МОДИФИЦИРОВАННАЯ ФУНКЦИЯ ЛАГРАНЖА*

М. В. Долгополик
maxim.dolgopolik@gmail.com

24 сентября 2015 г.

Аннотация. В докладе обсуждается один из способов построения модифицированной функции Лагранжа для задач нелинейного программирования. Данный способ опирается на теорию двойственности и позволяет распространить двойственные методы выпуклого программирования на случай невыпуклых задач.

1°. Двойственность в выпуклом программировании. Модифицированная функция Лагранжа для задач нелинейного программирования, обсуждаемая в данном докладе, появилась в связи с желанием распространить теорию двойственности (и двойственные методы) для задач выпуклого программирования на невыпуклый случай. Поэтому прежде чем перейти к рассмотрению модифицированной функции Лагранжа, мы кратко обсудим некоторые результаты теории двойственности для выпуклых экстремальных задач, важные для понимания невыпуклого случая.

Рассмотрим следующую задачу выпуклого программирования, которую мы будем называть *прямой задачей*:

$$f(x) \to \min \tag{1}$$
$$g_i(x) \leqslant 0, \quad i \in 1:m, \tag{2}$$
$$h_j(x) = 0, \quad j \in 1:\ell, \tag{3}$$
$$x \in X. \tag{4}$$

Здесь f, $g_i \colon \mathbb{R}^n \to \mathbb{R}$ — выпуклые функции, $h_i \colon \mathbb{R}^n \to \mathbb{R}$ — аффинные функции, $X \subset \mathbb{R}^n$ — непустое выпуклое множество. Обозначим через Ω множество планов прямой задачи.

---

*Семинар по конструктивному негладкому анализу и недифференцируемой оптимизации «CNSA & NDO»: `http://www.apmath.spbu.ru/cnsa/`

Для любых $\lambda_i \geqslant 0$, $i \in 1\colon m$ и $\mu_j \in \mathbb{R}$, $j \in 1\colon \ell$, введём *функцию Лагранжа* прямой задачи

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x).$$

Заметим, что неявно заданное (нефункциональное) ограничение $x \in X$ задачи (1)–(4) соответствует тем ограничениям исходной задачи, которые по какой-либо причине не включены в функцию Лагранжа.

Свяжем с задачей (1)–(4) *двойственную задачу* вида

$$\Theta(\lambda, \mu) \to \sup$$
$$\lambda \in \mathbb{R}^m_+,$$

где $\mathbb{R}^m_+ = \{\lambda \in \mathbb{R}^m \mid \lambda \geqslant 0\}$, а функция Θ выражается через функцию Лагранжа следующим образом:

$$\Theta(\lambda, \mu) = \inf_{x \in X} L(x, \lambda, \mu).$$

Функция Θ является полунепрерывной сверху вогнутой функцией, как точная нижняя грань семейства аффинных функций вида

$$(\lambda, \mu) \to f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x), \quad x \in X.$$

Таким образом двойственная задача является задачей вогнутого программирования.

Легко проверить, что для любого плана x прямой задачи и для любого плана (λ, μ) двойственной задачи справедливы неравенства

$$f(x) \geqslant L(x, \lambda, \mu) \geqslant \Theta(\lambda, \mu), \tag{5}$$

из которых вытекает, что оптимальное значение двойственной задачи всегда не превосходит оптимального значения прямой задачи. В случае когда эти оптимальные значения совпадают, то есть когда

$$\inf_{x \in \Omega} f(x) = \sup_{(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^\ell} \Theta(\lambda, \mu), \tag{6}$$

говорят, что между прямой и двойственной задачами *нулевой зазор двойственности*. В противном случае говорят, что существует (ненулевой) *зазор двойственности*. Отметим, что формулу (6) называют также *соотношением двойственности*.

Для задач выпуклого программирования существует простое необходимое и достаточное условие того, что зазор двойственности равен нулю.

ТЕОРЕМА 1 ([1, 2]). *Пусть оптимальное значение прямой задачи конечно. Для того чтобы между прямой и двойственной задачами был нулевой зазор двойственности, необходимо и достаточно, чтобы оптимальное значение двойственной задачи было конечно и функция чувствительности прямой задачи*

$$S(p,q) = \inf\{f(x) \mid x \in X\colon g(x) \leqslant p, h(x) = q\}$$

была полунепрерывна снизу в нуле. В частности, если в прямой задаче выполнено условие Слейтера, то между прямой и двойственной задачами нулевой зазор двойственности.

Следует заметить, что нулевой зазор двойственности является важным свойством для построения численных методов оптимизации. А именно, нулевой зазор двойственности является необходимым условием применимости *двойственных методов.*

Опишем кратко основной подход к построению двойственных методов. Следующая теорема является базовым результатом в данном подходе.

ТЕОРЕМА 2. *Пусть между прямой и двойственной задачами нулевой зазор двойственности. Предположим, что* (λ^*, μ^*) *— это оптимальный план двойственной задачи. Тогда для того чтобы план прямой задачи* x^**, удовлетворяющий условию дополнительности*

$$\lambda_i^* g_i(x^*) = 0 \quad \forall\, i \in 1\colon m,$$

был оптимальным планом, необходимо и достаточно, чтобы выполнялось равенство

$$L(x^*, \lambda^*, \mu^*) = \inf_{x \in X} L(x, \lambda^*, \mu^*) = \Theta(\lambda^*, \mu^*),$$

то есть чтобы вектор x^* *был точкой глобального минимума функции Лагранжа* $L(x, \lambda^*, \mu^*)$ *на множестве* X.

Доказательство. Поскольку между прямой и двойственной задачами нулевой зазор двойственности и (λ^*, μ^*) — это оптимальный план двойственной задачи, то

$$\Theta(\lambda^*, \mu^*) = \sup_{(\lambda,\mu) \in \mathbb{R}^m_+ \times \mathbb{R}^\ell} \Theta(\lambda, \mu) = \inf_{x \in \Omega} f(x). \tag{7}$$

Пусть план x^* прямой задачи удовлетворяет равенству $\Theta(\lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*)$. Воспользовавшись условием дополнительности и равенством (7), получим

$$\inf_{x \in \Omega} f(x) = \Theta(\lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) = f(x^*),$$

откуда вытекает, что x^* — это оптимальный план прямой задачи.

Пусть теперь x^* — это оптимальный план прямой задачи. Тогда учитывая (7), имеем

$$f(x^*) = \inf_{x \in \Omega} f(x) = \Theta(\lambda^*, \mu^*).$$

Отсюда с учётом условия дополнительности получим

$$L(x^*, \lambda^*, \mu^*) = f(x^*) = \Theta(\lambda^*, \mu^*),$$

что и требовалось доказать. $\square$

СЛЕДСТВИЕ 1. *Пусть между прямой и двойственной задачами нулевой зазор двойственности. Пусть также (λ^*, μ^*) — это оптимальный план двойственной задачи, а x^* — это оптимальный план прямой задачи, удовлетворяющий условию дополнительности*

$$\lambda_i^* g_i(x^*) = 0 \quad \forall i \ \in 1 \colon m.$$

Тогда

$$\inf_{x \in X} L(x, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) = \sup_{(\lambda,\mu) \in \mathbb{R}_+^m \times \mathbb{R}^\ell} L(x^*, \lambda, \mu),$$

то есть (x^*, λ^*, μ^*) — это седловая точка функции Лагранжа.

Замечание 1. Отметим, что достаточным условием существования оптимального плана двойственной задачи, в предположении что оптимальный план прямой задачи существует, является условие Слейтера. Действительно, если в прямой задаче выполнено условие Слейтера и x^* — это оптимальный план прямой задачи, то, как хорошо известно, существуют множители Лагранжа $\lambda_i^* \geqslant 0$ и $\mu_j^* \in \mathbb{R}$ такие, что

$$L(x, \lambda^*, \mu^*) \geqslant L(x^*, \lambda^*, \mu^*) \quad \forall\, x \in X,$$
$$\lambda_i^* g_i(x^*) = 0 \quad \forall\, i \in 1 \colon m.$$

Нетрудно проверить, что в этом случае вектор (λ^*, μ^*) будет оптимальным планом двойственной задачи. Отсюда, в частности, следует, что если оптимальный план двойственной задачи (λ^*, μ^*) единственен и функция Лагранжа $L(x, \lambda^*, \mu^*)$ строго выпукла по x, то её единственная точка глобального минимума (если она существует) будет оптимальным планом прямой задачи.

Из теоремы 2 вытекает следующая возможная схема решения прямой задачи:

1) Найти решение (λ^*, μ^*) двойственной задачи.

2) Найти точку минимума x^* функции Лагранжа $L(x, \lambda^*, \mu^*)$ на множестве X, которая является планом прямой задачи и удовлетворяет условию дополнительности. Тогда x^* — это оптимальный план прямой задачи.

Таким образом, подход основанный на двойственности позволяет свести задачу выпуклого программирования к задаче безусловной минимизации функции Лагранжа (точнее, к задаче минимизации функции Лагранжа на выпуклом множестве X), которая тоже является выпуклой. Однако, для этого необходимо сначала решить двойственную задачу, для вычисления значения целевой функции которой также требуется решать задачу минимизации функции Лагранжа. Значит, в двойственных методах решение исходной задачи оптимизации с ограничениями заменяется на многократное решение задачи минимизации функции Лагранжа без ограничений.

Разделение двойственного метода на два этапа — решение двойственной задачи и минимизация функции Лагранжа — является формальным. На практике, двойственная задача решается некоторым итеративным методом, где на каждой итерации необходимо вычислить значение двойственной функции $\Theta(\lambda^{(k)}, \mu^{(k)})$, т. е. найти минимум функции Лагранжа $x \to L(x, \lambda^{(k)}, \mu^{(k)})$, и затем по какому-либо правилу перейти к новому приближению множителей Лагранжа $(\lambda^{(k+1)}, \mu^{(k+1)})$. В итоге, в качестве приближённого решения прямой задачи берётся точка минимума функции Лагранжа, полученная на последней итерации данного метода.

Следует заметить, что при вычислении значения двойственной функции $\Theta(\lambda, \mu)$ необходимо найти вектор x такой, что $\Theta(\lambda, \mu) = L(x, \lambda, \mu)$. Можно показать, что, при небольших дополнительных предположениях, для любого такого x вектор

$$v = \big(g_1(x), \dots, g_m(x), h_1(x), \dots, h_\ell(x)\big)$$

будет суперградиентом вогнутой функции Θ в точке (λ, μ). Таким образом, в процессе вычисления значения функции Θ также можно вычислить и один из её суперградиентов в рассматриваемой точке. Поэтому для решения задачи максимизации функции Θ удобными оказываются, например, bundle методы (см., например, [3]).

Наконец, отметим, что двойственные методы являются эффективными только для определённых классов задач выпуклого программирования, в которых задача вычисления значения двойственной функции $\Theta(\lambda, \mu)$ (т. е. задача безусловной минимизации функции Лагранжа) является существенно более “простой” задачей, чем прямая задача. Одним из таких классов, является класс задач в которых целевая функция f и ограничения g_i являются *сепарабельными функциями*, то есть имеют вид

$$f(x) = \sum_{k=1}^{n} f_k(x_k), \quad g_i(x) = \sum_{k=1}^{n} g_{ik}(x_k) \quad \forall i \in 1\colon m,$$

где все функции f_k и g_{ik} выпуклы (конкретные примеры задач такого типа см. в [3, 4]). В данном случае, задача вычисления значения двойственной функции $\Theta(\lambda, \mu)$ сводится к решению n задач одномерной минимизации выпуклых функций вида

$$f_k(x_k) + \sum_{i=1}^{m} \lambda_i g_{ik}(x_k) + \sum_{j=1}^{\ell} \mu_j a_{ik} x_k$$

(напомним, что функции h_j аффинны и поэтому заведомо сепарабельны).

2°. Модифицированная функция Лагранжа. Двойственные методы, обсуждавшиеся в предыдущем разделе, предназначены в первую очередь для решения задач выпуклого программирования, хотя формально ничто не мешает применить такую же схему к задачам невыпуклого программирования. Однако, в невыпуклом случае данные методы, как правило[1], не дают желаемого результата, поскольку для невыпуклых задач зазор двойственности чаще всего не равен нулю. Данную проблему можно избежать, если использовать *модифицированную* функцию Лагранжа вместо стандартной.

Для того чтобы понять каким именно образом следует модифицировать стандартную функцию Лагранжа, чтобы сохранить свойство нулевого зазора двойственности в невыпуклом случае, нам необходимо обратиться к характеризации решений двойственной задачи в терминах функции чувствительности прямой задачи. Далее мы следуем идеям работы [5].

Напомним, что функцией чувствительности прямой задачи называется функция вида

$$S(p, q) = \inf\{f(x) \mid x \in X : g(x) \leqslant p, h(x) = q\} \quad (p, q) \in \mathbb{R}^m \times \mathbb{R}^\ell.$$

Покажем, что функция чувствительности в случае задачи выпуклого программирования является выпуклой.

ЛЕММА 1. *Функция чувствительности задачи* (1)–(4) *выпукла.*

[1] Следует заметить, что существуют исключения из этого правила. Например, двойственные методы выпуклого программирования часто эффективно работают для невыпуклых задач с сепарабельной целевой функцией и ограничениями. Данное явление объясняется неявной выпуклостью многих сепарабельных задач оптимизации, которая вытекает из Теоремы Шепли–Фолкмана. См. [2], Приложение 1.

Доказательство. Пусть векторы (p_1, q_1), $(p_2, q_2) \in \mathbb{R}^m \times \mathbb{R}^\ell$ и $\alpha \in (0,1)$ произвольны. Необходимо доказать справедливость неравенства

$$S(\alpha p_1 + (1-\alpha)p_2, \alpha q_1 + (1-\alpha)q_2) \leqslant \alpha S(p_1, q_1) + (1-\alpha)S(p_2, q_2). \tag{8}$$

Если $S(p_1, q_1) = +\infty$ или $S(p_2, q_2) = +\infty$, то данное неравенство очевидно. Поэтому будем предполагать, что $S(p_i, q_i) < +\infty$, $i \in 1\colon 2$.

Выберем произвольные $a, b \in \mathbb{R}$ такие, что $S(p_1, q_1) < a$ и $S(p_2, q_2) < b$. По определению функции чувствительности существуют $x_i \in X$, $i \in 1\colon 2$ такие, что $g(x_i) \leqslant p_i$, $h(x_i) = q_i$ и

$$S(p_1, q_1) \leqslant f(x_1) < a, \quad S(p_2, q_2) \leqslant f(x_2) < b.$$

Введём вектор $x = \alpha x_1 + (1-\alpha)x_2$. Воспользовавшись выпуклостью функций f и g_i и аффинностью функций h_j, получим, что

$$f(x) \leqslant \alpha f(x_1) + (1-\alpha)f(x_2) < \alpha a + (1-\alpha)b,$$
$$g(x) \leqslant \alpha p_1 + (1-\alpha)p_2, \quad h(x) = \alpha q_1 + (1-\alpha)q_2).$$

Отсюда имеем

$$S(\alpha p_1 + (1-\alpha)p_2, \alpha q_1 + (1-\alpha)q_2) \leqslant f(x) \leqslant \alpha a + (1-\alpha)b.$$

Переходя к пределу при a стремящемся к $S(p_1, q_1)$ и b стремящемся к $S(p_2, q_2)$, получаем справедливость неравенства (8). $\square$

Поскольку функция S выпукла, то для неё корректно определено понятие субдифференциала. Оказывается, что оптимальные планы двойственной задачи можно полностью охарактеризовать с помощью субдифференциала функции чувствительности в начале координат.

ТЕОРЕМА 3. *Предположим, что оптимальное значение прямой задачи конечно и между прямой и двойственной задачами нулевой зазор двойственности. Тогда для того чтобы вектор (λ^*, μ^*) являлся оптимальным планом двойственной задачи, необходимо и достаточно, чтобы*

$$(-\lambda^*, -\mu^*) \in \partial S(0,0).$$

Доказательство. Для любого вектора $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^\ell$ справедлива следующая цепочка равенств

$$\begin{aligned}
&\inf_{(p,q)\in\mathbb{R}^{m+\ell}} \big(S(p,q) + \langle(\lambda,\mu),(p,q)\rangle\big) = \\
&\quad = \inf_{(p,q)\in\mathbb{R}^{m+\ell}} \Big(\inf\Big\{f(x) + \langle(\lambda,\mu),(p,q)\rangle \;\Big|\; x \in X,\ g(x) \leqslant p,\ h(x) = q\Big\}\Big) = \\
&\quad = \inf_{x\in X} \Big(\inf\Big\{f(x) + \langle(\lambda,\mu),(p,q)\rangle \;\Big|\; (p,q) \in \mathbb{R}^{m+\ell},\ g(x) \leqslant p,\ h(x) = q\Big\}\Big) = \\
&\qquad = \inf_{x\in X} \Big(f(x) + \big\langle(\lambda,\mu),(g(x),h(x))\big\rangle\Big) = \inf_{x\in X} L(x,\lambda,\mu) = \Theta(\lambda,\mu).
\end{aligned} \tag{9}$$

Здесь

$$\big\langle (\lambda, \mu), (p, q) \big\rangle = \sum_{i=1}^{m} \lambda_i p_i + \sum_{j=1}^{\ell} \mu_j q_j.$$

Пусть (λ^*, μ^*) — это оптимальный план двойственной задачи. Тогда учитывая равенства (9) и тот факт, что между прямой и двойственной задачами нулевой зазор двойственности, имеем

$$S(0,0) = \inf_{x \in \Omega} f(x) = \Theta(\lambda^*, \mu^*) = \inf_{(p,q) \in \mathbb{R}^{m+\ell}} \Big(S(p,q) + \big\langle (\lambda^*, \mu^*), (p,q) \big\rangle \Big),$$

откуда следует, что

$$S(p,q) \geqslant S(0,0) - \big\langle (\lambda^*, \mu^*), (p,q) \big\rangle \quad \forall\, (p,q) \in \mathbb{R}^{m+\ell},$$

то есть $(-\lambda^*, -\mu^*) \in \partial S(0,0)$.

Пусть теперь $(-\lambda^*, -\mu^*) \in \partial S(0,0)$. Покажем сначала, что $\lambda^* \geqslant 0$. Действительно, пусть для некоторого $i \in 1\colon m$ будет $\lambda_i^* < 0$. Тогда для любого вектора $p \in \mathbb{R}^m$ такого, что $p_k = 0$ при $k \neq i$ и $p_i > 0$ справедливы неравенства

$$S(0,0) \geqslant S(p,0) \geqslant S(0,0) - \big\langle (\lambda^*, \mu^*), (p,0) \big\rangle = S(0,0) - \lambda_i^* p_i,$$

из которых следует неравенство $\lambda_i^* p_i \geqslant 0$, противоречащее предположению. Значит $\lambda^* \geqslant 0$. Здесь мы воспользовались включением

$$\{x \in X \mid g(x) \leqslant 0,\ h(x) = 0\} \subset \{x \in X \mid g(x) \leqslant p,\ h(x) = 0\},$$

непосредственно вытекающим из определения вектора p.

Снова применяя равенства (9) и пользуясь определением субградиента, получим, что

$$S(0,0) = \inf_{(p,q) \in \mathbb{R}^{m+\ell}} \Big(S(p,q) + \big\langle (\lambda^*, \mu^*), (p,q) \big\rangle \Big) = \Theta(\lambda^*, \mu^*),$$

то есть

$$\Theta(\lambda^*, \mu^*) = S(0,0) = \inf_{x \in \Omega} f(x).$$

Поскольку оптимальное значение двойственной задачи всегда не превосходит оптимального значения прямой задачи, то из последнего равенства следует, что (λ^*, μ^*) — это оптимальный план двойственной задачи. $\square$

Замечание 2. (i) Из второй части доказательства предыдущей теоремы вытекает, что если субдифференциал функции чувствительности в нуле непуст, то между прямой и двойственной задачами нулевой зазор двойственности и существует оптимальный план двойственной задачи. Таким образом, если оптимальное значение прямой задачи конечно, то для того чтобы между прямой и двойственной задачами был нулевой зазор двойственности и существовал оптимальный план двойственной задачи, необходимо и достаточно, чтобы субдифференциал функции чувствительности в нуле был непуст.

(ii) Отметим, что если функцию чувствительности прямой задачи определять с помощью равенства

$$\widehat{S}(p,q) = \inf\{f(x) \mid x \in X,\ g(x) + p \leqslant 0,\ h(x) + q = 0\},$$

(т. е. $\widehat{S}(p,q) = S(-p,-q)$), то оптимальные планы двойственной задачи будут в точности элементами субдифференциала функции $\widehat{S}$ в нуле.

Нетрудно заметить, что выпуклость прямой задачи не использовалась в доказательстве предыдущей теоремы. Поэтому данная теорема справедлива для любой задачи нелинейного программирования вида

$$f(x) \to \min \tag{10}$$

$$g_i(x) \leqslant 0, \quad i \in 1\colon m, \tag{11}$$

$$h_j(x) = 0, \quad j \in 1\colon \ell, \tag{12}$$

$$x \in X, \tag{13}$$

где f, g_i, $h_j \colon \mathbb{R}^n \to \mathbb{R}$ заданные функции, а $X \subset \mathbb{R}^n$ непустое множество. В частности, из предыдущей теоремы следует, что если оптимальное значение задачи (10)–(13) конечно, то для того чтобы между задачей (10)–(13) и двойственной к ней был нулевой зазор двойственной и существовал оптимальный план двойственной задачи, необходимо и достаточно, чтобы субдифференциал функции чувствительности $\partial S(0,0)$ был непуст. Здесь

$$\partial S(0,0) = \big\{(\lambda,\mu) \in \mathbb{R}^{m+\ell} \mid S(p,q) \geqslant S(0,0) + \langle(\lambda,\mu),(p,q)\rangle \quad \forall\, (p,q) \in \mathbb{R}^{m+\ell}\big\},$$

то есть субдифференциал определяется точно также, как и в выпуклом случае, несмотря на возможное отсутствие выпуклости[2].

В случае, когда задача (10)–(13) не является задачей выпуклого программирования, её функция чувствительности, вообще говоря, не является выпуклой, и поэтому субдифференциал $\partial S(0,0)$ может оказаться пустым. Данный факт не позволяет напрямую воспользоваться двойственными методами выпуклого программирования в невыпуклом случае.

[2]Нетрудно проверить, что если определённым таким образом субдифференциал $\partial S(0,0)$ непуст, то справедливо равенство $\partial S(0,0) = \partial S^{**}(0,0)$, где S^{**} — это дважды сопряжённая функция к функции S (см., например, [2]). Отметим, что S^{**} является наибольшей собственной выпуклой функцией, не превосходящей функцию S.

ПРИМЕР 1. Рассмотрим следующую задачу математического программирования

$$-x_1^2 + x_2^2 \to \min \tag{14}$$

$$x_1 = 0, \quad x_2 \leqslant 0. \tag{15}$$

Функция чувствительности данной задачи имеет вид

$$S(p,q) = \inf\big\{ -x_1^2 + x_2^2 \mid x_2 \leqslant p,\ x_1 = q\big\} = -q^2 + \min\{p, 0\}^2 \quad \forall\, (p,q) \in \mathbb{R}^2.$$

Отсюда, учитывая что неравенство

$$-q^2 \geqslant \mu q \quad \forall\, q \in \mathbb{R}$$

не выполнено ни для какого $\mu \in \mathbb{R}$, получаем, что $\partial S(0,0) = \varnothing$. Заметим также, что для данной задачи $\Theta(\lambda, \mu) \equiv -\infty$.

Поскольку субдифференциал (в смысле выпуклого анализа) функции чувствительности в невыпуклом случае может оказаться пустым, попытаемся рассмотреть какой-либо другой субдифференциал данной функции. Например, можно ослабить определение субдифференциала следующим образом.

Зафиксируем произвольную выпуклую функцию $\sigma \colon \mathbb{R}^{m+\ell} \to [0, +\infty)$ такую, что $\sigma(p,q) = 0$ тогда и только тогда, когда $(p,q) = 0$. Для любого $r > 0$ введём множество

$$\partial_{\sigma,r} S(0,0) = \Big\{ (\lambda,\mu) \in \mathbb{R}^{m+\ell} \ \Big| \ S(p,q) \geqslant S(0,0) + \big\langle (\lambda,\mu), (p,q) \big\rangle - r\sigma(p,q) \quad \forall\, (p,q) \in \mathbb{R}^{m+\ell} \Big\}.$$

В частности, при $r = 0$ будет $\partial_{\sigma,r} S(0,0) = \partial S(0,0)$. Заметим также, что для любых $0 < r_1 < r_2$ будет

$$\partial S(0,0) \subseteq \partial_{\sigma,r_1} S(0,0) \subseteq \partial_{\sigma,r_2} S(0,0),$$

то есть множество $\partial_{\sigma,r} S(0,0)$ “увеличивается” при возрастании r.

Замечание 3. По поводу простой геометрической интерпретации субдифференциала $\partial_{\sigma,r} S(0,0)$ и его связи с нулевым зазором двойственности см. [1], параграф 9.3 и [5], параграф 11.K.

Субдифференциал $\partial_{\sigma,r} S(0,0)$ может оказаться непуст для гораздо более широкого класса задач нелинейного программирования, чем субдифференциал в смысле выпуклого анализа. Однако, возникает естественный вопрос: какой смысл несут в себе элементы субдифференциала $\partial_{\sigma,r} S(0,0)$? Для того чтобы ответить на этот вопрос необходимо обратиться к равенствам (9). Заменим в них выражение

$$S(p,q) + \big\langle (\lambda,\mu), (p,q) \big\rangle,$$

отвечающее субдифференциалу в смысле выпуклого анализа, на выражение

$$S(p,q) + \big\langle(\lambda,\mu),(p,q)\big\rangle + r\sigma(p,q),$$

соответствующее субдифференциалу $\partial_{\sigma,r}S(0,0)$. В итоге получим,

$$\begin{aligned}
&\inf_{(p,q)\in\mathbb{R}^{m+\ell}} \big(S(p,q) + \big\langle(\lambda,\mu),(p,q)\big\rangle + r\sigma(p,q)\big) = \\
&= \inf_{(p,q)\in\mathbb{R}^{m+\ell}} \Big(\inf \Big\{ f(x) + \big\langle(\lambda,\mu),(p,q)\big\rangle + r\sigma(p,q) \;\Big|\; x\in X, g(x)\leqslant p, h(x)=q \Big\}\Big) = \\
&= \inf_{x\in X} \Big(\inf \Big\{ f(x) + \big\langle(\lambda,\mu),(p,q)\big\rangle + r\sigma(p,q) \;\Big|\; (p,q)\in\mathbb{R}^{m+\ell}, g(x)\leqslant p, h(x)=q \Big\}\Big).
\end{aligned} \tag{16}$$

Обозначим

$$\begin{aligned}
&L(x,\lambda,\mu,r) = \\
&\quad = \inf \Big\{ f(x) + \big\langle(\lambda,\mu),(p,q)\big\rangle + r\sigma(p,q) \;\Big|\; (p,q)\in\mathbb{R}^{m+\ell},\ g(x)\leqslant p,\ h(x)=q \Big\}.
\end{aligned}$$

Тогда равенства (16) примут вид

$$\inf_{(p,q)\in\mathbb{R}^{m+\ell}} \big(S(p,q) + \big\langle(\lambda,\mu),(p,q)\big\rangle + r\sigma(p,q)\big) = \inf_{x\in X} L(x,\lambda,\mu,r). \tag{17}$$

Нетрудно заметить, что в случае $\sigma \equiv 0$ (или $r = 0$) функция $L(x,\lambda,\mu,r)$ преобразуется в стандартную функцию Лагранжа

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x).$$

Таким образом, функция $L(x,\lambda,\mu,r)$ представляется из себя *модифицированную функцию Лагранжа*[3] со штрафным параметром $r > 0$.

По аналогии со случаем стандартной функции Лагранжа, определим модифицированную двойственную целевую функцию, соответствующую модифицированной функции Лагранжа $L(x,\lambda,\mu,r)$, следующим образом:

$$\Theta(\lambda,\mu,r) = \inf_{x\in X} L(x,\lambda,\mu,r).$$

Введём также модифицированную двойственную задачу

$$\begin{gathered}
\Theta(\lambda,\mu,r) \to \sup \\
\lambda \geqslant 0, \quad r > 0.
\end{gathered}$$

[3]В англоязычной литературе данная функция называется “augmented Lagrangian”, то есть увеличенная (или расширенная) функция Лагранжа.

Нетрудно проверить, что модифицированная функция Лагранжа $L(x, \lambda, \mu, r)$ вогнута и полунепрерывна сверху по переменным (λ, μ, r). Поэтому функция $\Theta(\lambda, \mu, r)$ также вогнута и полунепрерывна сверху. Следовательно, модифицированная двойственная задача является задачей вогнутого программирования.

Как и в случае стандартной двойственной задачи, оптимальное значение модифицированной двойственной задачи всегда не превосходит оптимального значения исходной задачи нелинейного программирования (10)–(13).

ЛЕММА 2. *Пусть x — это допустимый план задачи* (10)–(13)*, а (λ, μ, r) — это допустимый план модифицированной двойственной задачи. Тогда*

$$f(x) \geqslant L(x, \lambda, \mu, r) \geqslant \Theta(\lambda, \mu, r).$$

Доказательство. Поскольку x — это допустимый план задачи (10)–(13), то $g(x) \leqslant 0$ и $h(x) = 0$. Следовательно,

$$\begin{aligned} &L(x, \lambda, \mu, r) = \\ &= \inf \Big\{ f(x) + \big\langle (\lambda, \mu), (p, q) \big\rangle + r\sigma(p, q) \;\Big|\; (p, q) \in \mathbb{R}^{m+\ell},\ g(x) \leqslant p,\ h(x) = q \Big\} \leqslant \\ &\qquad\qquad \leqslant f(x) + \big\langle (\lambda, \mu), (0, 0) \big\rangle + r\sigma(0, 0) = f(x). \end{aligned}$$

В свою очередь, неравенство $L(x, \lambda, \mu, r) \geqslant \Theta(\lambda, \mu, r)$ непосредственно вытекает из определения двойственной целевой функции. □

Если оптимальные значения задачи (10)–(13) и модифицированной двойственной задачи совпадают, то есть, если

$$\begin{aligned} &\inf \big\{ f(x) \;\big|\; x \in X,\ g(x) \leqslant 0,\ h(x) = 0 \big\} = \\ &\qquad\qquad = \sup \big\{ \Theta(\lambda, \mu, r) \;\big|\; \lambda \geqslant 0,\ \mu \in \mathbb{R}^{\ell},\ r > 0 \big\}, \end{aligned}$$

то будем говорить, что между задачей (10)–(13) и модифицированной двойственной задачей *нулевой зазор двойственности*.

Оказывается, что теорема 1 о необходимом и достаточном условии того, чтобы между прямой и двойственной задачами выпуклого программирования был нулевой зазор двойственности, может быть без изменения перенесена на случай задачи нелинейного программирования и модифицированной двойственной задачи.

ТЕОРЕМА 4 ([6])**.** *Пусть оптимальное значение задачи* (10)–(13) *конечно. Для того чтобы между задачей* (10)–(13) *и модифицированной двойственной задачей был нулевой зазор двойственности, необходимо и достаточно,* *чтобы оптимальное значение модифицированной двойственной задачи было конечно и функция чувствительности*

$$S(p,q) = \inf\big\{f(x) \mid x \in X,\ g(x) \leqslant p,\ h(x) = q\big\},$$

задачи (10)–(13) *была полунепрерывна снизу в нуле.*

Теперь мы можем ответить на вопрос о том, какой смысл в себе несут элементы субдифференциала $\partial_{\sigma,r}S(0,0)$. Повторяя доказательство теоремы 3 с заменой равенств 9 на определение модифицированной функции Лагранжа, получаем справедливость следующего утверждения.

ТЕОРЕМА 5. *Пусть оптимальное значение задачи* (10)–(13) *конечно и между задачей* (10)–(13) *и модифицированной двойственной задачей нулевой зазор двойственности. Тогда для того, чтобы вектор* (λ^*,μ^*,r^*) *был оптимальным планом модифицированной двойственной задачи, необходимо и достаточно, чтобы* $(-\lambda^*,-\mu^*) \in \partial_{\sigma,r^*}S(0,0)$.

СЛЕДСТВИЕ 2. *Пусть оптимальное значение задачи* (10)–(13) *конечно. Тогда для того чтобы между задачей* (10)–(13) *и модифицированной двойственной задачей был нулевой зазор двойственности и существовал оптимальный план модифицированной двойственной задачи, необходимо и достаточно, чтобы* $\partial_{\sigma,r}S(0,0) \neq \varnothing$ *для некоторого* $r > 0$.

Модифицированная функция Лагранжа может быть использована для построения двойственных методов решения задач нелинейного программирования. Поскольку субдифференциал $\partial_{\sigma,r}S(0,0)$ непуст для более широкого класса задач, чем субдифференциал в смысле выпуклого анализа, то двойственные методы основанные на модифицированной функции Лагранжа применимы к более широкому классу задач, чем двойственные методы выпуклого программирования.

Приведём здесь теорему, аналогичную теореме 2, на основе которой можно построить двойственные методы решения задач нелинейного программирования, использующие модифицированную функцию Лагранжа.

ТЕОРЕМА 6 ([5])**.** *Пусть оптимальное значение задачи* (10)–(13) *конечно и между данной задачей и модифицированной двойственной задачей нулевой зазор двойственности. Предположим, что* (λ^*,μ^*,r^*) *— это оптимальный план двойственной задачи. Тогда для любого* $r > r^*$ *будет*

$$\inf_{x\in\Omega} f(x) = \inf_{x\in X} L(x,\lambda^*,\mu^*,r),$$
$$\operatorname*{arg\,min}_{x\in\Omega} f(x) = \operatorname*{arg\,min}_{x\in X} L(x,\lambda^*,\mu^*,r).$$

З а м е ч а н и е 4. Пусть оптимальное значение задачи (10)–(13) конечно и между данной задачей и модифицированной двойственной задачей нулевой зазор двойственности. Из предыдущей теоремы следует, что если (λ^*, μ^*, r^*) — это оптимальный план модифицированной двойственной задачи, то для любого $r > r^*$ вектор (λ^*, μ^*, r) также будет оптимальным планом модифицированной двойственной задачи. Следовательно, выбрав достаточно большое $r > 0$ вместо модифицированной двойственной задачи можно решать задачу

$$\Theta(\lambda, r) \to \sup_{\lambda}$$
$$\lambda_i \geqslant 0, \quad i \in 1 : m.$$

Таким образом, переменная $r > 0$ играет роль, схожую с ролью штрафного параметра штрафной функции.

Приведём два конкретных примера модифицированных функций Лагранжа.

ПРИМЕР 2. Пусть

$$\sigma(p, q) = \frac{1}{2}\big\|(p, q)\big\|^2 = \frac{1}{2}\sum_{i=1}^{m} p_i^2 + \frac{1}{2}\sum_{j=1}^{\ell} q_j^2.$$

Для того чтобы выписать модифицированную функцию Лагранжа для данной функции σ в явном виде необходимо решить следующую задачу:

$$\big\langle (\lambda, \mu), (p, q) \big\rangle + \frac{r}{2}\big\|(p, q)\big\|^2 \to \inf_{p,q} \tag{18}$$

$$p \geqslant g(x), \quad q = h(x). \tag{19}$$

Поскольку функция $\lambda_i p_i + r p_i^2/2$ достигает глобального минимума на множестве $[g_i(x), +\infty)$ в точке

$$\max\left\{ g_i(x), -\frac{\lambda_i}{r} \right\},$$

то оптимальным планом задачи (18)–(19) является вектор

$$\left(\max\left\{ g_1(x), -\frac{\lambda_1}{r} \right\}, \ldots, \max\left\{ g_m(x), -\frac{\lambda_1}{r} \right\}, h_1(x), \ldots, h_\ell(x) \right).$$

Поэтому модифицированная функция Лагранжа имеет вид

$$L(x, \lambda, \mu, r) = f(x) + \sum_{i=1}^{m} \lambda_i \max\left\{ g_i(x), -\frac{\lambda_i}{r} \right\} +$$
$$+ \frac{r}{2}\sum_{i=1}^{m} \max\left\{ g_i(x), -\frac{\lambda_i}{r} \right\}^2 + \langle \mu, h(x) \rangle + \frac{r}{2}\|h(x)\|^2$$

для всех $x \in X$, $\lambda \geqslant \mathbb{R}^m_+$, $\mu \in \mathbb{R}^\ell$ и $r > 0$. Заметим, что в случае когда ограничения типа неравенств отсутствуют, данная модифицированная функция Лагранжа принимает особенно простой вид

$$L(x, \mu, r) = f(x) + \langle \mu, h(x) \rangle + \frac{r}{2} \|h(x)\|^2$$

для всех $x \in X$, $\mu \in \mathbb{R}^\ell$ и $r > 0$.

ПРИМЕР 3. Пусть

$$\sigma(p, q) = |p_1| + \ldots + |p_m| + |q_1| + \ldots + |q_m|.$$

В данном случае задача вычисления значения модифицированной функции Лагранжа сводится к решению следующей задачи:

$$\sum_{i=1}^{m} (\lambda_i p_i + r|p_i|) + \sum_{j=1}^{\ell} (\mu_j q_j + r|q_j|) \to \inf_{p,q}$$
$$p \geqslant g(x), \quad q = h(x).$$

Нетрудно проверить, что

$$\operatorname*{arg\,min}_{p_i \in [g_i(x), +\infty)} (\lambda_i p_i + r|p_i|) = \begin{cases} \{\max(g_i(x), 0)\}, & \text{если } r \geqslant \lambda_i, \\ \{g_i(x)\}, & \text{если } 0 < r < \lambda_i. \end{cases}$$

Обозначим

$$I_1(\lambda, r) = \big\{ i \in 1\colon m \mid r \geqslant \lambda_i \big\}, \quad I_2(\lambda, r) = \big\{ i \in 1\colon m \mid r < \lambda_i \big\}.$$

Тогда

$$L(x, \lambda, \mu, r) = f(x) + \sum_{i \in I_1(\lambda, r)} \big(\lambda_i + r\big) \max\{g_i(x), 0\} +$$
$$+ \sum_{i \in I_2(\lambda, r)} \big(\lambda_i g_i(x) + r|g_i(x)|\big) + \sum_{j=1}^{\ell} \big(\mu_j h_j(x) + r|h_j(x)|\big).$$

для всех $x \in X$, $\lambda \geqslant \mathbb{R}^m_+$, $\mu \in \mathbb{R}^\ell$ и $r > 0$.

ЛИТЕРАТУРА

1. Bazaraa M.S., Sherali H.D., Shetty C.M. *Nonlinear Programming: Theory and Algorithms.* New Jersey, John Wiley & Sons, Inc., 2006.

2. Экланд И., Темам Р. *Выпуклый анализ и вариационные проблемы.* М.: Мир, 1979. 399 с.

3. Hiriart-Urruty J.B., Lemaréchal C. *Convex Analysis and Minimization Algorithms II.* Berlin Heidelberg, Spriger-Verlag, 1993.

4. Patriksson M. *A survey on the continuous nonlinear resource allocation problem* // European Journal of Operational Research, 2008, vol. 185, pp. 1–46.

5. Rockafellar R.T., Wets R.J.-B. *Variational Analysis.* Berlin Heidelberg, Springer-Verlag, 1998.

6. Huang X.X., Yang X.Q. *Further Study on Augmented Lagrangian Duality Theory* // Journal of Global Optimization, 2005, vol. 31, pp. 193–210.