_2 - \frac{1}{2} \mu_2 \|u - u_0\|_2^2 \Big\} \\[2ex]
(\hat{\phi} \text{ is convex и } (5.2.4)) & \geq & \max\limits_{u \in Q_2} \Big\{ -\hat{\phi}(u) + \hat{f}(x_0(u_0)) + \langle A x_0(u_0), u \rangle_2 - \mu_2 d_2(u) \Big\} \\[2ex]
(\text{из } (5.2.5)) & = & f_{\mu_2}(x_0(u_0)).
\end{array}$$

□

Теорема 5.2.4 *Предположим, что при некотором положительном* μ_2 *точки* $\bar{x} \in Q_1$ *и* $\bar{u} \in Q_2$ *удовлетворяют условию обратного зазора (5.2.39). Зафиксируем* $\tau \in (0,1)$ *и выберем* $\mu_2^+ = (1-\tau)\mu_2$,

$$\begin{aligned} \hat{u} &= (1-\tau)\bar{u} + \tau u_{\mu_2}(\bar{x}), \\ \bar{x}_+ &= (1-\tau)\bar{x} + \tau x_0(\hat{u}), \\ \bar{u}_+ &= V(\hat{u}). \end{aligned} \tag{5.2.43}$$

Тогда пара $(\bar{x}_+, \bar{u}_+)$ *удовлетворяет условию (5.2.39) с параметром сглаживания* μ_2 *при условии, что* τ *достаточно мало:*

$$\tfrac{\tau^2}{1-\tau} \le \tfrac{\mu_2}{L_2(\phi)}. \tag{5.2.44}$$

Доказательство.

Обозначим $\hat{x} = x_0(\hat{u})$ и $u_2 = u_{\mu_2}(\bar{x})$. В силу второй строки в правилах (5.2.43) и представления (5.2.5) имеем

$$\begin{aligned} f_{\mu_2^+}(\bar{x}_+) &= \hat{f}(\bar{x}_+) + \max_{u\in Q_2}\Big\{\langle A((1-\tau)\bar{x}+\tau\hat{x}), u\rangle_2 - \hat{\phi}(u) - (1-\tau)\mu_2 d_2(u)\Big\} \\ (\hat{f}\ \text{выпукла}) &\le \max_{u\in Q_2}\Big\{(1-\tau)\Big[\hat{f}(\bar{x}) + \langle A\bar{x}, u\rangle_2 - \hat{\phi}(u) - \mu_2 d_2(u)\Big] \\ &\quad +\tau[\hat{f}(\hat{x}) + \langle A\hat{x}, u\rangle_2 - \hat{\phi}(u)]\Big\} \\ (\text{из } (1.1.30)) &\le \max_{u\in Q_2}\big\{(1-\tau)\left[f_{\mu_2}(\bar{x}) - \tfrac{1}{2}\mu_2\|u-u_2\|_2^2\right] \\ (\text{из } (5.2.38)) &\quad +\tau[\phi(\hat{u}) + \langle \nabla\phi(\hat{u}), u-\hat{u}\rangle_2]\big\}. \end{aligned}$$

Поскольку функция ϕ вогнута, в силу неравенства (5.2.39) получаем:

$$\begin{aligned} f_{\mu_2}(\bar{x}) &\le \phi(\bar{u}) \le \phi(\hat{u}) + \langle \nabla\phi(\hat{u}), \bar{u}-\hat{u}\rangle_2 \\ (\text{строка 1, } (5.2.43)) &= \phi(\hat{u}) + \tau\langle \nabla\phi(\hat{u}), \bar{u}-u_2\rangle_2. \end{aligned}$$

Поэтому можно завершить доказательство следующим образом:

$$\begin{array}{rcl}
f_{\mu_2^+}(\bar{x}_+) & \leq & \max\limits_{u \in Q_2} \left\{ \phi(\hat{u}) + \tau \langle \nabla \phi(\hat{u}), u - u_2 \rangle_2 - \frac{1}{2}(1-\tau)\mu_2 \| u - u_2 \|_2^2 \right\} \\
(\text{из } (5.2.44)) & \leq & \max\limits_{u \in Q_2} \left\{ \phi(\hat{u}) + \tau \langle \nabla \phi(\hat{u}), u - u_2 \rangle_2 - \frac{1}{2}\tau^2 L_2(\phi) \| u - u_2 \|_2^2 \right\} \\
(v = \bar{u} + \tau(u - \bar{u})) & = & \max\limits_{v \in \bar{u} + \tau(Q_2 - \bar{u})} \left\{ \phi(\hat{u}) + \langle \nabla \phi(\hat{u}), v - \hat{u} \rangle_2 - \frac{1}{2} L_2(\phi) \| v - \hat{u} \|_2^2 \right\} \\
(Q_2 \text{ выпукло}) & \leq & \max\limits_{v \in Q_2} \left\{ \phi(\hat{u}) + \langle \nabla \phi(\hat{u}), v - \hat{u} \rangle_2 - \frac{1}{2} L_2(\phi) \| v - \hat{u} \|_2^2 \right\} \\
(\text{из } (5.2.41)) & = & \phi(\hat{u}) + \langle \nabla \phi(\hat{u}), \bar{u}_+ - \hat{u} \rangle_2 - \frac{1}{2} L_2(\phi) \| \bar{u}_+ - \hat{u} \|_2^2 \\
(\text{из } (1.1.15)) & \leq & \phi(\bar{u}_+).
\end{array}$$

□

Теперь мы можем обосновать следующий метод минимизации.

1. **Инициализация**
   Полагаем $\mu_{2,0} = 2L_2(\phi)$, $\bar{x}_0 = x_0(u_0)$ и $\bar{u}_0 = V(u_0)$.

2. **Для $k \geq 0$ повторяем:**

$$\begin{array}{rl}
\text{полагаем } \tau_k = \frac{2}{k+3} \text{ и } \hat{u}_k = (1-\tau_k)\bar{u}_k + \tau_k u_{\mu_{2,k}}(\bar{x}_k); & \\
\text{пересчитываем } \quad \mu_{2,k+1} & = (1-\tau_k)\mu_{2,k}, \\
\bar{x}_{k+1} & = (1-\tau_k)\bar{x}_k + \tau_k x_0(\hat{u}_k), \\
\bar{u}_{k+1} & = V(\hat{u}_k).
\end{array} \tag{5.2.45}$$

Теорема 5.2.5 *Пусть задача (5.2.1) удовлетворяет предположению 5.2.1. Тогда пары точек* $(\bar{x}_k, \bar{u}_k)$*, сформированные методом (5.2.45), удовлетворяют следующему неравенству:*

$$f(\bar{x}_k) - \phi(\bar{u}_k) \leq \frac{4L_2(\phi)D_2}{(k+1)(k+2)}, \tag{5.2.46}$$

где константа $L_2(\phi)$ *определена в формуле (5.2.37).*

Доказательство.

Действительно, в силу теоремы 5.2.4 и леммы 5.2.9 нам нужно только доказать, что последовательности $\{\mu_{2,k}\}_{k=0}^{\infty}$ и $\{\tau_k\}_{k=0}^{\infty}$ удовлетворяют соотношению (5.2.44). Это действительно так поскольку

$$\mu_{2,k} = \frac{4L_2(\phi)}{(k+1)(k+2)\sigma_2}$$

при всех $k \geq 0$. □

В заключение этого параграфа приведем один пример. Рассмотрим задачу

$$f(x) = \tfrac{1}{2}\|x\|_1^2 + \max_{1\leq j\leq m}[f_j + \langle g_j, x - x_j\rangle_1] \quad \rightarrow \quad \min: \ x \in E_1. \tag{5.2.47}$$

Задачи такого типа возникают, например, как вспомогательные задачи в методе пучков [59]. Пусть $E_1 = \mathbb{R}^n$ и мы выбрали

$$\|x\|_1^2 = \sum_{i=1}^{n}(x^{(i)})^2, \quad x \in E_1.$$

Тогда эта задача может быть решена методом (5.2.45).

Действительно, целевая функция в (5.2.47) может быть представлена в форме (5.2.2) следующим образом. Выберем

$$E_2 = \mathbb{R}^m, \quad Q_2 = \Delta_m = \{u \in \mathbb{R}^m_+ : \ \sum_{j=1}^{m} u^{(j)} = 1\},$$

$$\hat{f}(x) = \tfrac{1}{2}\|x\|_1^2, \quad \hat{\phi}(u) = \langle b, u\rangle_2, \quad b^{(j)} = \langle g_j, x_j\rangle_1 - f_j, \ j = 1, \ldots, m,$$

$$A^T = (a_1, \ldots, a_m).$$

Тогда $\hat{\sigma} = 1$ и $L_2(\hat{\phi}) = 0$. В пространстве E_2 выберем норму

$$\|u\|_2 = \sum_{j=1}^{m} |u^{(j)}|.$$

Тогда можно использовать энтропийное расстояние

$$d_2(u) = \ln m + \sum_{j=1}^{m} u^{(j)} \ln u^{(j)}, \quad u_0 = (\tfrac{1}{m}, \ldots, \tfrac{1}{m}),$$

для которого $\sigma_2 = 1$ и $D_2 = \ln m$. Заметим, что в этом случае

$$\|A\|_{1,2} = \max_{1\leq j\leq m} \|g_j\|_1^*.$$

Таким образом, метод (5.2.45), применяемый к задаче (5.2.47), сходится со следующей скоростью:

$$f(\bar{x}_k) - \phi(\bar{u}_k) \leq \tfrac{4\ln m}{(k+1)(k+2)} \cdot \max_{1\leq j\leq m} \left(\|g_j\|_1^*\right)^2.$$

Остановимся на вычислительных затратах этого метода применительно к нашей задаче. На каждой итерации им производятся следующие операции.

1. **Вычисление точки** $u_{\mu_2}(\bar{x})$. Это решение следующей задачи:

$$\max_u \left\{ \sum_{j=1}^{m} u^{(j)} s^{(j)}(\bar{x}) - \mu_2 d_2(u) : \ u \in Q_2 \right\},$$

где $s^{(j)}(\bar{x}) = f_j + \langle g_j, \bar{x} - x_j \rangle$, $j = 1, \dots, m$. Как мы уже видели, ее решение может быть найдено по формуле

$$u^{(j)}_{\mu_2}(\bar{x}) = e^{s^{(j)}(\bar{x})/\mu_2} \cdot \left[\sum_{l=1}^{m} e^{s^{(l)}(\bar{x})/\mu_2} \right]^{-1}, \quad j = 1, \dots, m.$$

2. **Вычисление точки** $x_0(\hat{u})$. В нашем случае это решение следующей задачи:

$$\min_x \{ \langle Ax, \hat{u} \rangle_2 + \tfrac{1}{2} \|x\|_1^2 : \ x \in E_1 \}.$$

   Поэтому ответ очень простой: $x_0(\hat{u}) = -A^T \hat{u}$.

3. **Вычисление** $V(\hat{u})$. В нашем случае

$$\begin{aligned} \phi(\bar{u}) &= \min_{x \in E_1} \left\{ \sum_{j=1}^{m} u^{(j)} [f_j + \langle g_j, x - x_j \rangle_1] + \tfrac{1}{2} \|x\|_1^2 \right\} \\ &= -\langle b, u \rangle_2 - \tfrac{1}{2} \left(\|A^T \hat{u}\|_1^* \right)^2. \end{aligned}$$

   Таким образом, $\nabla \phi(\bar{u}) = -b - AA^T \hat{u}$. Теперь можно вычислить $V(\hat{u})$ с помощью формулы (5.2.41). В конце параграфа 5.1 мы показали, как можно вычислить $V(\bar{u})$ за $O(m \ln m)$ операций.

Таким образом, мы видим, что все операции, необходимые методу (5.2.45) для решения задачи (5.2.47), очень простые. Наиболее трудной из них является умножение матрицы A на вектор. Прямолинейная реализация метода потребует три таких умножения. Однако небольшая модификация схемы позволяет сократить их число до двух.

5.3 Техника сглаживания в полуопределенной оптимизации

5.3.1 Введение

Мотивировка

В параграфах 5.1 и 5.2 было показано, что правильное использование внутренней структуры негладких выпуклых задач позволяет получить эффективные градиентные методы, чья скорость сходимости на порядок превосходит максимальные скорости, устанавливаемые теорией сложности для черноящичных задач. Для такого ускорения необходимо сформировать вычисляемую гладкую аппроксимацию целевой функции и применить быстрый градиентный метод (1.2.18).

Предыдущие результаты касались в основном кусочно-линейных функций. В этом параграфе мы распространим их на задачи полуопределенной оптимизации. Для этого мы покажем, как формировать гладкие аппроксимации для двух наиболее важных симметричных функций собственных значений симметрических матриц, *максимального собственного значения* и *спектрального радиуса.* Наши аппроксимации основываются на энтропийном сглаживании и на аппроксимации равномерной нормы с помощью подходящих l_p-норм. Конечно же, техника сглаживания могла бы применяться к этим задачам и на основе фробениусовой нормы. Однако в этом случае качество получаемых аппроксимаций сильно зависело бы от размерности пространства.

Содержание

В п. 5.3.2 мы изучаем гладкие аппроксимации симметричных функций собственных значений симметрических матриц (так называемых *спектральных функций*). Для нас основным вопросом является липшицевость градиента такой функции. В принципе это свойство можно было бы вывести из общей формулы для гессиана спектральных функций. К сожалению, насколько нам известно, это еще никому не удавалось сделать без существенной потери в качестве оценки. Поэтому мы приводим простую технику, которая позволяет получать оценки константы Липшица градиента для некоторых важных частных случаев. А именно, значение такой спектральной функции должно получаться как сумма одной и той же абсолютно монотонной функции от одной переменной, примененной ко всем собственным значениям (см. теорему 5.3.1). Показывается, что такие функции хорошо аппроксимируют максимальное собственное значение и спектральный радиус симметрической матрицы. Более того, их можно легко минимизировать быстрым градиентным методом (1.2.18).

В п. 5.3.3 мы показываем, как применять технику сглаживания для минимизации максимального собственного значения матриц, линейно зависящих от нескольких переменных. Оценка эффективности нашего метода оказывается пропорциональной $\frac{1}{\epsilon}$, где ϵ – требуемая абсолютная точность решения задачи. В п. 5.3.4 техника сглаживания применяется для минимизации спектрального радиуса. Число итераций в данном случае оказывается ограниченным числом $\frac{4}{\delta}\sqrt{(1+\delta)r\ln r}$, где δ – требуемая *относительная точность*, а r – максимальный ранг матриц, возникающих в соответствующих линейных матричных неравенствах. Одним из преимуществ этого метода является то, что в нем не надо вычислять полного разложения матрицы по базису сингулярных векторов.

Обозначения

Напомним, что через $\mathcal{M}_n$ мы обозначаем пространство $n \times n$-матриц, а через $\mathcal{S}_n \subset \mathcal{M}_n$ – пространство симметрических матриц. Матрицы всегда обозначаются большими латинскими буквами. В пространствах $\mathbb{R}^n$ и $\mathcal{M}_n$ используются стандартные скалярные произведения

$$\begin{aligned}\langle x,y\rangle &= \sum_{i=1}^{n} x^{(i)}y^{(i)},\ x,y\in\mathbb{R}^n,\\ \langle X,Y\rangle_M &= \sum_{i,j=1}^{n} X^{(i,j)}Y^{(i,j)},\ X,Y\in\mathcal{M}_n,\end{aligned}$$

Для матрицы $X\in\mathcal{S}_n$ вектор $\lambda(X)\in\mathbb{R}^n$ содержит все ее собственные значения, записанные в порядке убывания:

$$\lambda^{(1)}(X)\ge\lambda^{(2)}(X)\ge\ldots\ge\lambda^{(n)}(X),\quad X\in\mathcal{S}_n.$$

Таким образом, $\lambda_{\max}(X)=\lambda^{(1)}(X)$. Обозначение $D(\lambda)\in\mathcal{S}_n$ используется для диагональной матрицы с вектором $\lambda\in\mathbb{R}^n$ на основной диагонали. Напомним, что для любой матрицы $X\in\mathcal{S}_n$ существует разложение по базису собственных векторов:

$$X=U(X)D(\lambda(X))U(X)^T,$$

где $U(X):U(X)U(X)^T=I_n$, а $I_n\in\mathcal{S}_n$ – единичная матрица.

Наконец, остановимся на обозначениях, имеющих различный смысл для векторов и для матриц. Для вектора $\lambda\in\mathbb{R}^n$ через $|\lambda|\in\mathbb{R}^n$ обозначается вектор с координатами $|\lambda^{(i)}|$, $i=1,\ldots,n$. Обозначение $\lambda^k\in\mathbb{R}^n$ используется для вектора с компонентами $(\lambda^{(i)})^k$, $i=1,\ldots,n$. Однако для матрицы $X\in\mathcal{S}_n$ мы полагаем

$$|X|\overset{\text{def}}{=}U(X)D(|\lambda(X)|)U(X)^T\succeq 0,$$

а обозначение X^k используется для обычной степени матрицы. Поскольку возведение в степень $k\ge 0$ любой матрицы $X\succeq 0$ не меняет порядок ее собственных чисел, имеем

$$\lambda^k(X)\ =\ \lambda(X^k). \tag{5.3.1}$$

Далее, в пространстве $\mathbb{R}^n$ мы используем стандартные обозначения для l_p-норм:

$$\|x\|_{(p)}=\left[\sum_{i=1}^{n}|x^{(i)}|^p\right]^{1/p},\quad x\in\mathbb{R}^n,$$

где $p\ge 1$, и $\|x\|_{(\infty)}=\max\limits_{1\le i\le n}|x^{(i)}|$. Соответствующие нормы в $\mathcal{S}_n$ вводятся следующим образом:

$$\|X\|_{(p)}=\|\lambda(X)\|_{(p)}=\|\lambda(|X|)\|_{(p)},\quad X\in\mathcal{S}_n. \tag{5.3.2}$$

5.3.2 Гладкие симметричные функции собственных значений

Для $k\ge 1$ рассмотрим следующую функцию:

$$\pi_k(X)=\langle X^k,I_n\rangle_M=\sum_{i=1}^{n}(\lambda^{(i)}(X))^k,\quad X\in\mathcal{S}_n.$$

Получим верхнюю оценку на ее вторую производную. Заметим, что она нетривиальна только при $k \geq 2$.

Производные по направлению $H \in \mathcal{S}_n$ этой функции вычисляются по формулам

$$\begin{array}{rcl} \langle \nabla \pi_k(X), H \rangle_M & = & k \langle X^{k-1}, H \rangle_M, \\ \langle \nabla^2 \pi_k(X) H, H \rangle_M & = & k \sum\limits_{p=0}^{k-2} \langle X^p H X^{k-2-p}, H \rangle_M. \end{array} \tag{5.3.3}$$

Нам потребуется следующий результат.

Лемма 5.3.1 *Для любых степеней $p, q \geq 0$ и матриц $X, H \in \mathcal{S}_n$ справедлива оценка*

$$\langle X^p H X^q + X^q H X^p, H \rangle_M \leq 2 \langle |X|^{p+q}, H^2 \rangle_M \leq 2 \langle \lambda^{p+q}(|X|), \lambda^2(|H|) \rangle. \tag{5.3.4}$$

Доказательство.
Действительно, обозначим $\lambda = \lambda(X)$, $D = D(\lambda)$, $U = U(X)$ и $\hat{H} = U^T H U$. Тогда

$$\begin{array}{rcl} \langle X^p H X^q + X^q H X^p, H \rangle_M & = & \langle U D^p U^T H U D^q U^T + U D^q U^T H U D^p U^T, H \rangle_M \\ & = & \langle D^p \hat{H} D^q + D^q \hat{H} D^p, \hat{H} \rangle_M \\ & = & \sum\limits_{i,j=1}^{n} (\hat{H}^{(i,j)})^2 \left((\lambda^{(i)})^p (\lambda^{(j)})^q + (\lambda^{(i)})^q (\lambda^{(j)})^p \right) \\ & \leq & \sum\limits_{i,j=1}^{n} (\hat{H}^{(i,j)})^2 \left(|\lambda^{(i)}|^p |\lambda^{(j)}|^q + |\lambda^{(i)}|^q |\lambda^{(j)}|^p \right). \end{array}$$

Заметим, что для любых неотрицательных чисел a и b выполнено неравенство

$$0 \leq (a^p - b^p)(a^q - b^q) = (a^{p+q} + b^{p+q}) - (a^p b^q + a^q b^p).$$

Таким образом, предыдущую цепочку неравенств можно продолжить следующим образом:

$$\begin{array}{rcl} \langle X^p H X^q + X^q H X^p, H \rangle_M & \leq & \sum\limits_{i,j=1}^{n} (\hat{H}^{(i,j)})^2 \left(|\lambda^{(i)}|^{p+q} + |\lambda^{(j)}|^{p+q} \right) \\ & = & 2 \sum\limits_{i,j=1}^{n} (\hat{H}^{(i,j)})^2 |\lambda^{(i)}|^{p+q} \\ & = & 2 \langle D(|\lambda|)^{p+q} \hat{H}, \hat{H} \rangle_M \\ & = & 2 \langle D^{p+q}(|\lambda|), \hat{H}^2 \rangle_M \\ & = & 2 \langle |X|^{p+q}, H^2 \rangle_M. \end{array}$$

Следовательно, мы получили первое из неравенств (5.3.4). Далее, в силу неравенства фон Ноймана

$$\langle |X|^{p+q}, H^2\rangle_M \ \leq \ \langle \lambda^{p+q}(|X|^{p+q}), \lambda(H^2)\rangle \ \overset{(5.3.1)}{=} \ \langle \lambda^{p+q}(|X|), \lambda^2(|H|)\rangle,$$

что полностью доказывает неравенства в (5.3.4). □

Следствие 5.3.1 *При любом $k \geq 2$ выполнено неравенство*

$$\langle \nabla^2 \pi_k(X)H, H\rangle_M \leq k(k-1)\langle \lambda^{k-2}(|X|), \lambda^2(|H|)\rangle. \tag{5.3.5}$$

Доказательство.
Для $k = 2$ оценка очевидна. Для $k \geq 3$ в представлении (5.3.3) можно объединить члены в выражении $\sum\limits_{p=0}^{k-2}\langle X^p H X^{k-2-p}, H\rangle_M$ в симметричные пары вида

$$\langle X^p H X^{k-2-p} + X^{k-2-p} H X^p, H\rangle_M.$$

Применяя к каждой паре неравенство (5.3.4), получаем оценку (5.3.5). □

Пусть функция одной вещественной переменной $f(\tau)$ задана рядом

$$f(\tau) = a_0 + \sum_{k=1}^{\infty} a_k \tau^k,$$

где $a_k \geq 0$ при всех $k \geq 2$. Предположим, что ее область определения $\mathrm{dom}\, f = \{\tau : \ |\tau| < R\}$ является непустой. Для матриц $X \in \mathcal{S}_n$ рассмотрим следующую симметричную функцию от собственных значений:

$$F(X) = \sum_{i=1}^{n} f(\lambda^{(i)}(X)).$$

Ясно что $\mathrm{dom}\, F = \{X \in \mathcal{S}_n : \ \lambda^{(1)}(X) < R, \ \lambda^{(n)}(X) > -R\}$.

Теорема 5.3.1 *Для любого $X \in \mathrm{dom}\, F$ и $H \in \mathcal{S}_n$ имеем*

$$\langle \nabla^2 F(X)H, H\rangle \leq \sum_{i=1}^{n} \nabla^2 f(\lambda^{(i)}(|X|))(\lambda^{(i)}(|H|))^2.$$

Доказательство.
Действительно,

$$\begin{aligned} F(X) &= n \cdot a_0 + \sum_{i=1}^{n}\sum_{k=1}^{\infty} a_k (\lambda^{(i)}(X))^k \\ &= n \cdot a_0 + \sum_{k=1}^{\infty} a_k \sum_{i=1}^{n} (\lambda^{(i)}(X))^k \ = \ n \cdot a_0 + \sum_{k=1}^{\infty} a_k \pi_k(X). \end{aligned}$$

Таким образом, в силу неравенства (5.3.5)

$$\begin{array}{rcl}\langle \nabla^2 F(X)H, H\rangle_M & = & \sum\limits_{k=2}^{\infty} a_k \langle \nabla^2 \pi_k(X)H, H\rangle_M \\ & \leq & \sum\limits_{k=2}^{\infty} k(k-1)a_k \langle \lambda^{k-2}(|X|), \lambda^2(|H|)\rangle \\ & = & \sum\limits_{i=1}^{n}\sum\limits_{k=2}^{\infty} k(k-1)a_k (\lambda^{(i)}(|X|))^{k-2}(\lambda^{(i)}(|H|))^2 \\ & = & \sum\limits_{i=1}^{n} \nabla^2 f(\lambda^{(i)}(|X|))(\lambda^{(i)}(|H|))^2. \qquad \Box\end{array}$$

Приведем два важных примера симметричных функций от собственных значений.

1. Квадрат матричной l_p-нормы. Для целого $p \geq 1$ рассмотрим следующую функцию:

$$F_p(X) = \tfrac{1}{2}\|\lambda(X)\|^2_{(2p)} = \tfrac{1}{2}\langle X^{2p}, I_n\rangle_M^{1/p}, \quad X \in \mathcal{S}_n. \tag{5.3.6}$$

Таким образом, $F_p(X) = \frac{1}{2}(\pi_{2p}(X))^{1/p}$. Следовательно, в силу неравенства (5.3.5) для любого $X, H \in \mathcal{S}_n$ имеем

$$\begin{array}{rcl}\langle \nabla F_p(X), H\rangle_M & = & \frac{1}{2p}(\pi_{2p}(X))^{\frac{1}{p}-1}\langle \nabla\pi_{2p}(X), H\rangle_M, \\ \langle \nabla^2 F_p(X)H, H\rangle_M & = & \frac{1}{2p}\cdot\left(\frac{1}{p}-1\right)\cdot(\pi_{2p}(X))^{\frac{1}{p}-2}\langle \nabla\pi_{2p}(X), H\rangle_M^2 \\ & & +\frac{1}{2p}(\pi_{2p}(X))^{\frac{1}{p}-1}\langle \nabla^2\pi_{2p}(X)H, H\rangle_M \qquad (5.3.7) \\ & \leq & (2p-1)(\pi_{2p}(X))^{\frac{1}{p}-1}\langle \lambda^{2p-2}(|X|), \lambda^2(|H|)\rangle.\end{array}$$

Применим неравенство Гельдера: $\langle x, y\rangle \leq \|x\|_{(\beta)}\|y\|_{(\gamma)}$ с $\beta = \frac{p}{p-1}$, $\gamma = \frac{\beta}{\beta-1} = p$, и

$$x^{(i)} = (\lambda^{(i)}(|X|))^{2p-2}, \quad y^{(i)} = (\lambda^{(i)}(|H|))^2, \quad i = 1, \ldots, n.$$

Тогда

$$\begin{array}{rcl}\langle x, y\rangle & \leq & \left[\sum\limits_{i=1}^{n}(\lambda^{(i)}(|X|))^{2p}\right]^{\frac{p-1}{p}}\cdot\left[\sum\limits_{i=1}^{n}(\lambda^{(i)}(|H|))^{2p}\right]^{\frac{1}{p}} \\ & \overset{(5.3.2)}{=} & \pi_{2p}(X)^{\frac{p-1}{p}}\cdot\|\lambda(H)\|^2_{(2p)}\end{array}$$

и можно продолжить:

$$\begin{array}{rcl}\langle \nabla^2 F_p(X)H, H\rangle_M & \leq & (2p-1)\|\lambda(H)\|^2_{(2p)} \\ & = & (2p-1)\|H\|^2_{(2p)}.\end{array} \tag{5.3.8}$$

2. Энтропийное сглаживание максимального собственного значения. Рассмотрим функцию

$$E(X) = \ln \sum_{i=1}^{n} e^{\lambda^{(i)}(X)} \stackrel{\mathrm{def}}{=} \ln F(X), \quad X \in \mathcal{S}_n. \tag{5.3.9}$$

Заметим, что

$$\begin{array}{rcl} \langle \nabla E(X), H \rangle_M & = & \frac{1}{F(X)} \langle \nabla F(X), H \rangle_M, \\ \langle \nabla^2 E(X) H, H \rangle_M & = & -\frac{1}{F^2(X)} \langle \nabla F(X), H \rangle_M^2 + \frac{1}{F(X)} \langle \nabla^2 F(X) H, H \rangle_M \\ & \leq & \frac{1}{F(X)} \langle \nabla^2 F(X) H, H \rangle_M. \end{array}$$

Предположим сначала, что $X \succeq 0$. Функция $F(X)$ сформирована с помощью вспомогательной функции $f(\tau) = e^\tau$, которая удовлетворяет предположениям теоремы 5.3.1. Таким образом,

$$\langle \nabla^2 E(X) H, H \rangle_M \ \leq \ \left[\sum_{i=1}^{n} e^{\lambda^{(i)}(X)} \right]^{-1} \sum_{i=1}^{n} e^{\lambda^{(i)}(X)} (\lambda^{(i)}(|H|))^2 \leq \| H \|_{(\infty)}^2. \tag{5.3.10}$$

Остается заметить, что $E(X + \tau I_n) = E(X) + \tau$. Поэтому гессиан $\nabla^2 E(X + \tau I_n)$ не зависит от τ и мы заключаем, что оценка (5.3.10) верна для произвольной матрицы $X \in \mathcal{S}_n$.

5.3.3 Минимизация максимального собственного значения симметрической матрицы

Рассмотрим следующую задачу:

$$\text{найти } \phi^* = \min_y \{ \phi(y) \stackrel{\mathrm{def}}{=} \lambda_{\max}(C + A(y)) : \ y \in Q \}, \tag{5.3.11}$$

где Q – выпуклое замкнутое множество в $\mathbb{R}^m$ и $A(\cdot)$ – линейный оператор из $\mathbb{R}^m$ в $\mathcal{S}_n$:

$$A(y) = \sum_{i=1}^{m} y^{(i)} A_i \in \mathcal{S}_n, \quad y \in \mathbb{R}^m.$$

Заметим, что целевая функция в задаче (5.3.11) является негладкой. Таким образом, эту задачу можно решать либо методами внутренней точки (см. п. 1.3.3), либо общими методами негладкой выпуклой оптимизации (см. п. 1.2.1). Однако, ввиду очень специальной структуры целевой функции в задаче (5.3.11), можно попытаться разработать для нее специальный метод.

Мы собираемся решать задачу (5.3.11) с помощью техники сглаживания, описанной в параграфе 5.1. Для этого нужно заменить функцию $\lambda_{\max}(X)$ на ее гладкую аппроксимацию $f_\mu(X) = \mu E(\frac{1}{\mu}X)$, определяемую с помощью формулы (5.3.9) и параметра сглаживания $\mu > 0$. Заметим, что

$$\begin{array}{rcl} f_\mu(X) & = & \mu \ln \left[\sum\limits_{i=1}^{n} e^{\lambda^{(i)}(X)/\mu} \right] \geq \lambda_{\max}(X), \\ f_\mu(X) & \leq & \lambda_{\max}(X) + \mu \ln n. \end{array} \qquad (5.3.12)$$

В то же время

$$\nabla f_\mu(X) = \left[\sum\limits_{i=1}^{n} e^{\lambda^{(i)}(X)/\mu} \right]^{-1} \cdot \sum\limits_{i=1}^{n} e^{\lambda^{(i)}(X)/\mu} u_i(X) u_i(X)^T. \qquad (5.3.13)$$

В каждой тестовой точке X градиент $\nabla f_\mu(X)$ учитывает все собственные вектора матрицы X. Поскольку множители $e^{\lambda^{(i)}(X)/\mu}$ убывают очень быстро, градиент фактически зависит только от небольшого количества наибольших собственных значений. Их отбор производится автоматически с помощью выражения (5.3.13), где используется логарифмическая шкала, определяемая параметром μ.

Оценим теперь эффективность техники сглаживания применительно к задаче (5.3.11). Нашей целью является нахождение ϵ-решения $\bar{x} \in Q$ задачи (5.3.11):

$$\phi(\bar{y}) - \phi^* \leq \epsilon. \qquad (5.3.14)$$

Для этого нам нужно найти $\frac{1}{2}\epsilon$-решение сглаженной задачи

$$\text{найти } \phi^*_\mu = \min_y \{ \phi_\mu(y) \stackrel{\text{def}}{=} f_\mu(C + A(y)) : \ y \in Q \}, \qquad (5.3.15)$$

где параметр $\mu = \mu(\epsilon)$ задан по формуле

$$\mu(\epsilon) = \tfrac{\epsilon}{2 \ln n}. \qquad (5.3.16)$$

Ясно, что если $\phi_\mu(\bar{y}) - \phi^*_\mu \leq \frac{1}{2}\epsilon$, то в силу (5.3.12) получим

$$\phi(\bar{y}) - \phi^* \leq \phi_\mu(\bar{y}) - \phi^*_\mu + \mu \ln n \leq \epsilon.$$

Оценим теперь сложность нахождения $\frac{1}{2}\epsilon$-решения задачи (5.3.15) быстрым градиентным методом (1.2.18).

Для точек $h \in \mathbb{R}^m$ зафиксируем некоторую норму $\|h\|$. Рассмотрим прокс-функцию $d(x)$ множества Q с прокс-центром $x_0 \in Q$. Предполагается, что эта функция сильно выпукла на Q с параметром $\sigma > 0$. Зададим

$$\|A\| \ = \ \max_{h \in \mathbb{R}^m} \{ \|A(h)\|_{(\infty)} : \ \|h\| = 1 \}.$$

Заметим, что эта норма не очень велика. Действительно,

$$\|A(h)\|_{(\infty)} = \lambda^{(1)}(|A(h)|) \leq \langle A(h), A(h) \rangle_M^{1/2}, \quad h \in \mathbb{R}^m.$$

Таким образом, например, $\|A\| \le \|A\|_G \overset{\text{def}}{=} \max\limits_{\|h\|=1} \langle A(h), A(h)\rangle_M^{1/2}$.

Оценим вторую производную функции $\phi_\mu(y)$. Для любых $y, h \in \mathbb{R}^m$ в силу неравенства (5.3.10) имеем

$$\begin{aligned} \langle \nabla \phi_\mu(y), h\rangle &= \langle \nabla f_\mu(C + A(y)), h\rangle \\ &= \langle \nabla E(\tfrac{1}{\mu}(C + A(y))), A(h)\rangle_M, \\ \langle \nabla^2 \phi_\mu(y)h, h\rangle &= \tfrac{1}{\mu}\langle \nabla^2 E(C + A(y))A(h), A(h)\rangle_M \\ &\le \tfrac{1}{\mu}\|A(h)\|^2_{(\infty)} \\ &\le \tfrac{1}{\mu}\|A\|^2 \cdot \|h\|^2. \end{aligned}$$

Таким образом, в силу теоремы 1.1.1 функция $\phi_\mu(y)$ имеет липшицев градиент с константой

$$L = \tfrac{1}{\mu}\|A\|^2 = \tfrac{2\ln n}{\epsilon}\|A\|^2.$$

Принимая во внимание оценку (1.2.20), мы видим, что метод (1.2.18), применяемый к задаче (5.3.15), имеет следующую скорость сходимости:

$$\phi_\mu(y_k) - \phi_\mu^* \le \tfrac{8\ln n\|A\|^2 d(y_\mu^*)}{\epsilon\cdot\sigma(k+1)(k+2)},$$

где $y_\mu^* \in Q$ – решение задачи (5.3.15). Поэтому он может вычислить $\frac{1}{2}\epsilon$-решение этой задачи (которое является ϵ-решением задачи (5.3.11)) не более чем за

$$\tfrac{4\|A\|}{\epsilon}\sqrt{\tfrac{\ln n}{\sigma}\, d(y_\mu^*)} \tag{5.3.17}$$

итераций.

5.3.4 Минимизация спектрального радиуса симметрических матриц

Спектральный радиус матрицы $X \in \mathcal{S}_n$ определяется следующим образом:

$$\rho(X) = \max_{1\le i\le n} |\lambda^{(i)}(X)| = \max\{\lambda^{(1)}(X), -\lambda^{(n)}(X)\}.$$

Ясно, что $\rho(X)$ – выпуклая функция на всем пространстве $\mathcal{S}_n$. В этом разделе мы рассматриваем следующую задачу минимизации:

$$\text{найти } \phi_* = \min_{y\in\mathbb{R}^m}\{\phi(y) \overset{\text{def}}{=} \rho(A(y)) : \ y \in Q\}, \tag{5.3.18}$$

где $Q \subset \mathbb{R}^m$ – выпуклое замкнутое множество, не содержащее нуль, и $A(\cdot)$ – линейный оператор из $\mathbb{R}^m$ в $\mathcal{S}_n$:

$$A(y) = \sum_{i=1}^m y^{(i)} A_i \in \mathcal{S}_n, \quad y \in \mathbb{R}^m.$$

Будем считать, что матрицы $\{A_i\}_{i=1}^m$ линейно независимы. Поэтому матрица симметрическая матрица G с элементами

$$G^{(i,j)} = \langle A_i, A_j \rangle_M, \quad i,j = 1,\ldots,m,$$

положительно определена. Обозначим через r максимальный ранг матриц $A(y)$:

$$r = \max_{y\in\mathbb{R}^m} \operatorname{rank} A(y) \le \min\left\{ n, \sum_{i=1}^m \operatorname{rank} A_i \right\}.$$

Мы собираемся решать задачу (5.3.18) с помощью одного из вариантов специального подхода, позволяющего получать приближенные решения с *относительной точностью*. Заметим, что в силу наших предположений оптимальное значение ϕ^* всегда положительно.

Прежде всего нам нужно приблизить негладкую целевую функцию в задаче (5.3.18) некоторой гладкой функцией. Для этого воспользуемся функцией $F_p(X)$, определенной в (5.3.6). Заметим, что

$$\begin{array}{rcl} F_p(X) & = & \frac{1}{2}\langle X^{2p}, I_n\rangle_M^{1/p} \ge \frac{1}{2}\rho^2(X), \\ & & \\ F_p(X) & \le & \frac{1}{2}\rho^2(X)\cdot(\operatorname{rank} X)^{1/p}. \end{array} \tag{5.3.19}$$

Рассмотрим задачу

$$\text{найти } f_p^* = \min_{y\in\mathbb{R}^m}\{ f_p(y) \overset{\mathrm{def}}{=} F_p(A(y)) :\ y \in Q\}. \tag{5.3.20}$$

Из соотношений (5.3.19) мы видим, что

$$\tfrac{1}{2}\phi_*^2 \le f_p^* \le \tfrac{1}{2}\phi_*^2 \cdot r^{1/p}. \tag{5.3.21}$$

Нашей целью является нахождение точки $\bar{y} \in Q$, решающей задачу (5.3.18) с относительной точностью $\delta > 0$:

$$\phi(\bar{y}) \le (1+\delta)\phi_*.$$

Выберем целое p, удовлетворяющее неравенству

$$p(\delta) \overset{\mathrm{def}}{=} \tfrac{1+\delta}{\delta}\ln r \le p \le 2p(\delta). \tag{5.3.22}$$

Пусть точка $\bar{y} \in Q$ решает задачу (5.3.20) с относительной точностью δ. Тогда в силу неравенств (5.3.19) и (5.3.21) имеем

$$\begin{array}{rcl} \phi(\bar{y})/\phi_* & \le & r^{\frac{1}{2p}} \cdot \sqrt{f_p(\bar{y})/f_p^*} \\ & & \\ & \le & r^{\frac{1}{2p}} \cdot \sqrt{1+\delta} \\ & & \\ & \le & e^{\frac{\delta}{2(1+\delta)}} \cdot \sqrt{1+\delta} \ \le \ 1+\delta. \end{array}$$

Таким образом, нам нужно оценить эффективность метода (1.2.18) применительно к задаче (5.3.20). Введем норму

$$\|h\|_G = \langle Gh, h\rangle^{1/2}, \quad h \in \mathbb{R}^m.$$

Предполагая, что $p(\delta) \geq 1$ и пользуясь оценкой (5.3.8), для любых $y, h \in \mathbb{R}^m$ получаем:

$$\begin{aligned}
\langle \nabla^2 f_p(y)h, h\rangle &= \langle \nabla^2 F_p(A(y))A(h), A(h)\rangle_M \\
&\leq (2p-1)\|A(h)\|^2_{(2p)} \\
&\leq (2p-1)\|A(h)\|^2_{(2)} \\
&= (2p-1)\langle A(h), A(h)\rangle_M \\
&= (2p-1)\langle Gh, h\rangle \\
&= (2p-1)\|h\|^2_G.
\end{aligned}$$

Таким образом, в силу теоремы 1.1.1 функция $f_p(y)$ имеет градиент, который является липшицевым на $\mathbb{R}^m$ относительно нормы $\|\cdot\|_G$ с константой

$$L = 2p - 1 \leq 4p(\delta). \tag{5.3.23}$$

С другой стороны, для любого $X \in \mathcal{S}_n$, $\operatorname{rank} X \leq r$, и любого $p \geq 1$ имеем

$$\tfrac{1}{r}\|X\|^2_{(2)} \leq \|X\|^2_\infty \leq \|X\|^2_{(2p)}.$$

Поэтому $\frac{1}{2r}\|y\|^2_G \leq f_p(y)$ при всех $y \in \mathbb{R}^m$. В частности,

$$\tfrac{1}{2r}\|y^*_p\|^2_G \leq f^*_p, \tag{5.3.24}$$

где y^*_p – решение задачи (5.3.20).

Обозначим $x_0 = \arg\min\limits_y\{\|y\|_G : \ y \in Q\}$. Поскольку норма $\|\cdot\|_G$ – евклидова и Q – выпуклое множество, имеем

$$\|y^*_p - x_0\|^2_G \leq \|y^*_p\|^2_G - \|x_0\|^2_G < \|y^*_p\|^2_G.$$

Учитывая это неравенство и оценку (5.3.24), получаем

$$\tfrac{1}{2}\|y^*_p - x_0\|^2_G \leq \tfrac{1}{2}\|y^*_p\|^2_G \leq r f^*_p. \tag{5.3.25}$$

Для решения задачи (5.3.20) методом (1.2.18), выберем в качестве прокс-функции

$$d(x) = \tfrac{1}{2}\|x - x_0\|^2_G. \tag{5.3.26}$$

Поскольку параметр выпуклости σ этой функции равен единице, в силу оценок (5.3.23) и (5.3.25), метод (1.2.18), стартовавший из точки x_0, сходится следующим образом:

$$f_p(y_k) - f_p^* \quad \leq \quad \tfrac{16(1+\delta)r\ln r}{\delta\cdot(k+1)(k+2)} \cdot f_p^*. \tag{5.3.27}$$

Поэтому для решения задачи (5.3.20) с относительной точностью δ (и, таким образом, для решения задачи (5.3.18) с той же относительной точностью), методу (1.2.18) потребуется не более чем

$$\tfrac{4}{\delta}\sqrt{(1+\delta)r\ln r} \tag{5.3.28}$$

итераций. Заметим, что эта оценка никак не зависит от свойств конкретной задачи.

На каждой итерации метода (1.2.18), применяемого к задаче (5.3.20) с прокс-функцией $d(x)$, заданной по формуле (5.3.26), необходимо дважды вычислить проекцию точки на выпуклое множество Q в евклидовой метрике $\|\cdot\|_G$. Эта операция является простой в двух случаях.

- Множество Q – аффинное подпространство в $\mathbb{R}^m$. Тогда для вычисления проекции достаточно обратить матрицу G. Важным примером такой задачи является следующая постановка:
$$\min_{y\in\mathbb{R}^m}\left\{\rho\left(\sum_{i=1}^m y^{(i)}A_i\right) : \; y^{(1)} = 1\right\}.$$
- Матрица G и множество Q являются простыми. Например, если $\langle A_i, A_j\rangle = 0$ для $i \neq j$, то G – диагональная матрица. Такая ситуация возникает, когда матрица $A(y)$ параметризуется непосредственно через свои элементы.

Наконец заметим, что вычисление значения и градиента функции $f_p(y)$ может осуществляться без нахождения базиса собственных векторов матрицы $A(y)$. Действительно, пусть $p = 2^k$ удовлетворяет условию (5.3.22). Рассмотрим следующую последовательность матриц:

$$\begin{aligned} X_0 \;&=\; A(y), \quad Y_0 \;=\; I_n, \\ X_i \;&=\; X_{i-1}^2, \quad Y_i \;=\; Y_{i-1}X_{i-1}, \quad i = 1,\ldots,k. \end{aligned} \tag{5.3.29}$$

По индукции нетрудно убедиться в том, что $X_k = A^p(y)$ и $Y_k = A^{p-1}(y)$. Поэтому в соответствии с формулами (5.3.3), (5.3.6) и определением функции $f_p(y)$ в (5.3.20) имеем

$$\begin{aligned} f_p(y) \;&=\; \tfrac{1}{2}\langle X_k, I_n\rangle_M^{2/p}, \\ \nabla f_p(y)^{(i)} \;&=\; \tfrac{2f_p(y)}{\langle X_k, I_n\rangle_M}\cdot\langle Y_k, A_i\rangle_M, \quad i = 1,\ldots,m. \end{aligned}$$

Заметим, что сложность вычисления матрицы $A(y)$ имеет порядок $O(n^2 m)$ арифметических операций. Вспомогательное вычисление (5.3.29) потребует

$$O(n^3\ln p) = O\left(n^3\ln\tfrac{\ln r}{\delta}\right)$$

операций. После этого вектор $\nabla f_p(y)$ может быть вычислен за $O(n^2m)$ операций. Ясно, что трудоемкость первого и последнего вычисления сильно снижается в случае разреженных матриц A_i.

Заметим также, что вычисление (5.3.29) будет осуществляться более эффективно если матрица $A(y)$ представлена в форме

$$A(y) = UTU^T, \quad UU^T = I_n,$$

где T – три-диагональная матрица. Вычисление такого представления потребует $O(n^3)$ арифметических операций.

Глава 6

Оптимизация с относительной точностью

6.1 Однородная модель

6.1.1 Введение

Мотивировка

Обычно при теоретическом обосновании эффективности методов выпуклой минимизации рассматриваются ограниченные допустимые множества. Помимо чисто технических удобств, это предположение позволяет ввести естественный масштаб для измерения абсолютной точности приближенного решения. Если ограниченности нет, то многие методы требуют введения искусственных ограничений (подход “большого M”). Такой подход, возможно, оправдан для полиномиальных методов, где “большое M” появляется в оценке трудоемкости только под логарифмом. Однако для методов градиентного типа эту идею вряд ли можно признать удачной.

На самом деле, в теории оптимизации давно обсуждается вопрос о том возникают ли на самом деле в практических приложениях задачи с неограниченными допустимыми множествами. А если и возникают, то можно ли их решать с гарантированными оценками сложности. Не вдаваясь в общую дискуссию, отметим, что существует по крайней мере один класс важных задач такого типа. Это двойственные задачи, полученные с помощью лагранжевой релаксации линейных ограничений в прямой задаче. Если бы имелись разумные границы на размер таких двойственных множителей, то их стоило бы учесть в формулировке прямой задачи, что перевело бы эти ограничения в целевую функцию. Поскольку этого не делается, надо полагать, что такой информации нет.

В этом параграфе мы исследуем альтернативный способ решения задач с неограниченными множествами. А именно, мы собираемся их решать с *относительной точностью*. Мы покажем, что эта идея применима по крайней мере к одному специальному классу задач *конической безусловной минимизации.* Это задачи минимизации однородной выпуклой функции на аффинных подпространствах. Для решения таких задач с относительной точностью необходимо сначала построить приближенный эллипсоид Джона для субдифференциала целевой функции в нуле.

Оказывается, что во многих случаях вся информация, необходимая для применения этого подхода, может быть извлечена из исходных данных. Пользуясь этим подходом, мы можем решить, например, задачу

$$\sum_{i=1}^{m} |\langle a_i, x\rangle + c^{(i)}| \quad \to \quad \min_{x \in \mathbb{R}^n},$$

с относительной точностью δ за $O\left(\frac{\sqrt{m}}{\delta} \ln m\right)$ итераций метода градиентного типа. Наиболее эффективные из применяемых методов используют технику сглаживания, описанную в параграфе 5.1.

Содержание

В п. 6.1.2 мы описываем модель задачи безусловной конической минимизации и приводим несколько примеров. Размер оптимального решения в таких задачах измеряется с помощью параметра $\alpha \in (0, 1)$, ответственного за асферичность субдифференциала целевой функции в нуле. В п. 6.1.3 мы доказываем оценки эффективности двух черонящичных субградиентных методов, применяемых к коническим задачам. Они могут найти приближенное решение с относительной точностью δ за $O\left(\frac{1}{\alpha^4\delta^2}\right)$ и $O\left(\frac{1}{\alpha^2\delta^2} \ln \frac{1}{\alpha}\right)$ итераций соответственно. Наиболее быстрая схема использует рекуррентную схему пересчета оценок расстояния до оптимального решения. В п. 6.1.4 мы вводим двухуровневую модель целевой функции. С одной стороны, эта модель используется для построения подходящей евклидовой метрики в прямом пространстве. С другой стороны, она позволяет задействовать технику сглаживания из параграфа 5.1. Мы приводим два алгоритма этого типа с оценками эффективности $O\left(\frac{1}{\alpha^2\delta}\right)$ и $O\left(\frac{1}{\alpha\delta} \ln \frac{1}{\alpha}\right)$ итераций. Вторая схема использует рекуррентную стратегию из п. 6.1.3. В последнем пункте 6.1.5 рассматриваются приложения. Мы приводим оценки наиболее быстрого метода из п. 6.1.4 для задачи линейного программирования, задачи минимизации спектрального радиуса аффинного семейства симметрических матриц и задачи расчета оптимальных механических конструкций. Мы сравниваем эти оценки с оценками для полиномиальных методов отслеживания траекторий (см. п. 1.3.3). Оказывается, например, что для задач линейного программирования оценки эффективности нового градиентного метода лучше, если требуемая относительная точность не слишком мала. А именно, если $\delta \geq O\left(\frac{1}{\min\{n,m\}}\right)$.

Обозначения

Мы говорим, что значение $f(\bar{x})$, вычисленное в допустимой точке $\bar{x}$, приближает минимальное значение $f^* > 0$ с *относительной точностью* δ если

$$f(\bar{x}) \le (1+\delta) f^*.$$

Следующее обозначение используется для шаров относительно нормы $\| \cdot \|$:

$$B_{\|\cdot\|}(r) = \{ x \in \mathbb{R}^n : \; \|x\| \le r \}.$$

Точка $\pi_{Q,\|\cdot\|}(x)$ является проекцией точки x на множество Q относительно нормы $\| \cdot \|$. Если не возникает двусмысленности, указатель на норму может быть опущен.

Через I_n обозначается единичная матрица в $\mathbb{R}^n$, а e_i обозначает i-й координатный вектор. Вектор $\bar{e}_n$ состоит из одних единиц. Для $n \times n$-матрицы X мы обозначаем через $\lambda_1(X), \dots, \lambda_n(X)$ ее собственные значения.

6.1.2 Коническая задача безусловной минимизации

Задача, рассматриваемая в этом параграфе, ставится следующим образом:

$$\begin{array}{rcl} \text{найти} \quad f^* & = & \min\limits_x \{ f(x) : \; x \in \mathcal{L} \}, \\ \mathcal{L} & = & \{ x \in \mathbb{R}^n : \; Cx = b \}, \end{array} \tag{6.1.1}$$

где выпуклая функция $f(x)$ является однородной степени единица, C – $p \times n$-матрица, и $b \neq 0$. Не ограничивая общности можно считать, что матрица C имеет полный строчный ранг. Наше основное предположение о задаче (6.1.1) выглядит так:

$$\operatorname{dom} f \equiv \mathbb{R}^n, \quad 0 \in \operatorname{int} \partial f(0). \tag{6.1.2}$$

Другими словами, мы предполагаем, что функция f является опорной функцией некоторого компактного выпуклого множества, у которого нуль является внутренней точкой. Отсюда следует, что $f^* > 0$ и задача отыскания решения задачи (6.1.1) с некоторой относительной точностью поставлена корректно. Всюду далее мы называем постановку (6.1.1), (6.1.2) *конической* задачей безусловной минимизации.

Заметим, что любая задача безусловной минимизации

$$\min_{y \in \mathbb{R}^{n-1}} \phi(y)$$

с выпуклой целевой функцией $\phi(\cdot)$ может быть переписана в конической форме (6.1.1) с помощью простого приема:

$$x = (y, \tau) \in \mathbb{R}^{n-1} \times R^1_+, \quad f(x) = \tau \phi(y/\tau), \quad Cx \equiv \tau, \quad b = 1,$$

(для детального обсуждения см., например, монографию [59]). Однако в общем случае мы не можем гарантировать, что эта функция удовлетворяет условию (6.1.2).

Рассмотрим несколько примеров.

Пример 6.1.1 Пусть наша исходная задача состоит в нахождении минимума функции

$$\phi_1(y) = \max_{1 \le i \le m} |\langle a_i, y\rangle + c^{(i)}|, \quad y \in \mathbb{R}^{n-1}.$$

Введем $x = \begin{pmatrix} y \\ \tau \end{pmatrix}$, и $\hat{a}_i = \begin{pmatrix} a_i \\ c^{(i)} \end{pmatrix}$, $i = 1, \ldots, m$. Пусть

$$A^T = (\hat{a}_1 \ldots, \hat{a}_m), \quad F_1(v) = \max_{1 \le i \le m} |v^{(i)}|,$$

$$p = 1, \quad C = (\underbrace{0, \ldots, 0}_{(n-1)\text{ раз}}, 1), \quad b = 1.$$

Тогда для положительного τ можно определить

$$f(x) = \tau\phi_1(y/\tau) \equiv F_1(Ax).$$

Таким образом, последнее представление функции $f(x)$ может быть расширено на все пространство. Аналогично, для функции

$$\phi_\infty(y) = \sum_{i=1}^{m} |\langle a_i, y\rangle + c^{(i)}|, \quad y \in \mathbb{R}^{n-1},$$

можно получить представление (6.1.1), которое удовлетворяет условию (6.1.2). В этом случае мы пользуемся функцией $f(x) = F_\infty(Ax)$ с

$$F_\infty(v) = \sum_{i=1}^{m} |v^{(i)}|.$$

Однако для функции

$$\phi(y) = \max_{1 \le i \le m} [\langle a_i, y\rangle + c^{(i)}], \quad y \in \mathbb{R}^{n-1},$$

простое добавление одной проектирующей переменной не сможет гарантировать выполнение условия (6.1.2).

Зафиксируем некоторую норму $\|\cdot\|$ в $\mathbb{R}^n$. Тогда предположение (6.1.2) можно выразить в количественной форме. Пусть положительные параметры $\gamma_0 \le \gamma_1$ удовлетворяют следующему условию:

$$B_{\|\cdot\|^*}(\gamma_0) \subseteq \partial f(0) \subseteq B_{\|\cdot\|^*}(\gamma_1). \tag{6.1.3}$$

Таким образом, условие (6.1.2) нам гарантирует, что эти значения корректно определены. Заметим, что они, конечно же, зависят от выбора нормы $\|\cdot\|$. Однако в дальнейшем этот выбор всегда будет ясен из контекста.

Обозначим

$$\alpha = \frac{\gamma_0}{\gamma_1} < 1.$$

Как мы увидим, этот параметр определяет сложность нахождения решения задачи (6.1.1) с некоторой относительной точностью.

Может показаться, что эти оценки зависят от специфики конкретной задачи. Однако это не совсем так. Прежде всего заметим, что во многих ситуациях целесообразно выбирать норму $\| \cdot \|$ евклидовой. В силу теоремы Джона, в таких случаях можно надеяться на выполнение неравенства

$$\alpha \geq \frac{1}{n}. \tag{6.1.4}$$

Более того, если субдифференциал $\partial f(0)$ симметричен относительно нуля или, что то же самое, $f(x) = f(-x)$, $x \in \mathbb{R}^n$, то нижняя оценка для евклидовых норм будет еще лучше:

$$\alpha \geq \frac{1}{\sqrt{n}}. \tag{6.1.5}$$

Конечно же, задача отыскания хорошей нормы для конкретной функции $f(x)$ может оказаться достаточно сложной. Однако для ее решения можно воспользоваться знанием структуры целевой функции.

Например, может случится что нам известен самосогласованный барьер $\psi(v)$ для выпуклого множества $\partial f(0)$ и $\psi'(0) = 0$. Тогда можно воспользоваться нормами

$$\|v\|^* = \langle v, \psi''(0) v \rangle^{1/2}, \quad \|x\| = \langle [\psi''(0)]^{-1} x, x \rangle^{1/2}.$$

В этом случае можно выбрать

$$\gamma_0 = 1, \quad \gamma_1 = \nu + 2\sqrt{\nu},$$

где ν - параметр барьера $\psi(\cdot)$ (см. теорему 1.3.20).

Может оказаться, что субдифференциал $\partial f(0)$ – многогранник, заданный вершинами. Для таких задач бывает полезным следующий факт.

Лемма 6.1.1 *Пусть* $f(x) = \max\limits_{1 \leq i \leq m} \langle a_i, x \rangle$. *Предположим, что матрица* $A = (a_1, \ldots, a_m)$ *имеет полный строчный ранг и* $\sum\limits_{i=1}^{m} a_i = 0$*; таким образом* $m > n$. *Тогда норма*

$$\|x\| = \left[\sum_{i=1}^{m} \langle a_i, x \rangle^2 \right]^{1/2}$$

корректно определена и можно выбрать $\gamma_1 = 1$ *и* $\gamma_0 = \frac{1}{\sqrt{m(m-1)}}$.

Доказательство.

Заметим, что матрица $G = \sum\limits_{i=1}^{m} a_i a_i^T$ невырождена. Тогда $\|v\|^* = \langle v, G^{-1}v\rangle^{1/2}$ и, следовательно,

$$\begin{aligned}(\|a_i\|^*)^2 &= \langle a_i, G^{-1}a_i\rangle = \max_{x\in\mathbb{R}^n}\{2\langle a_i, x\rangle - \langle Gx, x\rangle\} = \max_{x\in\mathbb{R}^n}\left\{2\langle a_i, x\rangle - \sum_{k=1}^{m}\langle a_k, x\rangle^2\right\}\\ &\le \max_{x\in\mathbb{R}^n}\{2\langle a_i, x\rangle - \langle a_i, x\rangle^2\} = 1.\end{aligned}$$

Поскольку $\partial f(0) = \mathrm{Conv}\,\{a_i,\ i = 1,\dots,m\}$, можно взять $\gamma_1 = 1$.

С другой стороны, для любого $x \in \mathbb{R}^n$ имеем $\sum\limits_{i=1}^{m}\langle a_i, x\rangle = 0$. Таким образом,

$$\langle Gx, x\rangle = \sum_{i=1}^{m}\langle a_i, x\rangle^2 \le \max_{s\in\mathbb{R}^m}\left\{\sum_{i=1}^{m}(s^{(i)})^2 :\ \sum_{i=1}^{m} s^{(i)} = 0,\ s^{(i)} \le f(x),\ i = 1,\dots,m\right\}.$$

В этой задаче максимизации решение достигается в нескольких точках, например, в

$$\hat{s} = f(x)\cdot(\bar{e}_m - me_1).$$

Это означает, что $\langle Gx, x\rangle \le m(m-1)f^2(x)$. То есть, $f(x) \ge \frac{\|x\|}{\sqrt{m(m-1)}}$, что позволяет выбрать $\gamma_0 = \frac{1}{\sqrt{m(m-1)}}$. □

Возможность использования другого структурного представления задачи (6.1.1) обсуждается в п. 6.1.4.

Завершим этот раздел утверждением, которое определяет наши возможности по решению задачи (6.1.1) с относительной точностью. Обозначим через x_0 проекцию начала координат на аффинное подпространство $\mathcal{L}$ относительно нормы $\|\cdot\|$:

$$\|x_0\| = \min_x\{\|x\| :\ Cx = b\}.$$

Теорема 6.1.1

1. Для любого $x \in \mathbb{R}^n$ имеем

$$\gamma_0\cdot\|x\| \le f(x) \le \gamma_1\cdot\|x\|. \tag{6.1.6}$$

Таким образом, функция $f(x)$ липшицева на $\mathbb{R}^n$ относительно нормы $\|\cdot\|$ с константой γ_1. Более того, справедливы неравенства

$$\alpha f(x_0) \le \gamma_0\cdot\|x_0\| \le f^* \le f(x_0) \le \gamma_1\cdot\|x_0\|. \tag{6.1.7}$$

2. Для любого оптимального решения x^* задачи (6.1.1) имеем

$$\|x_0 - x^*\| \le \tfrac{2}{\gamma_0}f^* \le \tfrac{2}{\gamma_0}f(x_0). \tag{6.1.8}$$

Если норма $\|\cdot\|$ евклидова, то это неравенство может быть усилено:

$$\|x_0 - x^*\| \le \tfrac{1}{\gamma_0}f^* \le \tfrac{1}{\gamma_0}f(x_0). \tag{6.1.9}$$

Доказательство.
Для любого $x \in \mathbb{R}^n$ имеем

$$f(x) \;=\; \max_v \{\langle v, x\rangle :\; v \in \partial f(0)\} \geq \max_u \{\langle v, x\rangle :\; v \in B_{\|\cdot\|^*}(\gamma_0)\} = \gamma_0 \cdot \|x\|,$$

$$f(x) \;=\; \max_v \{\langle v, x\rangle :\; v \in \partial f(0)\} \leq \max_u \{\langle v, x\rangle :\; v \in B_{\|\cdot\|^*}(\gamma_1)\} = \gamma_1 \cdot \|x\|.$$

Таким образом, для любых $x, h \in \mathbb{R}^n$ выполнено неравенство

$$f(x+h) \leq f(x) + f(h) \leq f(x) + \gamma_1 \cdot \|h\|.$$

Более того,

$$f^* \;=\; \min_x \{f(x) :\; Cx = b\} \geq \min_x \{\gamma_0 \|x\| :\; Cx = b\} = \gamma_0 \cdot \|x_0\|.$$

Поэтому в силу (6.1.6) имеем

$$f^* \;\geq\; \gamma_0 \cdot \|x_0\| \geq \alpha f(x_0),$$

$$f^* \;\leq\; f(x_0) \leq \gamma_1 \cdot \|x_0\|.$$

Для доказательства второго утверждения заметим, что первое утверждение теоремы дает оценку

$$\|x_0 - x^*\| \leq \|x_0\| + \|x^*\| \leq \tfrac{2}{\gamma_0} \cdot f^*.$$

Для евклидовой нормы $\|x\| = \langle Gx, x\rangle^{1/2}$, $G \succ 0$, эта оценка может быть усилена. Действительно, в этом случае $\langle Gx_0, x^* - x_0\rangle = 0$. Таким образом,

$$\|x_0 - x^*\|^2 = \|x^*\|^2 - \|x_0\|^2 < \|x^*\|^2. \qquad \Box$$

6.1.3 Субградиентная аппроксимирующая схема

Обсудим теперь различные стратегии решения задачи (6.1.1). Для простоты предположим, что норма $\|\cdot\|$ евклидова.

Первая схема основывается на стандартном субградиентном методе для минимизации негладких выпуклых функций. Обозначим через $g(x)$ произвольный субградиент функции $f(x)$ в точке x. Рассмотрим простейший вариант этого метода, применяемый к задаче (6.1.1).

```
Субградиентный метод $G_N(R)$

Для $k = 0 \dots N$ повторяем:

    Вычисляем $f(x_k)$ и $g(x_k)$.

    $x_{k+1} = \pi_{\mathcal{L}} \left( x_k - \frac{R}{\sqrt{N+1}} \cdot \frac{g(x_k)}{\|g(x_k)\|^*} \right).$

Результат
$\bar{x} = \arg\min\{f(x) : \ x = x_0, \dots, x_N\}.$
```

(6.1.10)

Обозначим результат $\bar{x}$ работы этого метода через $G_N(R)$. В силу теоремы 3.2.2 из монографии [83] этот метод сходится следующим образом:

$$f(G_N(R)) - f^* \leq \frac{\gamma_1}{\sqrt{N+1}} \cdot \frac{\|x_0 - x^*\|^2 + R^2}{2R}. \tag{6.1.11}$$

Для этого метода очень важно иметь хорошую оценку расстояния между начальной точкой x_0 и решением x^*:

$$R \ \approx \ \|x_0 - x^*\|.$$

В нашем случае эта оценка могла бы быть получена из первой части неравенства в (6.1.9). Однако, так как заранее значение f^* обычно неизвестно, мы будем пользоваться его второй частью:

$$\hat{\rho} \overset{\text{def}}{=} \frac{1}{\gamma_0} f(x_0). \tag{6.1.12}$$

Эффективность такого выбора описывается следующим утверждением.

Теорема 6.1.2 *Для фиксированного δ из $(0,1)$ выберем*

$$N = \left\lfloor \frac{1}{\alpha^4 \delta^2} \right\rfloor. \tag{6.1.13}$$

Тогда $f(G_N(\hat{\rho})) \leq (1+\delta) \cdot f^*$.

Доказательство.
В силу неравенства (6.1.11), выбора (6.1.13) и неравенств (6.1.9), (6.1.7) имеем

$$f(G_N(\hat{\rho})) - f^* \leq \alpha^2 \delta \gamma_1 \cdot \frac{\|x_0 - x^*\|^2 + \hat{\rho}^2}{2\hat{\rho}} \leq \alpha^2 \delta \gamma_1 \hat{\rho} = \alpha \delta f(x_0) \leq \delta \cdot f^*. \qquad \square$$

Заметим, что мы заплатили довольно высокую цену за плохую оценку начального расстояния до решения. Если бы мы могли воспользоваться первой частью неравенства

(6.1.9), то оценка эффективности нашего метода была бы гораздо лучше. Покажем, что нашу оценку можно также улучшить пользуясь тривиальным наблюдением, что $f^* \leq f(x)$ для любой точки x из $\mathcal{L}$.

Пусть $\delta \in (0,1)$ – требуемая относительная точность. И пусть

$$\hat{N} = \left\lfloor \tfrac{e}{\alpha^2} \cdot \left(1 + \tfrac{1}{\delta}\right)^2 \right\rfloor,$$

где e – число Эйлера. Рассмотрим следующий процесс. Полагаем $\hat{x}_0 = x_0$, и для $t \geq 1$ повторяем следующие операции:

$$\boxed{\begin{array}{l} \hat{x}_t := G_{\hat{N}}\left(\tfrac{1}{\gamma_0} f(\hat{x}_{t-1})\right); \\[2ex] \textbf{если } f(\hat{x}_t) \geq \tfrac{1}{\sqrt{e}} f(\hat{x}_{t-1}) \textbf{ то } T := t \textbf{ и СТОП}. \end{array}} \qquad (6.1.14)$$

Теорема 6.1.3 *Число точек в процессе (6.1.14) ограничено:*

$$T \leq 1 + 2 \ln \tfrac{1}{\alpha}. \qquad (6.1.15)$$

Последняя из этих точек удовлетворяет неравенству $f(\hat{x}_T) \leq (1+\delta) f^*$. *Общее число градиентных шагов в процессах нижнего уровня (6.1.14) не превосходит*

$$\tfrac{e}{\alpha^2} \cdot \left(1 + \tfrac{1}{\delta}\right)^2 \cdot \left(1 + 2 \ln \tfrac{1}{\alpha}\right). \qquad (6.1.16)$$

Доказательство.

Пользуясь индукцией, легко доказать, что в начале каждого этапа t в методе (6.1.14) выполняется следующее неравенство:

$$\left(\tfrac{1}{\sqrt{e}}\right)^{t-1} f(x_0) \geq f(\hat{x}_{t-1}), \quad t \geq 1.$$

Таким образом, в силу неравенства (6.1.7) на последнем этапе T этого процесса имеем

$$\left(\tfrac{1}{\sqrt{e}}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \alpha f(x_0).$$

Это дает нам неравенство (6.1.15).

В силу (6.1.9) выполнена оценка $\| x_0 - x^* \| \leq \frac{1}{\gamma_0} f^* \leq \frac{1}{\gamma_0} f(\hat{x}_{T-1})$. Таким образом, на последнем этапе этого процесса, пользуясь (6.1.11) и критерием прерывания (6.1.14), получаем

$$f(\hat{x}_T) - f^* \leq \frac{\gamma_1}{\sqrt{\hat{N}+1}} \cdot \frac{1}{\gamma_0} \cdot f(\hat{x}_{T-1}) \leq \frac{\sqrt{e}}{\alpha\sqrt{\hat{N}+1}} \cdot f(\hat{x}_T) \leq \frac{\delta}{1+\delta} \cdot f(\hat{x}_T). \qquad \Box$$

6.1.4 Минимизация составной функции

В п. 6.1.3 внутренняя и внешняя эллипсоидальная аппроксимация множества $\partial f(0)$ являлась ключевым элементом в построении методов решения задачи (6.1.1) с относительной точностью. Однако для построения евклидовой нормы, которая хорошо подходит к нашей задаче, необходимо каким-то образом учесть ее структуру. В этом пункте мы рассматриваем модель задачи (6.1.1), в которой возможно как явное указание такой нормы, так и применение техники сглаживания из параграфа 5.1. Заметим, что оценки эффективности этого подхода существенно превосходят оценки субградиентного метода.

Поскольку целевая функция $f(x)$ в задаче (6.1.1) однородна, в простейших случаях она может иметь следующую структуру. Предположим, что функция $f(x)$ представлена в виде суперпозиции двух функций, линейной функции $A(x)$ и *простой* нелинейной выпуклой однородной функции $F(v)$. Таким образом, $f(x) = F(A(x))$. Приведем формальное определение этого объекта. В этом пункте мы используем некоторые обозначения из параграфа 5.1.

Пусть выпуклое замкнутое множество Q_2 из $\mathbb{R}^m$ ограничено и содержит нуль как внутреннюю точку. Зададим выпуклую однородную функцию $F(v)$ следующим образом:

$$F(v) = \max_u \{\langle v, u\rangle :\ u \in Q_2\}. \tag{6.1.17}$$

Далее, пусть матрица A размером $m \times n$ имеет полный столбцовый ранг (таким образом, $m \geq n$). Определим нашу целевую функцию следующим образом:

$$f(x) = F(Ax), \quad x \in \mathbb{R}^n. \tag{6.1.18}$$

Ясно, что $f(x)$ – выпуклая однородная функция со степенью однородности единица. Мы собираемся решать задачу (6.1.1), которую для удобства переписываем здесь:

$$\text{найти } f^* = \min_{x \in \mathbb{R}^n} \{f(x) :\ Cx = b\}. \tag{6.1.19}$$

Поскольку $\partial F(0) \equiv Q_2$, имеем $\partial f(0) = A^T Q_2$. Таким образом, задача (6.1.19) удовлетворяет основному предположению (6.1.2).

Обозначим через $\|\cdot\|_2$ стандартную евклидову норму в $\mathbb{R}^m$:

$$\|u\|_2 = \left[\sum_{i=1}^m \left(u^{(i)}\right)^2\right]^{1/2}, \quad u \in \mathbb{R}^m.$$

Введем следующие характеристики функции F:

$$\begin{aligned}
\gamma_0(F) &= \max_{r>0}\{r :\ B_{\|\cdot\|_2}(r) \subseteq \partial F(0)\}, \\
\gamma_1(F) &= \min_{r>0}\{r :\ B_{\|\cdot\|_2}(r) \supseteq \partial F(0)\}, \\
\alpha(F) &= \tfrac{\gamma_1(F)}{\gamma_0(F)} \geq 1.
\end{aligned}$$

Для множеств из примера 6.1.1 эти значения выглядят так:

$$\begin{array}{lll} \gamma_0(F_1) = \frac{1}{\sqrt{m}}, & \gamma_1(F_1) = 1, & \alpha(F_1) = \sqrt{m}, \\ \gamma_0(F_\infty) = 1, & \gamma_1(F_\infty) = \sqrt{m}, & \alpha(F_\infty) = \sqrt{m}. \end{array} \tag{6.1.20}$$

Определим теперь специальную евклидову норму в прямом пространстве:

$$\|x\|_1 = \|Ax\|_2^*, \quad x \in \mathbb{R}^n.$$

Поскольку A – невырожденная матрица, эта норма корректно определена. Полагая $G = A^T A$, получаем следующее представление:

$$\begin{array}{rcl} \|x\|_1 & = & \langle Gx, x\rangle^{1/2} = \left[\sum\limits_{i=1}^{m} \langle a_i, x\rangle^2\right]^{1/2}, \\ \|g\|_1^* & = & \langle g, G^{-1}g\rangle^{1/2}, \end{array} \tag{6.1.21}$$

где через a_i, $i = 1, \ldots, m$, обозначены столбцы матрицы A^T.

Лемма 6.1.2 *Для нормы* $\|\cdot\|_1$, *условие (6.1.3) выполняется с параметрами*

$$\gamma_0 = \gamma_0(F), \quad \gamma_1 = \gamma_1(F).$$

Таким образом, можно взять $\alpha = \alpha(F) = \frac{\gamma_0(F)}{\gamma_1(F)}$.

Доказательство.
Поскольку $\partial f(0) = A^T Q_2$, опорная функция этого множества может быть представлена следующим образом:

$$\xi(x) \overset{\text{def}}{=} \max_{s \in \partial f(0)} \langle s, x\rangle = \max_{u \in Q_2} \langle A^T u, x\rangle = \max_{u \in Q_2} \langle Ax, u\rangle.$$

Следовательно,

$$\begin{array}{rcl} \xi(x) & \leq & \max\limits_{\|u\|_2 \leq \gamma_1(F)} \langle Ax, u\rangle = \gamma_1(F)\|Ax\|_2^* = \gamma_1(F)\|x\|_1, \\ \xi(x) & \geq & \max\limits_{\|u\|_2 \leq \gamma_0(F)} \langle Ax, u\rangle = \gamma_0(F)\|Ax\|_2^* = \gamma_0(F)\|x\|_1. \end{array}$$

Поэтому $\partial f(0) \subseteq B_{\|\cdot\|_1^*}(\gamma_1(F))$, и $\partial f(0) \supseteq B_{\|\cdot\|_1^*}(\gamma_0(F))$. □

Заметим, что для многих простых множеств параметры $\gamma_1(F)$ и $\gamma_0(F)$ легко оцениваются (см., например, (6.1.20)). Таким образом, метрика (6.1.21) может использоваться для решения соответствующих задач субградиентным методом (6.1.14). Однако основное преимущество представления (6.1.18) заключается в том, что в него легко вставляется техника сглаживания из параграфа 5.1. Покажем как это можно сделать.

По сравнению с постановкой (5.1.3), в задаче (6.1.19) есть лишь одно существенное отличие – наличие неограниченного допустимого множества. Таким образом, непосредственное применение техники сглаживания к задаче (6.1.19) невозможно. Однако мы можем ввести искусственное ограничение, пользуясь неравенством (6.1.9). Обозначим

$$Q_1(R) = \{x \in \mathbb{R}^n : \ Cx = b, \ \|x - x_0\|_1 \leq R\}.$$

В силу (6.1.9) имеем $x^* \in Q_1(\hat{\rho})$, где $\hat{\rho} = \frac{1}{\gamma_0(F)} f(x_0)$. Таким образом, задача (6.1.19) эквивалентна следующей:

$$\begin{aligned} \text{найти } f^* &= \min_{x \in \mathbb{R}^n} \{f(x) : \ x \in Q_1(\hat{\rho})\} \\ &= \min_{x \in Q_1(\hat{\rho})} \max_{u \in Q_2} \langle Ax, u \rangle \\ &= \max_{u \in \mathbb{R}^m} \{\phi_{\hat{\rho}}(u) : \ u \in Q_2\}, \end{aligned} \tag{6.1.22}$$

где $\phi_R(u) = \min\limits_{x \in Q_1(R)} \langle Ax, u \rangle$. Таким образом, нам удалось представит нашу задачу в форме, необходимой для применения результатов из параграфа 5.1.

Введем объекты, необходимые для применения техники сглаживания. В прямом пространстве выберем прокс-функцию $d_1(x) = \frac{1}{2}\|x - x_0\|_1^2$ с параметром выпуклости $\sigma_1 = 1$. Ее максимальное значение на допустимом множестве $Q_1(\hat{\rho})$ равно $D_1 = \frac{1}{2}\hat{\rho}^2$.

Аналогично, для двойственного множества выбирается $d_2(u) = \frac{1}{2}\|u\|_2^2$ с параметром $\sigma_2 = 1$ и максимальным значением на двойственном допустимом множестве Q_2 равным $D_2 = \frac{1}{2}\gamma_1^2(F)$. Остается заметить, что

$$\begin{aligned} \|A\|_{1,2} &= \max_{x,u} \{\langle Ax, u \rangle : \ \|x\|_1 \leq 1, \ \|u\|_2 \leq 1\} = \max_x \{\|Ax\|_2^* : \ \|x\|_1 \leq 1\} \\ &= \max_x \{\|x\|_1 : \ \|x\|_1 \leq 1\} = 1. \end{aligned} \tag{6.1.23}$$

Приведем здесь для удобства версию метода (1.2.18), приспособленную к нашей ситуации. Этот метод будет применяться для минимизации гладкой аппроксимации целевой функции $f(x)$:

$$f_\mu(x) = \max_u \{\langle Ax, u \rangle - \mu d_2(u) : \ u \in Q_2\}. \tag{6.1.24}$$

Заметим, что эта функция имеет липшицев градиент

$$\nabla f_\mu(x) = A^T u_\mu(x),$$

где $u_\mu(x)$ - это единственное решение оптимизационной задачи (6.1.24). В силу формулы

(6.1.23) константа Липшица для этого градиента равна $\frac{1}{\mu}$ (см. теорему 5.1.1).

```
Метод $S_N(R)$

Полагаем $\mu = \frac{2R}{\gamma_1(F)\cdot(N+1)}$.

Для $k = 0 \dots N$ повторяем
    $u_\mu(x_k) = \arg\max_u \{\langle Ax_k, u\rangle - \frac{\mu}{2}\|u\|_2^2 : \ u \in Q_2\}$,
    $y_k = \arg\min_x \{\langle A(x - x_k), u_\mu(x_k)\rangle + \frac{1}{2\mu}\|x - x_k\|_1^2 : \ x \in Q_1(R)\}$,        (6.1.25)
    $z_k = \arg\min_x \left\{\frac{1}{2\mu}\|x - x_0\|_1^2 + \langle A(x - x_0), \sum_{i=0}^{k} \frac{i+1}{2} u_\mu(x_i)\rangle : \ x \in Q_1(R)\right\}$,
    $x_{k+1} = \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$.

Результат   $\bar{x} := y_N$.
```

В дальнейшем мы обозначаем результат $\bar{x}$ работы этого метода через $S_N(R)$. Легко видеть, что все условия теоремы 5.1.2 выполнены. Таким образом, если $\|x_0 - x^*\|_1 \le R$, то результат работы этого метода удовлетворяет неравенству

$$f(S_N(R)) - f^* \le \frac{2\gamma_1(F)R}{N+1}. \tag{6.1.26}$$

Из этого наблюдения вытекает важное следствие.

Теорема 6.1.4 *Для* $\delta \in (0, 1)$, *положим*

$$N = \left\lfloor \frac{2}{\alpha^2(F)\,\delta} \right\rfloor. \tag{6.1.27}$$

Тогда $f\left(S_N\left(\frac{1}{\gamma_0(F)} f(x_0)\right)\right) \le (1 + \delta) f^*$.

Доказательство.

Поскольку в силу неравенства (6.1.9) $\|x_0 - x^*\|_1 \le \frac{1}{\gamma_0(F)} f(x_0)$ и в силу выбора (6.1.27) $N + 1 \ge \frac{2}{\alpha^2(F)\,\delta}$, из неравенств (6.1.26) и (6.1.7) следует, что

$$f(S_N(R)) - f^* \le \delta \cdot \alpha(F) f(x_0) \le \delta \cdot f^*. \qquad \square$$

Заметим, что оценка эффективности (6.1.27) метода (6.1.25) лучше, чем оценка субградиентного метода (6.1.14) с рекурсивной настройкой расстояния до оптимальной точки. Покажем, что аналогичный прием ускоряет также и метод (6.1.25).

Обозначим через $\delta \in (0,1)$ требуемую относительную точность. Положим

$$\tilde{N} = \left\lfloor \tfrac{2e}{\alpha(F)} \cdot \left(1 + \tfrac{1}{\delta}\right) \right\rfloor .$$

Рассмотрим следующий процесс. Положим $\hat{x}_0 = x_0$. Для $t \geq 1$ повторяем следующие операции:

$$\boxed{\begin{array}{l} \hat{x}_t := S_{\tilde{N}} \left(\frac{1}{\gamma_0(F)} f(\hat{x}_{t-1}) \right); \\ \\ \textbf{если } f(\hat{x}_t) \geq \frac{1}{e} f(\hat{x}_{t-1}) \textbf{ то } T := t \textbf{ и СТОП}. \end{array}} \qquad (6.1.28)$$

Теорема 6.1.5 *Число точек T в процессе (6.1.28) ограничено:*

$$T \leq 1 + \ln \tfrac{1}{\alpha(F)}. \qquad (6.1.29)$$

Последняя сформированная точка удовлетворяет неравенству $f(\hat{x}_T) \leq (1+\delta) f^*$. Общее число шагов в процессах (6.1.28) нижнего уровня не будет больше чем

$$\tfrac{2e}{\alpha(F)} \cdot \left(1 + \tfrac{1}{\delta}\right) \cdot \left(1 + \ln \tfrac{1}{\alpha(F)}\right). \qquad (6.1.30)$$

Доказательство.

С помощью индукции нетрудно доказать, что в начале любого этапа t выполняется следующее неравенство:

$$\left(\tfrac{1}{e}\right)^{t-1} f(x_0) \geq f(\hat{x}_{t-1}), \quad t \geq 1.$$

Таким образом, в силу пункта 1 теоремы 6.1.1 на последнем этапе T этого процесса имеем

$$\left(\tfrac{1}{e}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \alpha(F) f(x_0),$$

откуда получаем неравенство (6.1.29).

Заметим, что $\| x_0 - x^* \| \leq \frac{1}{\gamma_0(F)} f^* \leq \frac{1}{\gamma_0(F)} f(\hat{x}_{T-1})$. Таким образом, на последней стадии этого процесса, в силу неравенства (6.1.26) и критерия остановки в (6.1.28) получаем

$$f(\hat{x}_T) - f^* \;\leq\; \tfrac{2\gamma_1(F)}{\tilde{N}+1} \cdot \tfrac{1}{\gamma_0(F)} \cdot f(\hat{x}_{T-1}) \leq \tfrac{2e}{\alpha(F)\cdot(\tilde{N}+1)} \cdot f(\hat{x}_T) \leq \tfrac{\delta}{1+\delta} \cdot f(\hat{x}_T). \qquad \Box$$

6.1.5 Примеры приложений

Обсудим оценки эффективности метода предложенного в п. 6.1.4 и применяемого к задачам оптимизации с разной структурой.

Линейное программирвание

Обозначим через $\hat{A}$ матрицу размером $m \times (n-1)$, $m \geq n$, имеющую полный столбцовый ранг. Для заданного вектора $c \in \mathbb{R}^m$ рассмотрим следующую задачу:

$$\text{найти } f^* = \max_{u \in \mathbb{R}^m} \{\langle c, u \rangle : \ \hat{A}^T u = 0, \ |u^{(i)}| \leq 1, \ i = 1, \ldots, m\}. \tag{6.1.31}$$

Эта задача нетривиальна только если ранг расширенной матрицы $A = (\hat{A}, c)$ равен n, что мы и предположим.

Задача (6.1.31) может быть переписана в двойственной форме. Положим

$$\phi_\infty(y) = \max_{u \in \mathbb{R}^m} \{\langle c, u \rangle + \langle y, \hat{A}^T u \rangle : \ |u^{(i)}| \leq 1, \ i = 1, \ldots, m\} = \sum_{i=1}^m |\langle a_i, y \rangle + c_i|,$$

где через a_i обозначены столбцы матрицы $\hat{A}^T$. Тогда

$$f^* = \min_{y \in \mathbb{R}^{n-1}} \phi_\infty(y).$$

В примере 6.1.1 мы уже видели, что последняя задача может быть представлена в виде (6.1.18)-(6.1.19) с $x = (y^T, \tau)^T$, и $F_\infty(v) = \sum\limits_{i=1}^m |v^{(i)}|$. Таким образом,

$$Q_2 = \{u \in \mathbb{R}^m : \ |u^{(i)}| \leq 1, \ i = 1, \ldots, m\}.$$

Выбирая $\|u\|_2 = \left[\sum\limits_{i=1}^m (u^{(i)})^2\right]^{1/2}$, получаем

$$\gamma_0(F_\infty) = 1, \quad \gamma_1(F_\infty) = \sqrt{m}, \quad \alpha(F_\infty) = \tfrac{1}{\sqrt{m}}.$$

Таким образом, в силу теоремы 6.1.5 для того, чтобы найти f^* с относительной точностью $\delta \in (0,1)$, нам потребуется не более

$$2e \cdot m^{1/2} \cdot \left(1 + \tfrac{1}{2} \ln m\right) \cdot \left(1 + \tfrac{1}{\delta}\right)$$

итераций метода $S_N(R)$.

В этом методе необходимо вычислить и обратить матрицу $G = A^T A$. Если матрица A сильно заполнена, для этого потребуется $O(n^2 m)$ операций. Далее, каждая итерация метода $S_N(R)$ требует порядка $O(nm)$ операций. Действительно,

- умножение матрицы A на вектор x_k потребует $O(mn)$ операций;
- поскольку множество Q_2 и норма $\|u\|_2$ имеют сепарабельную структуру, для вычисления точки $u_\mu(x_k)$ необходимо $O(m)$ операций;
- вычисление точек y_k и z_k требует два умножения матрицы A^T на вектор и нахождения двух проекций на множество

  $$Q_1(R) = \{x : \ Cx = 1, \ \|x\|_1 \leq R\}$$

  в евклидовой метрике $\|\cdot\|_1$. Поскольку $C \in \mathbb{R}^{1 \times n}$, такая проекция может быть найдена по формуле.

Таким образом, общее количество вычислений в этой схеме составит

$$O\left(n^2 m + \tfrac{1}{\delta} \cdot n m^{1.5} \ln m\right) \tag{6.1.32}$$

операций. Первый член в этой сумме доминирует при $\delta > \frac{\sqrt{m}}{n} \ln m$.

Заметим, что для решения задачи (6.1.31) можно применить стандартный метод отслеживания траектории (см. раздел 1.3.3). Каждая итерация этого метода требует $O(n^2 m)$ операций. Таким образом, ее оценка эффективности выглядит так:

$$O\left(n^2 m^{1.5} \ln \tfrac{m}{\delta}\right) \tag{6.1.33}$$

Другая возможность состоит в решении этой задачи методом эллипсоидов [79]. В этом случае оценка эффективности составит

$$O\left(n^3 m \ln \tfrac{m}{\delta}\right). \tag{6.1.34}$$

операций. Сравнивая оценки (6.1.32), (6.1.33), и (6.1.34) заключаем, что метод (6.1.28) является наилучшим если требуемая относительная точность δ не очень мала, скажем

$$\delta > O\left(\tfrac{1}{n} \max\left\{1, \tfrac{\sqrt{m}}{n}\right\}\right).$$

Минимизация спектрального радиуса

Обозначим через S^n пространство симметрических $n \times n$-матриц. Для матрицы $X \in S^n$ рассмотрим ее спектральный радиус:

$$\rho(X) = \max_{1 \le i \le n} |\lambda_i(X)|.$$

Заметим, что эта функция выпукла на S^n. Для вектора переменных $x \in \mathbb{R}^p$ введем линейный оператор $A(x)$:

$$A(x) = \sum_{i=1}^{p} x^{(i)} A_i \in S^n.$$

Теперь можно определить целевую функцию задачи (6.1.19):

$$f(x) = \rho(A(x)). \tag{6.1.35}$$

Предположим также, что линейные ограничения в задаче (6.1.19), (6.1.35) очень простые. Например, это может быть просто одно ограничение $x^{(1)} = 1$.

Для решения задачи (6.1.19), (6.1.35) нам нужно представить функцию верхнего уровня $\rho(X)$ в специальной форме (6.1.17). Пусть

$$Q_2 = \{X \in S^n : \sum_{i=1}^{n} |\lambda_i(X)| \le 1\}.$$

Введем в пространстве S^n стандартную фробениусову норму

$$\|X\|_2 = \langle X, X \rangle^{1/2} = \left[\sum_{i,j=1}^{n} \left(X^{(i,j)}\right)^2\right]^{1/2}.$$

Лемма 6.1.3 *Множество* Q_2 *выпукло и замкнуто, причем*

$$B_{\|\cdot\|_2}(\tfrac{1}{\sqrt{n}}) \subset Q_2 \subset B_{\|\cdot\|_2}(1). \tag{6.1.36}$$

Более того, $\rho(X) = \max\limits_{U\in S^n}\{\langle X,U\rangle :\ U \in Q_2\}$.

Доказательство.

Для любого $X \in S^n$ имеем

$$\begin{aligned}\rho(X) &= \min_{\tau\in\mathbb{R}}\{\tau :\ \tau I_n \succeq X,\ \tau I_n \succeq -X\}\\ &= \min_{\tau\in\mathbb{R}}\max_{Y_1,Y_2\succeq 0}[\tau + \langle X - \tau I_n, Y_1\rangle - \langle X + \tau I_n, Y_2]\\ &= \max_{Y_1,Y_2\succeq 0}\{\langle X, Y_1 - Y_2\rangle :\ \langle I_n, Y_1 + Y_2\rangle = 1\}.\end{aligned}$$

Пусть $U = Y_1 - Y_2$ и $V = Y_1 + Y_2$. Тогда

$$\rho(X) = \max_{U\in S^n}\{\langle X,U\rangle,\ U \in \hat{Q}\},$$

где $\hat{Q} = \{U :\ \exists V \succeq \pm U,\ \langle I_n, V\rangle = 1\}$. Ясно, что множество $\hat{Q}$ замкнуто, выпукло и ограничено. Докажем что $\hat{Q} = Q_2$.

Действительно, представим матрицу U в базисе ее собственных векторов:

$$U = B\Lambda B^T, \quad BB^T = I_n,$$

где Λ – диагональная матрица. Пусть $U \in Q_2$. Определим диагональную матрицу $\hat{\Lambda}$, имеющую следующие диагональные элементы:

$$\hat{\Lambda}^{(i,i)} = |\Lambda^{(i,i)}|/[\sum_{j=1}^{n} |\Lambda^{(j,j)}|], \quad i = 1,\dots,n.$$

Тогда $V = B\hat{\Lambda}B^T \succeq \pm U$ и $\langle I_n, V\rangle = 1$. Таким образом, $Q_2 \subset \hat{Q}$.

Обратно, если $U \in \hat{Q}$, то существует такая матрица $V \in S^n$, что $B^TVB \succeq \pm\Lambda$. Таким образом,

$$\langle Vb_i, b_i\rangle \geq |\Lambda^{(i,i)}|, \quad i = 1,\dots,n,$$

где b_i – столбцы матрицы B. Поэтому

$$1 = \langle I_n, V\rangle = \langle BB^T, V\rangle = \langle I_n, B^TVB\rangle = \sum_{i=1}^{n}\langle Vb_i, b_i\rangle \geq \sum_{i=1}^{n}|\lambda_i(U)|.$$

Таким образом, $\hat{Q} \subseteq Q_2$ и мы заключаем, что $\hat{Q} = Q_2$.

Остается доказать включение (6.1.36). Действительно, неравенство $\|U\|_2^2 \leq \frac{1}{n}$ можно переписать в форме $\sum\limits_{i=1}^{n}\lambda_i^2(U) \leq \frac{1}{n}$. Тогда

$$\sum_{i=1}^{n}|\lambda_i(U)| \leq \sqrt{n}\cdot\left[\sum_{i=1}^{n}\lambda_i^2(U)\right]^{1/2} \leq 1.$$

Обратно, если $\sum\limits_{i=1}^{n} |\lambda_i(U)| \leq 1$, то $\sum\limits_{i=1}^{n} \lambda_i^2(U) \leq \left[\sum\limits_{i=1}^{n} |\lambda_i(U)|\right]^2 \leq 1$. □

Таким образом, в силу включения (6.1.36) имеем

$$\gamma_0(\rho) = \tfrac{1}{\sqrt{n}}, \quad \gamma_1(\rho) = 1, \quad \alpha(\rho) = \tfrac{1}{\sqrt{n}}.$$

Поэтому в силу теоремы 6.1.5 общее число итераций метода $S_N(R)$ не превзойдет

$$2e\sqrt{n}\left(1 + \tfrac{1}{2}\ln n\right) \cdot \left(1 + \tfrac{1}{\delta}\right).$$

Для применения нашего подхода нужно вычислить и обратить матрицу G. В рассматриваемой ситуации G является матрицей квадратичной формы

$$\langle Gx, x\rangle = \langle A(x), A(x)\rangle.$$

Таким образом, $G^{(i,j)} = \langle A_i, A_j\rangle$, $i, j = 1, \ldots, p$. Если матрицы A_i сильно заполнены, то вычисление элементов этой матрицы стоит $O(p^2n^2)$ операций, а обращение потребует $O(p^3)$ операций. Поскольку мы предполагаем, что $p < \frac{n(n+1)}{2}$, общая цена предварительных вычислений составит $O(p^2n^2)$ операций.

Далее, наиболее трудоемкими операциями на каждом шаге метода $S_N(R)$ являются следующие.

- Вычисление значения билинейной формы $\langle A(x), U\rangle$ и ее градиента требует $O(pn^2)$ операций.
- Нахождение проекции точки X на множество Q_2 относительно стандартной фробениусовой нормы. Наиболее сложная часть этой операции состоит в нахождении базиса из собственных векторов матрицы X. Это может быть сделано за $O(n^3)$ операций.
- Общая стоимость всех операций, производимых в пространстве $\mathbb{R}^p$ не превосходит $O(p^2)$ операций.

Таким образом, сложность каждой итерации метода $S_N(R)$ не превосходит $O(n^2(n+p))$ операций. Поэтому всего в методе (6.1.28) требуется

$$O\left(n^2p^2 + \tfrac{1}{\delta} \cdot n^{2.5}(p+n)\ln n\right) \tag{6.1.37}$$

арифметических операций.

Приведем оценку эффективности метода отслеживания траектории применительно к задаче (6.1.19)-(6.1.35). Для этого метода наиболее трудоемкой операцией на каждой итерации является вычисление гессиана барьерной функции. Его элементы задаются формулами

$$\langle X^{-1}A_iX^{-1}, A_j\rangle, \quad i, j = 1, \ldots, p.$$

Поэтому требуется выполнить $O(pn^2(p+n))$ операций. Таким образом, общая оценка эффективности метода отслеживания траектории имеет порядок

$$O\left(pn^{2.5}(p+n)\ln\frac{n}{\delta}\right)$$

операций. Сравнивая ее с оценкой (6.1.37), мы видим, что градиентный метод лучше если относительная точность не слишком мала:

$$\delta \geq O\left(\tfrac{1}{p}\right).$$

Расчет оптимальных механических конструкций

В этой задаче (см., например, [39]), имеется множество точек

$$x_i \in \mathbb{R}^2, \quad i = 1, \dots, n+p,$$

соединенных множеством дуг (i_k, j_k), $k = 1, \dots, m$. Мы всегда предполагаем, что $j_k > i_k$. Каждая дуга имеет неотрицательный вес $t^{(k)}$, сумма которых равна единице. Точки $x_{n+1}, \dots, x_{n+p}$ зафиксированы. К другим точкам приложены внешние силы

$$f_i \in \mathbb{R}^2, \quad i = 1, \dots, n, \quad f \stackrel{\mathrm{def}}{=} (f_1, \dots, f_n)^T \in \mathbb{R}^{2n}$$

(ненулевых сил обычно немного).

Нашей целью является нахождение оптимального вектора весов

$$t \stackrel{\mathrm{def}}{=} (t^{(1)}, \dots, t^{(m)})^T \in \Delta_m \equiv \left\{ t \in \mathbb{R}^m_+ : \sum_{i=1}^m t^{(i)} = 1 \right\},$$

который минимизирует общую энергию $\psi(t)$, накапливаемую в конструкции.

Для определения этой энергии, предположим сначала что $i_k < n$, $k = 1, \dots, m$. Для каждой дуги k, положим

$$d_k = \tfrac{x_{i_k} - x_{j_k}}{\|x_{i_k} - x_{j_k}\|^2} \in \mathbb{R}^2, \quad k = 1, \dots, m,$$

где $\|\cdot\|$ – стандартная евклидова норма в $\mathbb{R}^2$. Определим теперь вектор $a_k = (a_{k,1}, \dots, a_{k,n})^T \in \mathbb{R}^{2n}$, составленный из следующих двумерных векторов:

$$a_{k,q} = \begin{cases} d_k, & \text{если } q = i_k, \\ -d_k, & \text{если } q = j_k \text{ и } j_k \leq n, \\ 0, & \text{в противном случае.} \end{cases} \qquad q = 1, \dots, n.$$

Пусть $B(t) = \sum\limits_{k=1}^m t^{(k)} a_k a_k^T$. Тогда задача расчета механических конструкций записывается так:

$$\text{найти } \psi^* = \inf_t \{ \langle [B(t)]^{-1} f, f \rangle : \ t \in \operatorname{rint} \Delta_m \}. \tag{6.1.38}$$

Эта задача корректно определена тогда и только тогда, когда матрица $G \stackrel{\mathrm{def}}{=} B(\bar{e}_m)$ положительно определена.

Покажем теперь как записать эту задачу в форме (6.1.18)-(6.1.19). Заметим, что

$$\begin{aligned}
\psi^* &= \inf_{t \in \operatorname{rint} \Delta_m} \langle [B(t)]^{-1} f, f \rangle \\
&= \inf_{t \in \operatorname{rint} \Delta_m} \max_{x \in \mathbb{R}^{2n}} [2\langle f, x \rangle - \langle B(t)x, x \rangle] \\
&= \max_{x \in \mathbb{R}^{2n}} \inf_{t \in \operatorname{rint} \Delta_m} \left[2\langle f, x \rangle - \sum_{k=1}^{m} t^{(k)} \langle a_k, x \rangle^2 \right] \\
&= \max_{x \in \mathbb{R}^{2n}} \left[2\langle f, x \rangle - \max_{1 \le k \le m} \langle a_k, x \rangle^2 \right] \\
&= \max_{x \in \mathbb{R}^{2n}} \left[\langle f, x \rangle^2 \cdot \left(\max_{1 \le k \le m} |\langle a_k, x \rangle| \right)^{-2} \right]
\end{aligned}$$

(на последнем шаге мы максимизируем целевую функцию вдоль направления x).

Таким образом, мы получили задачу

$$\text{найти } f^* = \min_{x \in \mathbb{R}^{2n}} \{ f(x) \stackrel{\mathrm{def}}{=} \max_{1 \le k \le m} |\langle a_k, x \rangle| : \ \langle f, x \rangle = 1 \}, \tag{6.1.39}$$

которая как раз имеет нужную нам форму (6.1.18)–(6.1.19). Пусть A – $m \times (2n)$-матрица со строками a_k^T. Тогда, пользуясь обозначениями примера 6.1.1, целевая функция этой задачи может быть записана так:

$$f(x) = F_1(Ax).$$

В силу (6.1.20) имеем $\alpha(F_1) = \frac{1}{\sqrt{m}}$. Таким образом, для того чтобы найти решение задачи (6.1.39) с относительной точностью δ, методу (6.1.28) нужно не более

$$2e\sqrt{m} \left(1 + \tfrac{1}{2} \ln m\right) \cdot \left(1 + \tfrac{1}{\delta}\right) \tag{6.1.40}$$

итераций схемы $S_N(R)$. Наиболее трудными операциями на каждой итерации этой схемы являются следующие.

- Вычисление значения и градиента билинейной формы $\langle Ax, u \rangle$, которое требует $O(m)$ операций (напомним что матрица A разрежена).
- Евклидова проекция на множество $Q_2 \subset \mathbb{R}^m$, требующая $O(m \ln m)$ операций.
- Все шаги в прямом пространстве требуют в сумме $O(n^2)$ операций.

Заметим, что предварительное вычисление матрицы G требует $O(m+n^2)$ операций, а для ее обращения их нужно $O(n^3)$. Поскольку $m \le \frac{n(n+1)}{2}$, мы приходим к следующей оценке вычислительных затрат метода (6.1.28):

$$O\left(n^3 + \frac{1}{\delta} \cdot (n^2 + m \ln m) \cdot \sqrt{m} \ln m\right) \tag{6.1.41}$$

операций. Для сильно заполненной структуры с $m = O(n^2)$ эта оценка переходит в следующую:

$$O\left(\frac{n^3}{\delta} \ln^2 n\right).$$

6.2 Эллипсоидальная аппроксимация выпуклых тел

6.2.1 Введение

Мотивировка

Среди современных методов решения задач линейного программирования (ЛП), методы внутренней точки (МВТ) рассматриваются как наиболее эффективные. Однако эти методы требуют очень много вспомогательных вычислений. Для ЛП-задач размера $m \times n$, $(m > n)$, они находят решение задачи с *абсолютной* точностью ϵ за

$$O(\sqrt{m} \ln \tfrac{m}{\epsilon})$$

итераций метода Ньютона. Заметим, что для задачи с сильно заполненной матрицей ограничений на каждой такой итерации потребуется выполнить $O(n^2 m)$ операций.

Понятно, что такая трудоемкость итерации дает большие шансы на успех градиентным методам, у которых каждая итерация намного дешевле. Однако основным недостатком градиентных методов является их медленная сходимость. Субградиентный метод для общих выпуклых задач требует $O(\frac{C_0}{\epsilon^2})$ итераций для нахождения решения с точностью ϵ. В этой оценке сильная зависимость от точности усугубляется присутствием некоторой константы C_0, которая зависит от нормы матрицы ограничений, размера решения, качества стартовой точки, и т.д. Эта константа может оказаться неожиданно большой. Поэтому в настоящее время градиентные методы составляют конкуренцию МВТ только на очень больших задачах.

Однако в гл. 5 мы показали, что правильное использование структуры ЛП-задач может дать алгоритмы градиентного типа с оценкой эффективности $O(\frac{C_1}{\epsilon})$ итераций. Более того, было показано что для некоторых задач константу C_1 можно найти в явном виде и что она достаточно мала. В параграфе 6.1 этот результат был распространен на методы решения некоторых задач минимизации с относительной точностью. А именно, было показано, что решение некоторых ЛП-задачи с относительной точностью δ может быть найдено за $O(\frac{\sqrt{m}}{\delta})$ итераций градиентного типа. Заметим, что во многих приложениях концепция относительной точности очень привлекательна, так как точность решения автоматически подстраивается к любому размеру оптимального решения. Таким образом, в этом случае не надо опасаться больших неизвестных констант. Более того, во многих случаях большая относительная точность просто не нужна - порядок в $0.5\% - 0.1\%$ часто вполне достаточен.

Подход параграфа 6.1 применим к специальным *коническим* задачам безусловной минимизации. Это задачи минимизации выпуклой неотрицательной однородной функции $f(x)$, $x \in \mathbb{R}^n$, на аффинном подпространстве не содержащем нуля. Для решения этой задачи с относительной точностью нужно знать пару подобных концентричных аппроксимирующих эллипсоидов (*эллипсоидов Джона*) для субдифференциала целевой функции в нуле. Было показано, что для некоторых ЛП-задач можно легко указать такие эллипсоиды с отношением радиусов порядка $O(\sqrt{m})$.

Однако хорошо известно, что для любого центрально симметричного тела в $\mathbb{R}^n$ существует пара эллипсоидов Джона с отношением радиусов $\sqrt{n}$. Более того, легко найти пару эллипсоидов с близким отношением радиусов (см. работы [63] и [27]). Оказывается, что такие эллипсоиды дают нам хорошую евклидову норму, позволяющую быстро находить решение соответствующей ЛП-задачи с нужной относительной точностью. В этом разделе мы рассматриваем два нетривиальных класса ЛП-задач, для которых решение с относительной точностью δ находится за $O(\frac{\sqrt{n \ln m}}{\delta} \ln n)$ итераций градиентного типа.

Заметим, что объем предварительных вычислений для обоих классов задач вполне приемлем: он находится на уровне $O(n^2 m \ln m)$ операций. С точностью до логарифмического фактора, эта оценка совпадает с трудоемкостью нахождения проекции точки в $\mathbb{R}^m$ на линейное подпространство, заданное системой из n уравнений. Последующий процесс минимизации как правило оказывается более дешевым. Действительно, одна итерация градиентного типа требует $O(nm)$ операций. Поэтому, всему методу нужно порядка $O(\frac{\sqrt{n \ln m}}{\delta} \cdot nm \cdot \ln n)$ операций.

Содержание

В п. 6.2.2 мы описываем эффективный алгоритм для вычисления эллипсоидов Джона для различных типов выпуклых множеств: центрально-симметричных множеств (п. 6.2.2A), множеств общего вида п. 6.2.2B), и знако-симметричных множеств (п. 6.2.2C). Результаты раздедов 6.2.2A и 6.2.2B могут быть найдены в работе [63]; однако мы приводим для них новые геометрические доказательства. В любом случае, вычисление эллипсоидов Джона нужного качества требует $O(n^2 m \ln m)$ операций. Очень важно, что для знако-симметричных множеств матрица эллипсоидов Джона может быть выбрана диагональной. В п. 6.2.3 мы рассматриваем задачу минимизации максимального модуля линейных форм на аффинном подпространстве. Задачи такого вида возникают в регрессионном анализе, при проектировании механических конструкций, и т.д. Мы показываем, что подходящая евклидова метрика (соответствующая приближенному эллипсоиду Джона для субдифференциала целевой функции) приводит к оценке эффективности в $O(\frac{\sqrt{n \ln m}}{\delta} \ln n)$ итераций градиентного типа. В п. 6.2.4 мы рассматриваем задачу нахождения решения билинейной матричной игры, у которой матрица имеет неотрицательные коэффициенты (*задача упаковки*). Мы показываем, что после диагонального масштабирования (которое может быть эффективно найдено с помощью техники представленной в п. 6.2.2), эта задача может быть решена за $O(\frac{\sqrt{n \ln m}}{\delta} \ln n)$ итераций градиентного типа.

Обозначения

Любой самосопряженный положительно определенный оператор G задает евклидову норму на E:

$$\|x\|_G = \langle Gx, x \rangle^{1/2}, \quad x \in E.$$

Тогда двойственной нормой будет

$$\|s\|_G^* = \sup_x \{ \langle s, x \rangle : \ \|x\|_G \le 1 \} = \langle s, G^{-1} s \rangle^{1/2}, \quad s \in E^*.$$

Для выпуклого замкнутого ограниченного множества $C \subset E^*$, обозначим через $\xi_C(x)$ его опорную функцию:

$$\xi_C(x) = \max_{s \in C} \langle s, x \rangle, \quad x \in E.$$

Таким образом, $\partial \xi_C(0) = C$.

Наконец, для $g \in E^*$, обозначим через gg^* следующий линейный оператор:

$$(gg^*)(x) = \langle g, x \rangle \cdot g \in E^*, \quad x \in E.$$

В координатном представлении, $D(a)$ обозначает диагональную $n \times n$-матрицу с вектором $a \in \mathbb{R}^n$ на диагонали. Обозначим через $e_k \in \mathbb{R}^n$ k-й базисный вектор, а через $\bar{e}_n \in \mathbb{R}^n$ – вектор из всех единиц. Таким образом, $I_n \equiv D(\bar{e}_n)$. Обозначение $\mathbb{R}^n_+$ используется для положительного ортанта, а $\Delta_n \equiv \{x \in \mathbb{R}^n_+ : \ \langle \bar{e}_n, x \rangle = 1\}$ – для стандартного симплекса.

6.2.2 Вычисление эллипсоидов Джона

В этом пункте мы приведем эффективные алгоритмы построения эллипсоидов Джона для различных типов выпуклых множеств. Всюду ниже эллипсоиды $W_r(v, G) \subset E^*$ представляются в следующей форме:

$$W_r(v, G) = \{ s \in E^* : \ \|s - v\|_G^* \equiv \langle s - v, G^{-1}(s - v) \rangle^{1/2} \le r \},$$

где $G \succ 0$ – линейный оператор из E в E^*. Если $v = 0$, то мы обычно переходим к обозначению $W_r(G)$. Эллипсоид $W_1(v, G)$ называется β-*аппроксимацией* выпуклого множества $C \subset E^*$, $\beta \ge 1$, если выполнены включения

$$W_1(v, G) \subseteq C \subseteq W_\beta(v, G).$$

Параметр β называется *радиусом* эллипсоидальной аппроксимации.

А. Центрально симметричные выпуклые множества

Для произвольного вектора $g \in E^*$ рассмотрим множество $C_{\pm g}(G) = \mathrm{Conv}\,\{W_1(G), \pm g\}$. Для $\alpha \in [0,1]$ положим

$$G(\alpha) = (1-\alpha)G + \alpha g g^*.$$

Лемма 6.2.1 *При любом $\alpha \in [0,1)$ выполнено включение*

$$W_1(G(\alpha)) \subset C_{\pm g}(G). \tag{6.2.1}$$

Для $\sigma \overset{\mathrm{def}}{=} \frac{1}{n}(\|g\|_G^*)^2 - 1 > 0$ функция

$$V(\alpha) \overset{\mathrm{def}}{=} \ln \tfrac{\det G(\alpha)}{\det G(0)} = \ln(1 + \alpha(n(1+\sigma)-1)) + (n-1)\ln(1-\alpha),$$

принимает максимальное значение в точке $\alpha^* = \frac{\sigma}{n(1+\sigma)-1}$. Более того,

$$V(\alpha^*) = \ln(1+\sigma) + (n-1)\ln \tfrac{(n-1)(1+\sigma)}{n(1+\sigma)-1} \geq \ln(1+\sigma) - \tfrac{\sigma}{1+\sigma} \geq \tfrac{\sigma^2}{2(1+\sigma)^2}. \tag{6.2.2}$$

Доказательство.

Для любого $x \in E$ имеем

$$\begin{aligned} \xi_{W_1(G(\alpha))}(x) &= \langle G(\alpha)x, x\rangle^{1/2} = [(1-\alpha)\langle Gx, x\rangle + \alpha\langle g, x\rangle^2]^{1/2} \\ &\leq \max\{\langle Gx, x\rangle^{1/2}, |\langle g, x\rangle|\} \\ &= \max\{\xi_{W_1(G)}(x), \xi_{\mathrm{Conv}\,\{\pm g\}}(x)\} = \xi_{C_{\pm g}(G)}(x). \end{aligned}$$

Поэтому выполняется включение (6.2.1). Далее,

$$\begin{aligned} V(\alpha) &= \ln\det(G^{-1/2}G(\alpha)G^{-1/2}) \\ &= \ln\det\left((1-\alpha)I_n + \alpha G^{-1/2} g g^* G^{-1/2}\right) \\ &= \ln(1 - \alpha + \alpha(\|g\|_G^*)^2) + (n-1)\ln(1-\alpha) \\ &= \ln(1 + \alpha(n(1+\sigma)-1)) + (n-1)\ln(1-\alpha). \end{aligned}$$

Следовательно, условие оптимальности для функции $V(\alpha)$ записывается так:

$$\tfrac{n-1}{1-\alpha} = \tfrac{n(1+\sigma)-1}{1+\alpha(n(1+\sigma)-1)}.$$

Решением этого уравнения является $\alpha^* = \frac{\sigma}{n(1+\sigma)-1}$. Заметим, что

$$\begin{aligned} V(\alpha^*) &= \ln(1+\sigma) + (n-1)\ln \tfrac{(n-1)(1+\sigma)}{n(1+\sigma)-1} \\ &= \ln(1+\sigma) - (n-1)\ln\left(1 + \tfrac{\sigma}{(n-1)(1+\sigma)}\right) \\ &\geq \ln(1+\sigma) - \tfrac{\sigma}{1+\sigma} \geq \tfrac{\sigma^2}{2(1+\sigma)^2}, \end{aligned}$$

а это и есть неравенство (6.2.2). □

В этом пункте мы решаем следующую задачу. Пусть выпуклое множество C является центрально симметричным, i.e. $\operatorname{int} C \neq \emptyset$, и $x \in C \Leftrightarrow -x \in C$. Для заданного числа $\gamma > 1$ нужно найти эллипсоид Джона для множества C с радиусом не больше $\gamma\sqrt{n}$. Начальная аппроксимация этой задачи дается такой матрицей $G_0 \succ 0$, что $W_1(G_0) \subseteq C$ и $C \subseteq W_R(G_0)$ при некотором $R \geq 1$.

Приведем пример такой задачи.

Пример 6.2.1 Рассмотрим множество векторов $a_i \in E^*$, $i = 1, \ldots, m$, чья линейная оболочка покрывает все двойственное пространство. Определи множество C следующим образом:

$$C = \operatorname{Conv}\{\pm a_i,\ i = 1, \ldots, m\}. \tag{6.2.3}$$

Положим $G_0 = \frac{1}{m}\sum\limits_{i=1}^{m} a_i a_i^*$. Заметим, что для любого $x \in E$ имеем $\xi_C(x) = \max\limits_{1 \leq i \leq m} |\langle a_i, x\rangle|$. Таким образом,

$$\xi_{W_1(G_0)}(x) = \left[\frac{1}{m}\sum_{i=1}^{m}\langle a_i, x\rangle^2\right]^{1/2} \leq \xi_C(x),$$

$$\xi_{W_{m^{1/2}}(G_0)}(x) = m^{1/2}\left[\frac{1}{m}\sum_{i=1}^{m}\langle a_i, x\rangle^2\right]^{1/2} \geq \xi_C(x).$$

Следовательно, $W_1(G_0) \subseteq C \subseteq W_{m^{1/2}}(G_0)$. □

Рассмотрим следующий метод.

```
Для $k \geq 0$ повторяем:

  1. Вычисляем $g_k \in C :\ \|g_k\|^*_{G_k} = r_k \overset{\text{def}}{=} \max\limits_g \{\|g\|^*_{G_k} :\ g \in C\}$.

  2. Если $r_k \leq \gamma n^{1/2}$ то СТОП. Иначе

       $\alpha_k := \frac{1}{n} \cdot \frac{r_k^2 - n}{r_k^2 - 1}, \quad G_{k+1} := (1 - \alpha_k) G_k + \alpha_k g_k g_k^*.$
```
(6.2.4)

Оценка эффективности этого метода дается в следующей теореме.

Теорема 6.2.1 *Пусть $W_1(G_0) \subseteq C \subseteq W_R(G_0)$ для некоторого $R \geq 1$. Тогда число итераций в методе (6.2.4) не превзойдет*

$$\frac{2n \ln R}{2\ln\gamma - 1 + \gamma^{-2}}. \tag{6.2.5}$$

Доказательство.
Заметим, что коэффициент α_k на шаге 2 в методе (6.2.4) выбран как в лемме 6.2.1. Поскольку метод работает пока $\sigma_k \overset{\mathrm{def}}{=} \frac{1}{n} r_k^2 - 1 \geq \gamma^2 - 1$, в силу неравенства (6.2.2) при любом $k \geq 0$ имеем

$$\ln \det G_{k+1} \geq \ln \det G_k + \gamma^{-2} - 1 + 2 \ln \gamma. \tag{6.2.6}$$

Заметим, что для любого $k \geq 0$ выполнено неравенство

$$\begin{aligned} \det(G_k)^{1/2} \cdot \operatorname{vol}_n(W_1(I_n)) &= \operatorname{vol}_n(W_1(G_k)) \leq \operatorname{vol}_n(C) \leq \operatorname{vol}_n(W_R(G_0)) \\ &= \mathbb{R}^n \cdot \det(G_0)^{1/2} \cdot \operatorname{vol}_n(W_1(I_n)). \end{aligned}$$

Поэтому $\ln \det G_k - \ln \det G_0 \leq 2n \ln R$, и мы получаем оценку (6.2.5) суммируя по k неравенства (6.2.6). □

Оценим общую трудоемкость метода (6.2.4) для множества (6.2.3). В этой ситуации лучше рекурсивно пересчитывать обратные матрицы $H_k \overset{\mathrm{def}}{=} G_k^{-1}$ вместе со значениями

$$\nu_k^{(i)} = \langle a_i, H_k a_i \rangle, \quad i = 1, \ldots, m,$$

которые удобно держать в векторе $\nu_k \in \mathbb{R}^m$. Модифицированный вариант метода (6.2.4) выглядит так.

A. Вычисляем $H_0 = \left[\frac{1}{m} \sum\limits_{i=1}^{m} a_i a_i^* \right]^{-1}$ и вектор $\nu_0 \in \mathbb{R}^m$.

B. Для $k \geq 0$ повторяем:

- **1.** Находим такой i_k, что $\nu_k^{(i_k)} = \max\limits_{1 \leq i \leq m} \nu_k^{(i)}$. Полагаем $r_k = [\nu^{(i_k)}]^{1/2}$.
- **2.** Если $r_k \leq \gamma n^{1/2}$ то СТОП. Иначе
  - 2.1. Полагаем $\sigma_k = \frac{1}{n} r_k^2 - 1$, $\alpha_k = \frac{\sigma_k}{r_k^2 - 1}$, и вычисляем $x_k = H_k a_{i_k}$.
  - 2.2. Пересчитываем $H_{k+1} := \frac{1}{1-\alpha_k} \left[H_k - \frac{\alpha_k}{1+\sigma_k} \cdot x_k x_k^* \right]$.
  - 2.3. Пересчитываем $\nu_{k+1}^{(i)} := \frac{1}{1-\alpha_k} \left[\nu_k^{(i)} - \frac{\alpha_k}{1+\sigma_k} \cdot \langle a_i, x_k \rangle^2 \right]$, $i = 1, \ldots, m$.

(6.2.7)

Оценим общую трудоемкость этой схемы. Для простоты предположим, что матрица $A = (a_1, \ldots, a_m)$ является сильно заполненной. Ниже мы выписываем только главные члены в оценках трудоемкости соответствующих вычислений.

- **Этап A** требует $\frac{mn^2}{2}$ операций для вычисления матрицы G_0, плюс $\frac{n^3}{6}$ операций для ее обращения и еще $\frac{mn^2}{2}$ операций для вычисления вектора ν_0.
- **Шаг 2.1** потребует n^2 операций.
- **Шаг 2.2** выполняется за $\frac{n^2}{2}$ операций.
- **Шаг 2.3** требует mn операций.

Учитывая теперь оценку (6.2.5) с $R = \sqrt{m}$ (см. пример 6.2.1), мы заключаем, что в случае $\gamma > 1$ и центрально-симметричного множества (6.2.3) метод (6.2.7) может найти эллипсоидальную аппроксимацию радиуса $\gamma\sqrt{n}$ за

$$\tfrac{n^2}{6}(n+6m) + \tfrac{n^2(2m+3n)\ln m}{2[2\ln\gamma - 1 + \gamma^{-2}]}$$

арифметических операций. Заметим, что для разреженной матрицы A сложность Этапа A и Шага 2.3 будет гораздо меньше.

B. Общие выпуклые множества

Для произвольного вектора g из E^* рассмотрим множество $C_g(G) = \operatorname{Conv}\{W_1(G), g\}$. Заметим, что его опорная функция записывается так:

$$\xi_{C_g(G)}(x) = \max\{\|x\|_G, \langle g, x\rangle\}, \quad x \in E.$$

Пусть $r = \|g\|^*_G$ и

$$G(\alpha) = (1-\alpha)G + \left(\tfrac{\alpha}{r} + \left(\tfrac{r-1}{2}\right)^2 \cdot \left(\tfrac{\alpha}{r}\right)^2\right) \cdot gg^*.$$

Лемма 6.2.2 *При любом $\alpha \in [0,1)$ эллипсоид*

$$E_\alpha = \{s \in E^* : \|s - \tfrac{r-1}{2r} \cdot \alpha g\|^*_{G(\alpha)} \le 1\}$$

принадлежит множеству $C_g(G)$. Если $r \ge n$, то функция

$$V(\alpha) \stackrel{\mathrm{def}}{=} \ln \tfrac{\det G(\alpha)}{\det G(0)} = 2\ln\left(1 + \alpha \cdot \tfrac{r-1}{2}\right) + (n-1)\ln(1-\alpha),$$

максимальна при $\alpha^* = \frac{2}{n+1} \cdot \frac{r-n}{r-1}$. Более того,

$$\begin{array}{rcl} V(\alpha^*) & = & 2\ln\tfrac{r-1}{n+1} + (n-1)\ln\tfrac{(n-1)(r+1)}{(n+1)(r-1)} \\ & \ge & 2\left[\ln(1+\sigma) - \tfrac{\sigma}{1+\sigma}\right] \ge \left(\tfrac{\sigma}{1+\sigma}\right)^2, \end{array} \qquad (6.2.8)$$

где $\sigma = \frac{r-n}{n+1}$.

Доказательство.

Нам нужно доказать, что при всех $x \in E$ выполнено неравенство

$$\begin{array}{rcl} \xi_{E_\alpha}(x) & \equiv & \alpha \cdot \frac{r-1}{2r} \cdot \langle g, x\rangle + \left[(1-\alpha)\|x\|^2_G + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2}\right)^2 \cdot \left(\frac{\alpha}{r}\right)^2\right) \langle g, x\rangle^2\right]^{1/2} \\ & \leq & \xi_{C_g(G)}(x) = \max\{\|x\|_G, \langle g, x\rangle\}. \end{array}$$

Прежде всего, если $\|x\|_G \leq \langle g, x\rangle$, то

$$\xi_{E_\alpha}(x) \leq \alpha \cdot \tfrac{r-1}{2r} \cdot \langle g, x\rangle + |1 - \alpha \cdot \tfrac{r-1}{2r}| \cdot \langle g, x\rangle = \langle g, x\rangle.$$

В противном случае имеем $-r\|x\|_G \leq \langle g, x\rangle \leq \|x\|_G$. Заметим, что значение $\xi_{E_\alpha}(x)$ выпуклым образом зависит от величины $\langle g, x\rangle$. Следовательно, его максимум достигается в граничных точках допустимого интервала для $\langle g, x\rangle$. Для граничной точки $\langle g, x\rangle = \|x\|_G$ мы уже доказали, что $\xi_{E_\alpha}(x) = \|x\|_G$. Рассмотрим теперь случай $\langle g, x\rangle = -r\|x\|_G$. Тогда

$$\xi_{E_\alpha}(x) = -\alpha \cdot \tfrac{r-1}{2} \cdot \|x\|_G + \left[(1-\alpha)\|x\|^2_G + \left(\tfrac{\alpha}{r} + \left(\tfrac{r-1}{2}\right)^2 \cdot \left(\tfrac{\alpha}{r}\right)^2\right) r^2 \|x\|^2_G\right]^{1/2} = \|x\|_G.$$

Таким образом, мы доказали, что $E_\alpha \subseteq C_g(G)$ при любом $\alpha \in [0, 1)$.

Далее,

$$\begin{array}{rcl} V(\alpha) & = & \ln\det(G^{-1/2}G(\alpha)G^{-1/2}) \\ & = & \ln\det\left((1-\alpha)I_n + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2}\right)^2 \cdot \left(\frac{\alpha}{r}\right)^2\right) G^{-1/2}gg^*G^{-1/2}\right) \\ & = & \ln\left(1 - \alpha + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2}\right)^2 \cdot \left(\frac{\alpha}{r}\right)^2\right) \cdot r^2\right) + (n-1)\ln(1-\alpha) \\ & = & 2\ln\left(1 + \alpha \cdot \frac{r-1}{2}\right) + (n-1)\ln(1-\alpha). \end{array}$$

Поэтому условия оптимальности для функции $V(\alpha)$ записываются так:

$$\tfrac{n-1}{1-\alpha} = \tfrac{r-1}{1+\alpha\cdot\frac{r-1}{2}}.$$

Таким образом, ее максимум достигается при $\alpha^* = \frac{2}{n+1} \cdot \frac{r-n}{r-1}$. Полагая $\sigma = \frac{r-n}{n+1}$, получаем

$$\begin{array}{rcl} V(\alpha^*) & = & 2\ln\left(1 + \alpha^* \cdot \frac{r-1}{2}\right) + (n-1)\ln(1-\alpha^*) \\ & = & 2\ln(1+\sigma) - (n-1)\ln\left(1 + \frac{2(r-n)}{(n-1)(r+1)}\right) \\ & \geq & 2\ln(1+\sigma) - \frac{2(r-n)}{r+1} = 2\left[\ln(1+\sigma) - \frac{\sigma}{1+\sigma}\right]. \qquad \Box \end{array}$$

В этом пункте мы решаем следующую задачу. Пусть выпуклое ограниченное множество $C \subset E^*$ имеет непустую внутренность. Для заданного $\gamma > 1$ нам нужно найти эллипсоидальную аппроксимацию множества C радиуса γn. Начальная аппроксимация этого тела дается центром v_0 и такой матрицей $G_0 \succ 0$, что $W_1(v_0, G_0) \subseteq Q \subseteq W_R(v_0, G_0)$ при некотором $R \geq 1$. Предполагается, что $n \equiv \dim E \geq 2$.

Рассмотрим следующую алгоритмическую схему.

```
Для $k \geq 0$ повторяем:

    1. Вычисляем $g_k \in C : \ \|g_k - v_k\|^*_{G_k} = r_k \overset{\text{def}}{=} \max\limits_{g} \{\|g - v_k\|^*_{G_k} : \ g \in C\}$.

    2. Если $r_k \leq \gamma n$ то СТОП. Иначе                                  (6.2.9)

        $\alpha_k := \frac{2}{n+1} \cdot \frac{r_k - n}{r_k - 1}, \quad v_{k+1} := v_k + \alpha_k \frac{r_k - 1}{2 r_k}(g_k - v_k)$,

        $G_{k+1} := (1 - \alpha_k) G_k + \left( \frac{\alpha_k}{r_k} + \left(\frac{r_k - 1}{2}\right)^2 \cdot \left(\frac{\alpha_k}{r_k}\right)^2 \right) \cdot (g_k - v_k)(g_k - v_k)^*$.
```

Оценка эффективности этой схемы дается следующей теоремой.

Теорема 6.2.2 *Пусть* $W_1(v_0, G_0) \subseteq C \subseteq W_R(v_0, G_0)$ *при некотором* $R \geq 1$. *Тогда схема (6.2.9) остановится не более чем после*

$$\tfrac{(1+2\gamma)^2}{2(\gamma-1)^2} \cdot n \ln R \tag{6.2.10}$$

итераций.

Доказательство.
Заметим, что коэффициент α_k, вектор v_{k+1} и матрица G_{k+1} на шаге 2 метода (6.2.9) выбраны в соответствии с леммой 6.2.2. Поскольку этот метод работает пока выполнено неравенство

$$\sigma_k \overset{\text{def}}{=} \tfrac{r_k - n}{n+1} \geq \tfrac{n}{n+1}(\gamma - 1) \geq \tfrac{2}{3}(\gamma - 1),$$

в силу неравенства (6.2.8) на каждом шаге $k \geq 0$ имеем

$$\ln \det G_{k+1} \geq \ln \det G_k + \left(\tfrac{\sigma_k}{1+\sigma_k}\right)^2 \geq \ln \det G_k + \left(\tfrac{2(\gamma-1)}{1+2\gamma}\right)^2. \tag{6.2.11}$$

Заметим, что при любом $k \geq 0$ выполняются равенства

$$\begin{aligned} \det(G_k)^{1/2} \cdot \operatorname{vol}_n(W_1(I_n)) &= \operatorname{vol}_n(W_1(G_k)) \leq \operatorname{vol}_n(C) \leq \operatorname{vol}_n(W_R(G_0)) \\ &= \mathbb{R}^n \cdot \det(G_0)^{1/2} \cdot \operatorname{vol}_n(W_1(I_n)). \end{aligned}$$

Поэтому $\ln \det G_k - \ln \det G_0 \leq 2n \ln R$ и мы получаем оценку (6.2.10) суммированием неравенств (6.2.11). □

Заметим, что в случае $C = \mathrm{Conv}\,\{a_i,\ i = 1, \ldots, m\}$ схема (6.2.9) может быть эффективно реализована в том же стиле, что и метод (6.2.7). Исходный аппроксимирующий эллипсоид для множества C может быть выбран следующим образом.

Лемма 6.2.3 *Пусть у множества* $C = \mathrm{Conv}\,\{a_i,\ i = 1, \ldots, m\}$ *непустая внутренность. Определим*

$$\hat{a} = \tfrac{1}{m} \sum_{i=1}^{m} a_i, \quad G = \tfrac{1}{R^2} \sum_{i=1}^{m} (a_i - \hat{a})(a_i - \hat{a})^*,$$

где $R = \sqrt{m(m-1)}$. *Тогда* $W_1(\hat{a}, G) \subset C \subset W_R(\hat{a}, G)$.

Доказательство.

Для любых точек $x \in E$ и значений $r > 0$ имеем

$$\xi_{W_r(\hat{a},G)}(x) = \langle \hat{a}, x \rangle + r \|x\|_G = \langle \hat{a}, x \rangle + \tfrac{r}{R} \left[\sum_{i=1}^{m} \langle a_i - \hat{a}, x \rangle^2 \right]^{1/2}.$$

Таким образом, $\xi_{W_R(\hat{a},G)}(x) \geq \max\limits_{1 \leq i \leq m} \langle a_i, x \rangle = \xi_C(x)$. Поэтому $W_R(\hat{a}, G) \supset C$. Далее, пусть

$$\begin{array}{rcl} \tau_i & = & \langle a_i - \hat{a}, x \rangle, \quad i = 1, \ldots, m, \quad \text{и} \\[2ex] \hat{\tau} & = & \max\limits_{1 \leq i \leq m} \langle a_i, x \rangle - \langle \hat{a}, x \rangle \geq 0. \end{array}$$

Заметим, что $\sum\limits_{i=1}^{m} \tau_i = 0$ и $\tau_i \leq \hat{\tau}$ при всех i. Таким образом,

$$\begin{array}{rcl} \xi_{W_1(\hat{a},G)}(x) - \langle \hat{a}, x \rangle & \leq & \tfrac{1}{R} \max\limits_{\tau_i} \left\{ \left[\sum\limits_{i=1}^{m} \tau_i^2 \right]^{1/2} :\ \sum\limits_{i=1}^{m} \tau_i = 0,\ \tau_i \leq \hat{\tau},\ i = 1, \ldots, m \right\} \\[3ex] & = & \tfrac{\hat{\tau}}{R} \sqrt{m(m-1)} = \max\limits_{1 \leq i \leq m} \langle a_i, x \rangle - \langle \hat{a}, x \rangle = \xi_C(x) - \langle \hat{a}, x \rangle. \end{array}$$

Следовательно, $W_1(\hat{a}, G) \subset C$. □

C. Знако-инвариантные множества

Множество $C \subset \mathbb{R}^n$ называется *знако-инвариантным* если для любой точки g из C произвольное изменение знаков ее координат оставляет эту точку в C. Другими словами, для любого $g \in C \bigcap \mathbb{R}^n_+$, имеем

$$B(g) \equiv \{ s \in \mathbb{R}^n :\ -g \leq s \leq g \} \subseteq C.$$

Примерами таких множеств могут служить единичные шары в l_p-нормах и евклидовы нормы, порождаемые диагональными матрицами (но не только они).

Ясно, что любое знако-инвариантное множество является центрально-симметричным. Таким образом, в силу леммы 6.2.1 для него существует эллипсоидальная аппроксимация радиуса $\sqrt{n}$. Однако, как мы покажем, для таких множеств матрица этого эллипсоида может быть выбрана диагональной.

Пусть $D \succ 0$ – диагональная матрица. Выберем произвольный вектор $g \in \mathbb{R}^n_+ \subset E^*$. Пусть

$$\begin{aligned} C &= \mathrm{Conv}\,\{W_1(D), B(g)\}, \quad \text{и} \\ \\ G(\alpha) &= (1-\alpha)D + \alpha D^2(g). \end{aligned}$$

Ясно, что C – знако-инвариантное множество. Положим

$$V(\alpha) = \ln \tfrac{\det G(0)}{\det G(\alpha)} = -\sum_{i=1}^{n} \ln\left(1+\alpha(\tau_i - 1)\right), \quad \alpha \in [0,1),$$

где $\tau_i = \frac{(g^{(i)})^2}{D^{(i)}}$, $i = 1,\ldots,n$. Заметим, что функция $V(\alpha)$ является самосогласованной (см. параграф 1.3). Вычислим ее производные:

$$\begin{aligned} V'(0) &= n - \sum_{i=1}^{n} \tau_i = n - \left(\|g\|^*_D\right)^2, \quad \text{and} \\ \\ V''(0) &= \sum_{i=1}^{n} (\tau_i - 1)^2. \end{aligned} \qquad (6.2.12)$$

Лемма 6.2.4 *При любом* $\alpha \in [0,1]$ *имеем* $W_1(G(\alpha)) \subseteq C$. *Предполагая, что* $(\|g\|^*_D)^2 > n$, *определим шаг*

$$\alpha^* \overset{\mathrm{def}}{=} \tfrac{(\|g\|^*_D)^2 - n}{(2(\|g\|^*_D)^2 - n)\cdot(\|g\|^*_D)^2}.$$

Тогда $\alpha^* \in (0, \frac{1}{n}]$ *и для любого* $\gamma \in \left(1, \frac{1}{\sqrt{n}}\|g\|^*_D\right]$ *выполнено неравенство*

$$V(\alpha^*) \le \ln\left(1 + \tfrac{\gamma^2 - 1}{\gamma^2}\right) - \tfrac{\gamma^2-1}{\gamma^2} < 0. \qquad (6.2.13)$$

Доказательство.
Для любых $\alpha \in [0,1]$ и $x \in \mathbb{R}^n \equiv E$ получаем:

$$\begin{aligned} [\xi_{W_1(G(\alpha))}(x)]^2 &= (1-\alpha)\langle Dx, x\rangle + \alpha \sum_{i=1}^{n} (g^{(i)} x^{(i)})^2 \\ \\ &\le (1-\alpha)\langle Dx, x\rangle + \alpha \left(\sum_{i=1}^{n} g^{(i)} \cdot |x^{(i)}|\right)^2 \\ \\ &\le \left[\max\{\xi_{W_1(D)}(x), \xi_{B(g)}(x)\}\right]^2 = [\xi_C(x)]^2. \end{aligned}$$

Далее, обозначим $S = \sum\limits_{i=1}^{n} \tau_i = (\|g\|^*_D)^2$. В силу предположений леммы, $S > n$. Таким образом,

$$V''(0) \ \le \ \max_{\tau} \left\{ \sum_{i=1}^{n} (\tau_i - 1)^2 : \ \sum_{i=1}^{n} \tau_i = S, \ \tau_i \ge 0, \ i = 1 \ldots n \right\}$$

$$= \ (S-1)^2 + n - 1 \ < \ S^2.$$

Обозначим $\omega_*(\tau) = -\tau - \ln(1-\tau)$. Поскольку $V(\cdot)$ – самосогласованная функция, из неравенства (1.3.8) получаем:

$$V(\alpha) \ \le \ V(0) + \alpha \cdot V'(0) + \omega_*(\alpha \cdot (V''(0))^{1/2}) \ \le \ -\alpha \cdot (S-n) + \omega_*(\alpha \cdot S). \qquad (6.2.14)$$

Минимум правой части этого неравенства может быть найден из уравнения

$$S - n \ = \ \tfrac{\alpha_* S^2}{1-\alpha_* S}.$$

Таким образом, $\alpha_* = \frac{S-n}{S\cdot(2S-n)} < \frac{1}{n}$. При этом, убывание правой части неравенства (6.2.14) составляет

$$\omega\left(1 - \tfrac{n}{S}\right) \ \ge \ \omega(1-\gamma^{-2}),$$

где $\omega(t) = \max\limits_{\tau}[\tau t - \omega_*(\tau)] = t - \ln(1+t)$. □

Следствие 6.2.1 *Для любого знако-симметричного множества $C \subset \mathbb{R}^n$ с непустой внутренностью существует такая диагональная матрица $D \succ 0$, что*

$$W_1(D) \subseteq C \subseteq W_{\sqrt{n}}(D).$$

Доказательство.
Заметим, что для достаточно большого R множество $\{D \succeq 0 : \ W_1(D) \subseteq C \subseteq W_R(D)\}$ непусто, выпукло и ограничено. Таким образом, существование эллипсоидальной аппроксимации радиуса $\sqrt{n}$ следует из неравенства (6.2.13). □

Последнее следствие будет важным для нас ввиду следующего факта.

Лемма 6.2.5 *Пусть все векторы $a_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, имеют неотрицательные коэффициенты. Предположим, что существует такая диагональная матрица $D \succ 0$, что*

$$W_1(D) \subseteq \mathrm{Conv}\,\{B(a_i), \ i = 1, \ldots, m\} \subseteq W_{\gamma\sqrt{n}}(D)$$

при некотором $\gamma \ge 1$. Тогда функция $f(x) = \max\limits_{1 \le i \le m} \langle a_i, x \rangle$ удовлетворяет неравенству

$$\|x\|_D \le f(x) \le \gamma\sqrt{n} \cdot \|x\|_D \quad \forall x \in \mathbb{R}^n_+. \qquad (6.2.15)$$

Доказательство.

Рассмотрим функцию $\hat{f}(x) = \max\limits_{1\le i\le m} \sum\limits_{j=1}^{n} a_i^{(j)} |x^{(j)}|$. Заметим, что ее субдифференциал представим в виде

$$\partial \hat{f}(0) = \mathrm{Conv}\,\{B(a_i),\ i = 1,\dots,m\}.$$

Таким образом, для любого $x \in \mathbb{R}^n$ имеем

$$\begin{aligned} \|x\|_D &= \max_s\{\langle s,x\rangle :\ s \in W_1(D)\} \le \max_s\{\langle s,x\rangle :\ s \in \partial\hat{f}(0)\} \equiv \hat{f}(x) \\ &\le \max_s\{\langle s,x\rangle :\ s \in W_{\gamma\sqrt{n}}(D)\} = \gamma\sqrt{n}\cdot\|x\|_D. \end{aligned}$$

Остается заметить, что $\hat{f}(x) \equiv f(x)$ при всех $x \in \mathbb{R}^n_+$. □

Следствие 6.2.2 *Пусть* $a_i \in \mathbb{R}^n_+$, $i = 1,\dots,m$. *Рассмотрим множество*

$$\mathcal{F} = \{x \in \mathbb{R}^n_+ :\ \langle a_i, x\rangle \le b_i,\ i = 1,\dots,m\}$$

с $b_i > 0$, $i = 1,\dots,m$. *Тогда существует такая диагональная матрица* $D \succ 0$, *что*

$$W_1(D) \bigcap \mathbb{R}^n_+ \subset \mathcal{F} \subset W_{\sqrt{n}}(D) \bigcap \mathbb{R}^n_+. \tag{6.2.16}$$

Доказательство.

Рассмотрим функцию $f(x) = \max\limits_{1\le i\le m} \frac{1}{b_i}\langle a_i, x\rangle$. В силу следствия 6.2.1 предположения леммы 6.2.5 выполняются с $\gamma = 1$. Поскольку $\mathcal{F} = \{x \in \mathbb{R}^n_+ :\ f(x) \le 1\}$, включения (6.2.16) следуют из неравенств (6.2.15). □

Покажем как найти диагональный аппроксимирующий эллипсоид для следующего знако-инвариантного множества:

$$C = \mathrm{Conv}\,\{B(a_i),\ i = 1,\dots,m\},$$

где $a_i \in \mathbb{R}^n_+ \setminus \{0\}$, $i = 1,\dots,m$. Основное предположение об этих векторах выглядит так:

$$\hat{a} \overset{\mathrm{def}}{=} \frac{1}{m}\sum_{i=1}^{m} a_i > 0.$$

Обозначим $\hat{D} = D^2(\hat{a})$.

Лемма 6.2.6 *Справедливо включение* $W_1(\hat{D}) \subset C \subset W_{m\sqrt{n}}(\hat{D})$.

Доказательство.

Поскольку $\hat{a} \in C$, то $W_1(\hat{D}) \subset B(\hat{a}) \subseteq C$. С другой стороны,

$$C \subseteq B(m\hat{a}) \subset \left\{x \in \mathbb{R}^n :\ \sum_{i=1}^{n} \left(\frac{x^{(i)}}{m\hat{a}^{(i)}}\right)^2 \le n\right\} = W_{m\sqrt{n}}(\hat{D}). \qquad \square$$

Для вышеприведенного знако-инвариантного множества $C \subset \mathbb{R}^n$ рассмотрим специальную схему, которая находит диагональную аппроксимацию радиуса $\gamma\sqrt{n}$, где $\gamma > \left[1 + \frac{1}{\sqrt{n}}\right]^{1/2}$.

Положим $D_0 = \hat{D}$.

Для $k \geq 0$ **повторяем:**

1. Вычисляем $i_k : \; \| a_{i_k} \|^*_{D_k} = r_k \overset{\text{def}}{=} \max\limits_{1 \leq i \leq m} \| a_i \|^*_{D_k}$.

2. Если $r_k \leq \gamma\sqrt{n}$ то СТОП. Иначе

$$\beta_k := \sum_{j=1}^{n} \left(\frac{(a_{i_k}^{(j)})^2}{D_k^{(j)}} - 1 \right)^2, \quad \alpha_k := \frac{r_k^2 - n}{\beta_k + (r_k^2 - n)\beta_k^{1/2}},$$

$$D_{k+1} := (1 - \alpha_k) D_k + \alpha_k D^2(a_{i_k}).$$

(6.2.17)

Заметим, что в этой схеме применяются правила леммы 6.2.4, используя обозначение β_k для $V''(0)$. Таким образом, в точности так же, как в теоремах 6.2.1 и 6.2.2, приходим к следующему утверждению.

Теорема 6.2.3 *Для* $\gamma \geq \left[1 + \frac{1}{\sqrt{n}}\right]^{1/2}$ *схема (6.2.17) останавливается не более чем через*

$$\left[\frac{\gamma^2 - 1}{\gamma^2} - \ln\left(1 + \frac{\gamma^2 - 1}{\gamma^2}\right) \right]^{-1} \cdot n(\ln n + 2 \ln m)$$

итераций.

Заметим, что количество операций на каждой итерации схемы (6.2.17) пропорционально числу ненулевых элементов в матрице $A = (a_1, \dots, a_m)$.

6.2.3 Минимизация максимального модуля линейных форм

Рассмотрим следующую задачу линейного программирования:

$$\max_{1 \leq i \leq m} |\langle \bar{a}_i, y \rangle - c_i| \quad \to \quad \min : \; y \in \mathbb{R}^{n-1}. \tag{6.2.18}$$

Обозначая $a_i = (\bar{a}_i^T, -c_i)^T$, $i = 1, \dots, m$, $x = (y^T, \tau)^T \in \mathbb{R}^n \equiv E$ и $d = e_n$ можно переписать эту задачу в конической форме:

$$\text{найти } f^* = \min_x \left\{ f(x) \overset{\text{def}}{=} \max_{1 \leq i \leq m} |\langle a_i, x \rangle| : \; \langle d, x \rangle = 1 \right\}. \tag{6.2.19}$$

В параграфе 6.1 для нахождения эллипсоидальной аппроксимации множества $\partial f(0)$ использовалась структура функции $f(x)$. Однако радиус этой аппроксимации был достаточно большим, порядка $O(\sqrt{m})$. Теперь с помощью схемы (6.2.4) мы можем эффективно вычислять эллипсоидальную аппроксимацию с радиусом порядка $O(\sqrt{n})$. Покажем, что это приводит к более эффективному методу минимизации.

Зафиксируем некоторое $\gamma > 1$. Предположим, что пользуясь процессом (6.2.4) мы построили эллипсоидальную аппроксимацию центрально симметричного множества $\partial f(0)$ с радиусом $\gamma\sqrt{n}$:

$$W_1(G) \subseteq \partial f(0) \equiv \mathrm{Conv}\,\{\pm a_i,\ i = 1,\dots,m\} \subseteq W_{\gamma\sqrt{n}}(G).$$

Из этого немедленно следует, что

$$\|x\|_G \le f(x) \equiv \sup_s\{\langle s, x\rangle :\ s \in \partial f(0)\} \le \gamma\sqrt{n}\cdot\|x\|_G, \tag{6.2.20}$$

$$\|a_i\|_G^* \le \gamma\sqrt{n}, \quad i = 1,\dots,m. \tag{6.2.21}$$

Зафиксируем некоторое значение параметра сглаживания $\mu > 0$. Рассмотрим следующую гладкую аппроксимацию функции $f(x)$:

$$f_\mu(x) = \mu \ln\left(\sum_{i=1}^m \left[e^{\langle a_i,x\rangle/\mu} + e^{-\langle a_i,x\rangle/\mu}\right]\right).$$

Ясно, что функция $f_\mu(x)$ выпукла и бесконечное число раз дифференцируема на E. Более того,

$$f(x) \le f_\mu(x) \le f(x) + \mu\ln(2m) \quad \forall x \in \mathbb{R}^n. \tag{6.2.22}$$

Наконец заметим, что для любой точки x и любого направления h из E имеем

$$\begin{array}{rcl}
\langle \nabla f_\mu(x), h\rangle & = & \sum\limits_{i=1}^m \lambda_\mu^{(i)}(x)\cdot\langle a_i, h\rangle,\\[2ex]
\lambda_\mu^{(i)}(x) & = & \frac{1}{\omega_\mu(x)}\cdot\left(e^{\langle a_i,x\rangle/\mu} - e^{-\langle a_i,x\rangle/\mu}\right), \quad i = 1,\dots,m,\\[2ex]
\omega_\mu(x) & = & \sum\limits_{i=1}^m \left(e^{\langle a_i,x\rangle/\mu} + e^{-\langle a_i,x\rangle/\mu}\right).
\end{array}$$

Таким образом, гессиан сглаженной функции вычисляется по формуле

$$\langle \nabla^2 f_\mu(x)h, h\rangle \;=\; \frac{1}{\mu}\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle/\mu}+e^{-\langle a_i,x\rangle/\mu}}{\omega_\mu(x)}\cdot\langle a_i,h\rangle^2 - \frac{1}{\mu}\left(\sum_{i=1}^m \lambda_\mu^{(i)}(x)\cdot\langle a_i,h\rangle\right)^2.$$

В силу неравенства (6.2.21) имеем

$$\langle \nabla^2 f_\mu(x)h, h\rangle \le \frac{1}{\mu}\left(\max_{1\le i\le m}\|a_i\|_G^*\right)^2\cdot\|h\|_G^2 \le \frac{\gamma^2 n}{\mu}\cdot\|h\|_G^2.$$

Это означает, что градиент функции $f_\mu(\cdot)$ липшицев в метрике $\|\cdot\|_G$ с константой $L_\mu = \frac{\gamma^2 n}{\mu}$:

$$\|\nabla f_\mu(x) - \nabla f_\mu(y)\|_G^* \le L_\mu \|x - y\|_G \quad \forall x, y \in E.$$

Наш подход очень похож на то, что делалось в параграфе 6.1. Рассмотрим задачу

$$\min_x \{\phi(x);\ x \in Q\}, \tag{6.2.23}$$

где Q – выпуклое замкнутое множество, а выпуклая дифференцируемая функция $\phi(x)$ имеет градиент, который липшицев в евклидовой норме $\|\cdot\|_G$ с констаной L. Запишем быстрый градиентный метод (1.2.18) применительно к задаче (6.2.23).

```
**Метод** $S(\phi, L, Q, G, x_0, N)$

**Для** $k = 0, \dots, N$ **повторяем:**

    1. Вычисляем $\nabla \phi(x_k)$.

    2. $y_k := \arg\min_{y \in Q} \left[ \langle \nabla \phi(x_k), y - x_k \rangle + \frac{L}{2} \|y - x_k\|_G^2 \right].$

    3. $z_k := \arg\min_{z \in Q} \left[ \langle \sum_{i=0}^{k} \frac{i+1}{2} \nabla \phi(x_i), z - x_0 \rangle + \frac{L}{2} \|z - x_0\|_G^2 \right].$

    4. $x_{k+1} := \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k.$

**Результат**   $S(\phi, L, Q, G, x_0, N) \equiv y_N.$
```

(6.2.24)

В соответствии с теоремой 1.2.3, результирующая точка y_N удовлетворяет неравенству

$$\phi(y_N) - \phi(x_\phi^*) \le \frac{2L\|x_0 - x_\phi^*\|_G^2}{(N+1)^2}, \tag{6.2.25}$$

где x_ϕ^* – оптимальное решение задачи (6.2.23).

Как и в параграфе 6.1, мы воспользуемся методом (6.2.24) для нахождения решения задачи (6.2.19) с относительной точностью $\delta > 0$. Положим

$$\begin{array}{rcl} Q(r) & = & \{x \in \mathbb{R}^n :\ \langle d, x \rangle = 1,\ \|x\|_G \le r\}, \\ x_0 & = & \frac{G^{-1} d}{\langle d, G^{-1} d \rangle}, \\ \tilde{N} & = & \left\lfloor 2e\gamma \sqrt{2n \ln(2m)} \left(1 + \frac{1}{\delta}\right) \right\rfloor. \end{array}$$

Рассмотрим следующий метод.

```
Полагаем $\hat{x}_0 = x_0$.

Для $t \geq 1$ повторяем:
    $\mu_t := \frac{\delta f(\hat{x}_{t-1})}{2e(1+\delta)\ln(2m)}$;   $L_{\mu_t} := \frac{\gamma^2 n}{\mu_t}$;
    $\hat{x}_t := S\left(f_{\mu_t}, L_{\mu_t}, Q(f(\hat{x}_{t-1})), G, x_0, \tilde{N}\right)$;
    Если $f(\hat{x}_t) \geq \frac{1}{e} f(\hat{x}_{t-1})$, то $T := t$ и СТОП.
```

(6.2.26)

Теорема 6.2.4 *Число этапов верхнего уровня в методе (6.2.26) ограничено:*

$$T \leq 1 + \ln(\gamma\sqrt{n}). \tag{6.2.27}$$

Последняя точка удовлетворяет неравенству $f(\hat{x}_T) \leq (1+\delta)f^*$. *Общее число итераций в процессах (6.2.26) нижнего уровня не превосходит*

$$2\gamma e(1 + \ln(\gamma\sqrt{n}))\sqrt{2n\ln(2m)}\left(1 + \tfrac{1}{\delta}\right). \tag{6.2.28}$$

Доказательство.

Пусть x^* – оптимальное значение задачи (6.2.19). Заметим, что все точки $\hat{x}_t$, сформированные методом (6.2.26), допустимы для задачи (6.2.19). Таким образом, в силу (6.2.20) имеем

$$f(\hat{x}_t) \geq f^* \geq \|x^*\|_G.$$

Следовательно, $x^* \in Q(f(\hat{x}_t))$ для любого $t \geq 0$. Положим

$$f_t^* = f_{\mu_t}(x_t^*) = \min_x \{f_{\mu_t}(x) : \ x \in Q(f(\hat{x}_{t-1}))\}.$$

Поскольку $x^* \in Q(f(\hat{x}_t))$, в силу (6.2.22) имеем

$$f_t^* \leq f_{\mu_t}(x^*) \leq f^* + \mu_t \ln(2m).$$

Из первой части соотношений (6.2.22) получаем $f(\hat{x}_t) \leq f_{\mu_t}(\hat{x}_t)$. Заметим, что

$$\|x_0 - x_t^*\|_G \leq \|x_t^*\|_G \leq f(\hat{x}_{t-1}), \quad t \geq 1.$$

В силу оценки (6.2.25), на последней итерации T выполнено неравенство

$$\begin{array}{rcl} f(\hat{x}_T) - f^* & \leq & f_{\mu_T}(\hat{x}_T) - f_T^* + \mu_T \ln(2m) \\ & \leq & \frac{2L_{\mu_T} f^2(\hat{x}_{T-1})}{\left(\tilde{N}+1\right)^2} + \mu_T \ln(2m) \\ & = & \frac{2\gamma^2 n f^2(\hat{x}_{T-1})}{\mu_T\left(\tilde{N}+1\right)^2} + \mu_T \ln(2m) \\ & \leq & \frac{f^2(\hat{x}_{T-1})\,\delta^2}{4\mu_T e^2 \ln(2m)(1+\delta)^2} + \mu_T \ln(2m) \\ & = & 2\mu_T \ln(2m). \end{array}$$

Далее, в силу выбора параметра μ_t и критерия остановки имеем

$$2\mu_T \ln(2m) = \frac{\delta f(\hat{x}_{T-1})}{e(1+\delta)} \leq \frac{\delta f(\hat{x}_T)}{1+\delta}.$$

Таким образом, $f(\hat{x}_T) \leq (1+\delta) f^*$.

Остается доказать верхнюю границу (6.2.27) для общего числа шагов в нашем процессе. Действительно, с помощью простой индукции легко доказать, что в начале каждого этапа t выполняется неравенство

$$\left(\tfrac{1}{e}\right)^{t-1} f(x_0) \geq f(\hat{x}_{t-1}), \quad t \geq 1.$$

Заметим, что x_0 – проекция нуля на гиперплоскость $\langle d, x\rangle = 1$. Таким образом, в силу неравенств (6.2.20) имеем

$$f^* \geq \|x^*\|_G \geq \|x_0\|_G \geq \tfrac{1}{\gamma\sqrt{n}} f(x_0).$$

Следовательно, на последнем шаге получаем

$$\left(\tfrac{1}{e}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \tfrac{1}{\gamma\sqrt{n}} f(x_0).$$

Отсюда легко получить оценку (6.2.27). □

Напомним, что предварительная часть метода (6.2.26), состоящая в вычислении эллипсоидальной аппроксимации множества $\partial f(0)$ с радиусом $\gamma\sqrt{n}$, осуществляется процедурой (6.2.4) за

$$\tfrac{n^2}{6}(n+6m) + \tfrac{n^2(2m+3n)\ln m}{2[2\ln\gamma - 1 + \gamma^{-2}]} = O(n^2(n+m)\ln m)$$

арифметических операций. Поскольку каждый шаг метода (6.2.24) требует $O(mn)$ операций, трудоемкость предварительной части превосходит трудоемкость оптимизационной части при не слишком маленьких δ, а именно, при $\delta > \frac{1}{\sqrt{n}}$.

6.2.4 Билинейные матричные игры с неотрицательными коэффициентами

Пусть $n \times m$-матрица $A = (a_1, \dots, a_m)$ имеет неотрицательные коэффициенты. Рассмотрим задачу

$$\text{найти } f^* = \min_x \{ f(x) \overset{\text{def}}{=} \max_{1 \le i \le m} \langle a_i, x \rangle : \ x \in \Delta_n \}. \tag{6.2.29}$$

Заметим, что в этом формате могут быть записаны разные стандартные постановки. Например, рассмотрим *линейную задачу упаковки*

$$\text{найти } \psi^* = \max_{y \ge 0 \in \mathbb{R}^n} \{ \langle c, y \rangle : \ \langle a_i, y \rangle \le b^{(i)}, \ i = 1, \dots, m \},$$

где все векторы a_i неотрицательны, $b > 0 \in \mathbb{R}^m$ и $c > 0 \in \mathbb{R}^n$. Тогда

$$\begin{array}{rcl} \psi^* & = & \max\limits_{y \ge 0 \in \mathbb{R}^n} \{ \langle c, y \rangle : \ \max\limits_{1 \le i \le m} \frac{1}{b^{(i)}} \langle a_i, y \rangle \le 1 \} \\ & = & \max\limits_{y \ge 0 \in \mathbb{R}^n} \frac{\langle c, y \rangle}{\max\limits_{1 \le i \le m} \frac{1}{b^{(i)}} \langle a_i, y \rangle} \\ & = & \left[\min\limits_{y \ge 0 \in \mathbb{R}^n} \left\{ \max\limits_{1 \le i \le m} \frac{1}{b^{(i)}} \langle a_i, y \rangle : \ \langle c, y \rangle = 1 \right\} \right]^{-1} \\ & = & \left[\min\limits_{x \in \mathbb{R}^n} \left\{ \max\limits_{1 \le i \le m} \frac{1}{b^{(i)}} \langle D^{-1}(c) a_i, x \rangle : \ x \in \Delta_n \right\} \right]^{-1}. \end{array}$$

Как обычно, давайте приблизим целевую функцию $f(x)$ в задаче (6.2.29) гладкой функцией

$$f_\mu(x) = \mu \ln \left(\sum_{i=1}^m e^{\langle a_i, x \rangle / \mu} \right).$$

В этом случае выполнены неравенства

$$f(x) \le f_\mu(x) \le f(x) + \mu \cdot \ln m, \quad \forall x \in \mathbb{R}^n. \tag{6.2.30}$$

Положим

$$\hat{f}(x) = \max_{1 \le i \le m} \sum_{j=1}^n a_i^{(j)} |x^{(j)}|.$$

Заметим, что субдифференциал однородной функции $\hat{f}(x)$ в нуле записывается так:

$$\partial f(0) \quad = \quad \mathrm{Conv} \{ B(a_i), \ i = 1, \dots, m \}.$$

В п. 6.2.2 мы показали, как можно вычислить такую диагональную матрицу $D \succ 0$, что

$$W_1(D) \subseteq \partial \hat{f}(0) \subseteq W_{2\sqrt{n}}(D),$$

(это соответствует выбору $\gamma = 2$ в схеме (6.2.17)). В силу леммы 6.2.5 с помощью этой матрицы можно определить такую евклидову норму $\| \cdot \|_D$, что

$$\|x\|_D \le f(x) \le 2\sqrt{n} \cdot \|x\|_D, \quad \forall x \in \mathbb{R}^n_+. \tag{6.2.31}$$

Более того, в этой норме размеры всех векторов a_i ограничены константой $2\sqrt{n}$.

Теперь, пользуясь такими же рассуждениями, как и в п. 6.2.3, можно показать, что для любых x и h из $\mathbb{R}^n$ выполнено неравенство

$$\langle \nabla^2 f_\mu(x)h, h\rangle \le \tfrac{4n}{\mu} \cdot \|h\|_D^2.$$

Поэтому градиент этой функции липшицев относительно нормы $\|\cdot\|_D$ с константой $\frac{4n}{\mu}$. Это означает, что функцию $f_\mu(x)$ можно минимизировать быстрым градиентным методом (6.2.24).

Зафиксируем относительную точность $\delta > 0$. Положим

$$\begin{aligned} Q(r) &= \{x \in \Delta_n : \ \|x\|_D \le r\}, \\ x_0 &= \tfrac{D^{-1}\bar{e}_n}{\langle \bar{e}_n, D^{-1}\bar{e}_n\rangle}, \\ \tilde{N} &= \left\lfloor 4e\sqrt{2n\ln m}\left(1 + \tfrac{1}{\delta}\right)\right\rfloor. \end{aligned}$$

Рассмотрим следующий метод.

```
Полагаем $\hat{x}_0 = x_0$.

Для $t \ge 1$ повторяем:                                                    (6.2.32)
    $\mu_t := \frac{\delta f(\hat{x}_{t-1})}{2e(1+\delta)\ln m}$;    $L_{\mu_t} := \frac{4n}{\mu_t}$;
    $\hat{x}_t := S\left(f_{\mu_t}, L_{\mu_t}, Q(f(\hat{x}_{t-1})), D, x_0, \tilde{N}\right)$;
    Если $f(\hat{x}_t) \ge \frac{1}{e} f(\hat{x}_{t-1})$, то $T := t$ и СТОП.
```

Обоснование этого метода очень похоже на обоснование метода (6.2.26).

Теорема 6.2.5 *Число точек, сформированных методом (6.2.26) ограничено:*

$$T \le 1 + \ln(2\sqrt{n}). \tag{6.2.33}$$

Последняя из этих точек удовлетворяет неравенству $f(\hat{x}_T) \le (1+\delta)f^*$. Общее число шагов в процессах нижнего уровня (6.2.26) не превосходит

$$4e(1 + \ln(2\sqrt{n}))\sqrt{2n\ln m}\left(1 + \tfrac{1}{\delta}\right). \tag{6.2.34}$$

Доказательство.

Пусть x^* – оптимальное решение задачи (6.2.29). Заметим, что все точки $\hat{x}_t$ сформированные методом (6.2.32) допустимы. Таким образом, в силу неравенства (6.2.31) имеем

$$f(\hat{x}_t) \geq f^* \geq \|x^*\|_D.$$

Следовательно, $x^* \in Q(f(\hat{x}_t))$ для любого $t \geq 0$. Положим

$$f_t^* = f_{\mu_t}(x_t^*) = \min_x \{f_{\mu_t}(x) : \; x \in Q(f(\hat{x}_{t-1}))\}.$$

Поскольку $x^* \in Q(f(\hat{x}_t))$, в силу неравенства (6.2.30) получаем

$$f_t^* \leq f_{\mu_t}(x^*) \leq f^* + \mu_t \ln m.$$

Из первой части неравенства (6.2.30) следует, что $f(\hat{x}_t) \leq f_{\mu_t}(\hat{x}_t)$. Заметим, что $\|x_0 - x_t^*\|_D \leq \|x_t^*\|_D \leq f(\hat{x}_{t-1})$ для $t \geq 1$. Таким образом, в силу оценки (6.2.25) на последней итерации T имеем

$$\begin{array}{rcl} f(\hat{x}_T) - f^* & \leq & f_{\mu_T}(\hat{x}_T) - f_T^* + \mu_T \ln m \\ & \leq & \frac{2L_{\mu_T} f^2(\hat{x}_{T-1})}{\left(\tilde{N}+1\right)^2} + \mu_T \ln m \\ & = & \frac{8n f^2(\hat{x}_{T-1})}{\mu_T \left(\tilde{N}+1\right)^2} + \mu_T \ln m \\ & \leq & \frac{f^2(\hat{x}_{T-1})\,\delta^2}{4\mu_T e^2 \ln m (1+\delta)^2} + \mu_T \ln m = 2\mu_T \ln m. \end{array}$$

Далее, в силу выбора параметра μ_T и критерия остановки получаем

$$2\mu_T \ln m = \frac{\delta f(\hat{x}_{T-1})}{e(1+\delta)} \leq \frac{\delta f(\hat{x}_T)}{1+\delta}.$$

Следовательно, $f(\hat{x}_T) \leq (1+\delta) f^*$.

Остается доказать оценку (6.2.33) для общего числа шагов в процессе. Действительно, с помощью простой индукции нетрудно показать, что в начале каждого этапа t выполнено неравенство

$$\left(\frac{1}{e}\right)^{t-1} f(x_0) \geq f(\hat{x}_{t-1}), \quad t \geq 1.$$

Заметим, что x_0 – проекция нуля на гиперплоскость $\langle \bar{e}_n, x \rangle = 1$. Таким образом, в силу неравенств (6.2.31) имеем

$$f^* \geq \|x^*\|_D \geq \|x_0\|_D \geq \frac{1}{2\sqrt{n}} f(x_0).$$

Следовательно, на последнем этапе метода выполнены соотношения

$$\left(\frac{1}{e}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \frac{1}{2\sqrt{n}} f(x_0).$$

Это неравенство дает нам оценку (6.2.33). □

Таким образом, мы видим, что методу (6.2.32) требуется $O\left(\frac{\sqrt{n \ln m}}{\delta} \ln n\right)$ градиентных итераций метода (6.2.24). Поскольку матрица D диагональная, каждая такая итерация будет очень дешевой. Ее трудоемкость будет пропорциональна числу ненулевых элементов матрицы A. Заметим также, что на шаге 2 и шаге 3 метода (6.2.24) надо вычислять проекцию на множество $Q(r)$, которое представляет собой пересечение симплекса с эллипсоидом. Однако, поскольку матрица D диагональная, это может быть сделано за $O(n \ln n)$ операций.

Заключение

В настоящей работе мы показали, что правильное использование модели экстремальной задачи существенно увеличивает возможности методов минимизации по генерации полезной информации. Это было доказано с помощью построения новых оптимизационных алгоритмов. Так, были построены новые прямо-двойственные методы минимизации негладких функций, в которых можно формировать приближенные решения сопряженных задач различного уровня и пользоваться точными критериями остановки. Были построены новые методы решения задач с относительной точностью. Это совершенно новый класс задач, в которых положительность оптимального значения целевой функции выводится из ее структуры. Более того, сложность решения таких задач оказалось зависимой от качества предварительного шкалирования, позволяющего найти правильную евклидову метрику, хорошо описывающую линии уровня целевой функции.

Еще одним важным примером являются задачи минимизации составных функций. Было показано что если аддитивная добавка к целевой функции является простой (т. е. она может быть непосредственно включена во вспомогательную задачу вычисления градиентного отображения), то ее невыгодные свойства, такие, как негладкость, разрывность, и т.д., не портят скорости сходимости методов минимизации. В этом случае скорость сходимости будет определяться свойствами основной части целевой функции, обычно задаваемой черноящичным оракулом.

В настоящей работе впервые удалось решить задачу построения методов второго порядка, обладающих глобальными гарантиями эффективности. Это было достигнуто с помощью построения кубической модели целевой функции, которая является суммой обычной квадратичной модели и кубического штрафа за отклонение от текущей точки. При правильном выборе коэффициента при штрафе такая модель дает верхнюю аппроксимацию целевой функции, минимизация которой определяет следующую точку итеративного процесса. Полученный метод может работать даже с невыпуклыми целевыми функциями и всегда сходится к стационарной точке, удовлетворяющей условию оптимальности второго порядка.

Было также показано, что информация полученная из "черного ящика позволяет ускорять методы выпуклой оптимизации, превышая даже максимально возможные скорости, полученные в теории сложности. По-видимому, в ближайшем будущем мы увидим дальнейшее уточнение моделей целевых функций, которое позволит эффективно решать задачи сверхбольших размеров, возникающих в больших социальных сетях (телекоммуникации, Интернет).

В настоящий момент теория выпуклой оптимизации впервые за долгое время оказалась не защищенной никакой теорией сложности. Для ее разработки и обновления, по-видимому, потребуются совсем другие принципы классификации задач оптимизации, разработка которых является одним из самых интересных направлений для будущих исследований.

Литература

[1] Александров А.Д. Выпуклые многогранники. - М.-Л.: Гостехиздат, 1950.

[2] Алексеев В.М., Тихомиров В.М., Фомин С.В. Оптимальное управление. - М.: Наука, 1979.

[3] Арутюнов А.В. Условия экстремума. Анормальные и вырожденные задачи. - М.: Факториал, 1997.

[4] Половинкин Е.С., Балашов М.В. Элементы выпуклого и сильно выпуклого анализа. - М.: Физматлит, 2004.

[5] Бляшке В. Круг и шар. - М.: Наука, 1967.

[6] Боннензен Т., Фенхель В. Теория выпуклых тел. - М.: ФАЗИС, 2002.

[7] Васильев Ф.П. Численные методы решения экстремальных задач. - М.: Наука, 1980.

[8] Галеев Э.М., Тихомиров В.М. Оптимизация: теория, примеры, задачи. - М.: Эдиториал УРСС, 2000.

[9] Гольштейн Е.Г., Третьяков Н.В. Модифицированная функция Лагранжа. Теория и методы оптимизации. - М.: Наука, 1989.

[10] Демьянов В.Ф. Негладкий анализ. - Л.: Изд. ЛГУ, 1982.

[11] Демьянов В.Ф., Рубинов А.М. Основы негладкого анализа и квазидифференциальное исчисление. - М.: Наука, 1990.

[12] Евтушенко Ю.Г. Методы решения экстремальных задач и их применение в системах оптимизации. - М.: Наука, 1982.

[13] Иоффе А.Д., Тихомиров В.М. Теория экстремальных задач. - М.: Наука, 1974.

[14] Мордухович - Б.Ш. Методы аппроксимации в задачах оптимизации и управления. - М.: Наука, 1988.

[15] Нестеров Ю.Е. Метод минимизации выпуклых функций со скоростью сходимости $O(1/k^2)$ // Докл. АН СССР - 1983 - Т.269, вып.3 - С. 543-547.

[16] Нестеров Ю.Е. О классе методов безусловной минимизации с высокой скоростью сходимости // Журн. выч.мат. и мат.физ. - 1984 - Т.24, вып.7 - С. 1090-1093.

[17] Нестеров Ю.Е. Метод линейного программирования с трудоемкостью $O(n^3L)$ операций // Экономика и мат. методы - 1988 - Т.24, вып.1 - С. 174-176.

[18] Нестеров Ю.Е. Полиномиальные методы для задач линейного и квадратичного программирования // Известия АН СССР - 1988 - вып.3 - С. 119 - 128.

[19] Нестеров Ю.Е. Об одном подходе к разработке оптимальных методов минимизации гладких выпуклых функций // Экономика и мат.методы - 1988 - Т.24, вып.3 - С. 509-517.

[20] Нестеров Ю.Е. Двойственные полиномиальные алгоритмы для линейного программирования // Кибернетика - 1989 - вып.1 - С. 34-54.

[21] Нестеров Ю.Е. Эффективные методы нелинейного программирования. М.: Радио и Связь, 1989 - 280с.

[22] Ю.Е.Нестеров. Введение в выпуклую оптимизацию. - М.: Изд. МЦНМО, 2010. - 263с.

[23] Пшеничный Б.Н. Выпуклый анализ и экстремальные задачи. - М.: Наука, 1980.

[24] Поляк Б.Т. Введение в оптимизацию. - М.: Наука, 1983.

[25] Рокафеллар Р.Т. Выпуклый анализ. - М.: Мир, 1973.

[26] Andersen S.P., de Palma A., Thisse J.-F. Discrete choice theory of product differentiation. - Cambridge: MIT Press, 1992.

[27] Anstreicher K.M. Ellipsoidal approximations of convex sets based on the volumetric barrier. CORE Discussion Paper 9745. - Louvain-la-Neuve: CORE, 1997.

[28] Anstreicher K., Wolsey L. On dual solutions in subgradient optimization // Mathematical Programming - 2011.

[29] d'Aspremont A., Banerjee O., El Ghaoui L. First-Order Methods for Sparse Covariance Selection // SIAM Journal on Matrix Analysis and its Applications - 2008 - V.30, N.1. - P.56-66.

[30] Auslender A., Teboulle M. Interior projection-like methods for monotone variational inequalities // Mathematical Programming - 2005 - V.104, N.1. - P.39-68.

[31] Baiocchi C., Capelo A. Variational and quasivariational inequalities: applications to Free boundary problems. - NY: Wiley, 1984.

[32] Barahona F., Anbil R. The volume algorithm: Producing primal solutions with a subgradient method // Mathematical Programming A - 2000 - V.87. - P.385–399.

[33] Beck A., Teboulle M. Mirror descent and nonlinear projected subgradient methods for convex optimization // Operations Research Letters - 2003 - V.31. - P. 167–175.

[34] Bennet A.A. Newton's method in general analysis // Proc. Nat. Ac. Sci. USA - 1916 - V.2, N.10. - P.592–598.

[35] Bensoussan A. Points de Nash dans le cas de fontionnelles quadratiques et jeux differentiels linéaires á N personnes // SIAM Journal on Control - 1974 - V.12. - P. 460Ц-499.

[36] Bensoussan A., Goursat M., Lions J.-L. Contrôle impulsionnel et inéquations quasi-variationnelle. // Compte rendu de l'Académie des Sciences Paris, Série A - 1973 - V.276, - P.1279–1284.

[37] Bliemer M., Bovy P. Quasi-Variational inequality formulation of the multiclass dynamic traffic assignment problem // Transportation Research B - 2003 - V.37. - P. 501–519.

[38] Ben-Tal A., Margalit T., Nemirovski A.. The Ordered Subsets Mirror Descent Optimization Method with Applications to Tomography // SIOPT - 2001 - V.12, N.1. - P. 79-108.

[39] Ben-Tal A., Nemirovskii A. Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. - Philadelphia: SIAM, 2001.

[40] Bertsekas D.P. Constrained optimization and Lagrange multiplier methods. - NY: Academic Press, 1982.

[41] Bienstock D. Potential Function Methods for Approximately Solving Linear Programming Problems. Theory and Practice. - Boston: Kluwer Academic Publishers, 2002.

[42] Chan D., Pang J.-S. The generalized quasi-variational inequality problem // Mathematics of Operations Research - 1982 - V.7, N.2. - P.211–222.

[43] Chen S., Donoho D., Saunders M. Atomic decomposition by basis pursuit // SIAM Journal of Scientific Computation - 1998 - V.20. - P.33-61.

[44] Claerbout J., Muir F. Robust modelling of eratic data. // Geophysics - 1973 - V.38. - P.826-844.

[45] Conn A.B., Gould N.I.M., Toint Ph.L. Trust Region Methods. - Philadelphia: SIAM, 2000.

[46] Dennis J.E., Schnabel R.B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. 2nd edition. - Philadelphia: SIAM, 1996.

[47] Figueiredo M., Novak R., Wright S.J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems // Submitted for publication.

[48] Fletcher R. Practical Methods of Optimization. V. 1, Unconstrained Minimization. - NY:John Wiley - 1980.

[49] Fukushima M., Mine H. A generalized proximal point algorithm for certain nonconvex problems // International Journal of Systems Sciences - 1981 - V.12, N.8. - P. 989-100.

[50] Ermoliev Yu.M. Methods for Solving Nonlinear Extremal Problems // Kibernetika - 1966 - V.4. - P.1-17.

[51] Goffin J.-L. On convergence rates of subgradient optimization methods // Mathematical Programming - 1977 - V.13. - P.329-347.

[52] Goldfeld S., Quandt R., Trotter H. Maximization by quadratic hill climbing // Econometrica - 1966 - V.34. - P. 541–551.

[53] Grigoriadis M.D., Khachiyan L.G. Fast approximation schemes for convex programs with many blocks and coupling constraints // SIAM J. Optimization - 1994 - V.4. - P.86–107.

[54] Grigoriadis M.D., Khachiyan L.G. Approximate minimum-cost multicommodity flows // Mathematical Programming - 1996 - V.75. - P.477–482.

[55] Harker P.T. Generalized Nash games and quasivariational inequalities // European Journal of Operations Research - 1991 - V.54. - P.81–94.

[56] Helly E. Über Systeme von abgeschlossenen Mengen mit gemeinschaftlichen Punkten // Mh. Math. Phys. - 1930 - V.37. - P.281-302.

[57] Helmberg C., Rendl F. A spectral bundle method for semidenite programming. Technical Report SC 97-37. - Berlin: Konrad-Zuse-Zentrum fur Informationstechnik, 1997.

[58] Helmberg C., Oustry F. Bundle methods to minimize the maximum eigenvalue function. In Lieven Vandenberghe R. Saigal and H. Wolkovicz, eds, *Hanbook on Semidefinite Programming. Theory, Algorithms and Applications.* - Boston: Kluwer Academic Publisher, 1999.

[59] Hiriart-Urruty J.-B., Lemarechal C. Convex Analysis and Minimization Algorithms. Part 1. A Series of Comprehensive Studies in Mathematics. - NY: Springer Verlag, 1993.

[60] John F. Extremum problems with inequalities as subsidiary conditions. In: *Studies and Essays, Presented to R. Courant on his 60th Birthday January 8, 1948.* - NY: Wiley Interscience, 1948. - P.187-204.

[61] Kantorovich L.V. Functional analysis and applied mathematics // Uspehi Matem. Nauk - 1948 - V.3, N.1. - P.89–185(in Russian). Translated as N.B.S. Report 1509, Washington D.C., 1952.

[62] Kantorovich L.V., Akilov G.P.. Functional analysis in normed spaces. - NY: Pergamon Press, 1964.

[63] Khachiyan L.G. Rounding of polytopes in the real number model of computation // Mathematics of Operations Research - 1996 - V.21 N.2. - P.307-320.

[64] M. Kocvara M., J.V. Outrata. On a class of quasi-variational inequalities // Optimization Methods и Software - 1995 - V.5. - P.275Ц-295.

[65] Korpelevich G. Extrapolation gradient methods and relation to Modified Lagrangeans // Ekonomika i Matematicheskie Metody - 1983 - V.19. - P.694-703 (in Russian.

[66] Kumar P., Yildirim E.A. Minimum volume enclosing ellipsoid and core sets // Journal of Optimization Theory and Applications - 2005 - V.126. N.1. - P.1-21.

[67] Kim S.-J., Koh K., Lustig M., Boyd S., Gorinevsky D. A method for large-scale l_1-regularized least-squares problems with applications in signal processing and statistics // Research Report, Stanford University - 2000 - March 20.

[68] Lemarechal C., Oustry F. Nonsmooth algorithms to solve semidefinite programs // L. El Ghaoui and S-I. Niculescu, eds. Recent Advances on LMI methods in Control - NY: Advances in Design and Control series, SIAM, 1999.

[69] Levenberg K. A method for the solution of certain problems in least squares // Quart. Appl. Math. - 1944 - V.2. - P.164–168.

[70] Levy S., Fullagar P. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution // Geophysics - 1981 - V.46. - P. 1235-1243.

[71] Lewis A.S., Sendov H.S. Twice Differentiable Spectral Functions // SIMAX - 2001 - V.23, N.2. - P.368-386.

[72] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters // SIAM J. Appl. Math. - 1963 - V.11. - P. 431–441.

[73] Miller A. Subset Selection in Regression. - London: Chapman and Hall, 2002.

[74] Minkowski H. Geometrie der Zahlen. - Leipzig-Berlin: Teubner, 1910.

[75] Mosco U. Implicit variational problems and quasi variational inequalities // Proc. Summer School (Bruxelles, 1975) 'Nonlinear operators and Calculus of variations', Lectures Notes Math, no. 543. - Berlin: Springer-Verlag, 1976. - P.83–156.

[76] Nayakkankuppam M.V., Tymofejev Y. A parallel implementation of the spectral bundle method for large-scale semidefinite programs // Proceedings of the 8th SIAM Conference on Applied Linear Algebra, - 2003 - Williamsburg (VA).

[77] Nemirovski A. Informational-based complexity of linear operator equations // Journal of Complexity - 1992 - V.8. P. 153-175.

[78] Nemirovski A. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems // SIAM J.Optim. - 2004 - V.15, N.1. - P. 229–251.

[79] Nemirovskij A.S., Yudin D.B. Problem complexity and method efficiency in optimization. - NY: Wiley-Interscience Series in Discrete Mathematics. A Wiley-Interscience Publication. John Wiley & Sons, 1983.

[80] Nesterov Yu. A method for unconstrained convex minimization problem with the rate of convergence $O(\frac{1}{k^2})$ // Doklady AN SSSR - 1983 - V.269. - P.543-547.

[81] Nesterov Yu. Semidefinite Relaxation and Nonconvex Quadratic Optimization // Optimization Methods and Software - 1998 - V.9. - P.141–160.

[82] Nesterov Yu. Minimizing функцияв with bounded variation of the gradient. CORE DP 2005/79 - Louvain-la-Neuve: CORE, 2005.

[83] Nesterov Yu. Introductory lectures on convex optimization. A basic course. - Boston: Kluwer, 2004.

[84] Nesterov Yu. Smooth minimization of non-smooth functions // Mathematical Programming - 2005 - V.103, N1.-P.127-152.

[85] Nesterov Yu. Excessive gap technique in non-smooth convex minimizarion // SIAM J. Optim. - 2005 - V.16, N1. - P.235-249.

[86] Nesterov Yu. Smoothing technique and its applications in semidefinite optimization // Mathematical Programming - 2007 - V.110, N2. - P.245-259.

[87] Nesterov Yu. Dual extrapolation and its application for solving variational inequalities and related problems // Mathematical Programming - 2007 - V.109, N2-3. - P.319-344.

[88] Nesterov Yu. Modified Gauss-Newton scheme with worst-case guarantees for its global performance // Optimization Methods and Software - 2007 - V.22, N3. - P.469-483.

[89] Nesterov Yu. Rounding of convex sets and efficient gradient methods for linear programming problems // Optimization Methods and Software - 2008 - V.23, N1. - P.109-128.

[90] Nesterov Yu. Accelerating the cubic regularization of Newton's method on convex problems // Mathematical Programming - 2008 - V.112, N1. - P.159-181.

[91] Nesterov Yu. Unconstrained Convex Minimization in Relative Scale // Mathematics of Operations Research - 2009 - V.34, N1. - P.180-193.

[92] Nesterov Yu. Primal-dual subgradient methods for convex problems // Mathematical Programming - 2009 - V.120, N1. - P.261-283.

[93] Nesterov Yu. Barrier subgradient method // Mathematical Programming - 2011 - V.127, N1. - P.31-56.

[94] Nesterov Yu. Gradient methods for minimizing composite functions // Mathematical Programming - 2013 - V.140, N1. - P.125-161.

[95] Nesterov Yu., Nemirovskii A. Interior point polynomial methods in convex programming: Theory and Applications. - Philadelphia: SIAM, 1994.

[96] Nesterov Yu., Polyak B. Cubic regularization of Newton's method and its global performance // Mathematical Programming - 2006 - V.108, N1. - P.177-205.

[97] Nesterov Yu., Scrimali L. Solving strongly monotone variational and quasi-variational inequalities // Discrete and Continuous Dynamical Systems - 2011 - V.31, N4. - P.1383-1296.

[98] Nesterov Yu., Todd M.J., Ye Y. Infeasible-start Primal-Dual Methods and Infeasibility Detectors // Mathematical Programming - 1999 - V.84, N.2. - P.227-267.

[99] Nesterov Yu., Vial J.-Ph. Homogeneous analytic center cutting plane methods for convex problems and variational inequalities // SIAM J.Optim. - 1999 - V.9, N.3. - P.707-728.

[100] Nesterov Yu., Vial J.-Ph. Confidence level solutions for stochastic programming // Automatica - 2008 - V.44, N.6. - P. 1559-1568.

[101] Nocedal J., Wright S.J. Numerical Optimization. - NY: Springer-Verlag, 1999.

[102] Noor M.A., Menon Z.A. Algorithms for general mixed quasi variational inequalities // Journal of Inequalities in Pure and Applied Mathematics - 2002 - V.3, N.4, - P.59.

[103] Noor M.A., Oettli W. On general nonlinear complementarity problems and quasi-equilibria // Le Matematiche - 1994 - V.XLIX. - P.313-331.

[104] Ortega J., Rheinboldt W. Iterative Solution of Nonlinear Equations in Several Variables. - NY: Academic Press, 1970.

[105] Oustry F. A second order bundle method to minimize the maximum eigenvalue функция // Mathematical Programming - 2000 - V.89. - P.1-33.

[106] Outrata J., Zowe J. A numerical approach to optimization problems with variational inequality constraints // Mathematical Programming - 1995 - V.68. - P.105–130.

[107] Pang J.-S., Fukushima M. Quasi-variational inequalities, generalized Nash equilibria and Multi-leader-follower games // Computational Management Science - 2005 - V.1. - P. 21–56.

[108] Polyak B.T. Gradient methods for minimization of functionals // USSR Comp. Math. Math. Phys. - 1963 - V.3, N.3. - P.643–653.

[109] Polyak B.T. A general method of solving extremum problems // Soviet Mathematics Doklady - 1967 - V.8. - P. 593–597.

[110] Polyak B.T. On Bertsekas' method for minimization of composite функция // Inter. Symp. Systems Opt. Analysis (A.Bensoussan и J.L.Lions, eds.). = Berlin: Springer, 1978. - P. 179-186 .

[111] Polyak B.T. Convexity of quadratic transformations and its use in control and optimization // J. Optim. Theory and Appl. - 1998 - V.99, N.3. - P.553–583.

[112] Polyak R.A. Nonlinear rescaling vs. smoothing technique in convex optimization // Mathematical Programming - 2002 - V.92. - P.197-235.

[113] Plotkin S.A., Shmoys D.B., Tardos E. Fast approximation algorithms for fractional packing and covering problems // Mathematics of Operations Research - 1995 - V.20. - P.257–301.

[114] Salahuddin. Projection methods for quasi-variational inequalities // Mathematical and Computational Applications - 2004 - V.9, N.2. - P.125–131.

[115] Santosa F., Symes V.W. Linear inversion of band-limited reflection histograms // SIAM Journal of Scientific and Statistical Computing - 1986 - V.7. - P.1307-1330.

[116] Shahrokhi F., Matula D.W. The maximum concurrent flow problem // Journal of the ACM - 1991 - V.37. - P.318-334.

[117] Shor N.Z. Minimization Methods for Nondifferentiable Functions. - Berlin: Springer-Verlag, 1985.

[118] Taylor H., Bank S., McCoy J. Deconvolution with the l_1 norm // Geophysics - 1979 - V.44. - P.39-52.

[119] Tibshirani R. Regression shrinkage and selection via the lasso // Journal Royal Statistical Society B - 1996 - V.58. - P. 267-288.

[120] Todd M.J., Yildirim E.A. On Khachiyan's algorithm for the computation of minimum volume enclosing ellipsoids. Technical Report, TR 1435. - Cornell University: School of Operations Research and Industrial Engineering, 2005.

[121] Tropp J. Just relax: convex programming methods for identifying sparse signals // IEEE Transactions on Information Theory - 2006 - V.51. - P.1030-1051.

[122] Vladimirov A., Nesterov Yu., Chekanov Yu. Uniformly convex функцияals // Vestnik Moskovskogo universiteta, ser. Vychislit. Matem. i Kibern.. - 1978 - V.4. - P. 18-27.

[123] Wei J.Y., Smeers Y. Spatial oligopolistic electricity models with Cournot generators and regulated transmission prices // Operations Research - 1999 - V.47. - P.102Ц-112.

[124] Wright S.J. Solving l_1-regularized regression problems. Talk at International Conference // Combinatorics and Optimization - 2007 - Waterloo.

[125] Yao J.C. The generalized quasi-variational inequality problem with applications // Journal of Mathematical Analysis and Applications - 1991 - V.158. - P.139–160.