Н. М. Новикова

ОСНОВЫ ОПТИМИЗАЦИИ

(курс лекций)

МОСКВА 1998

Ответственный редактор
академик РАН П. С. Краснощеков

В сжатой форме дается изложение основ теории сложности, линейного программирования (ЛП) — с описанием полиномиальных алгоритмов, целочисленного ЛП, математического программирования (необходимые условия экстремума при ограничениях-неравенствах, локальные методы безусловной оптимизации, метод штрафов, идеи глобальной оптимизации), схем методов динамического программирования и ветвей и границ.

Работа написана на базе семестрового курса лекций, читаемого автором студентам 4-го курса программистского потока факультета ВМиК МГУ, с учетом дополнений и замечаний, указанных студентами. Автор благодарит всех студентов, содействовавших изданию этого курса и предложивших исправления, способствующие его улучшению, в том числе, Ласкавого Сергея, Санникова Андрея и Свахина Николая. Замеченные опечатки и неточности просьба сообщать автору по адресу nnovik@ccas.ru

Работа частично поддержана грантом РФФИ No.96-01-00786.

Рецензенты: С. К. Завриев,
А. В. Лотов

3. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ (МП)

Литература:

4. Карманов В. Г. Математическое программирование. М.: Наука, 1986.

5. Сухарев А. Г., Тимохов А. В., Федоров В. В. Курс методов оптимизации. М.: Наука, 1985.

6. Мину М. Математическое программирование. М.: Наука, 1990.

§8. Обзор идей МП

Классификация задач МП. Преимущества выпуклого случая. Понятие о градиентных и Ньютоновских методах минимизации. Условная оптимизация, способы освобождения от ограничений (методы барьеров и штрафов).

1. Задача ЛП, как и задача минимизации функции Кармаркара, является частным случаем задачи МП:

$$\min_{x \in X} f(x). \tag{1}$$

Здесь требуется найти $\arg\min_{x \in X} f(x) \in \mathrm{Arg}\min_{x \in X} f(x)$, т.е.

$$x^* \in X^* \doteq \{x^* \in X \mid f(x^*) \le f(x) \quad \forall x \in X\}, \quad \text{и } f^* = f(x^*). \tag{2}$$

Любой такой x^* называется *решением* (1); f^* — *значение* (1), или *оптимальное значение* целевой функции f в задаче (1), X — *множество ограничений* или *допустимое множество*.

В зависимости от природы множества X задачи оптимизации классифицируются как: дискретные (комбинаторные) — X конечно или счетно, целочисленные — $x_j \in \mathbf{Z}$, булевы — $x_j \in \mathbf{B}$, вещественные (непрерывные) — $X \subseteq \mathbf{R^n}$, бесконечномерные или в функциональном пространстве, например, когда X — подмножество гильбертова пространства $\mathbf{L_2}$, и т.п. В данном разделе будем по преимуществу рассматривать задачи с вещественными переменными, которые собственно и называются (традиционно) *задачами математического программирования (ЗМП)*. Если $X \subset \mathbf{R^n}$, то говорим о задаче *условной* оптимизации (при условии $x \in X$), иначе ($X = \mathbf{R^n}$) получаем задачу *безусловной* оптимизации.

Для ЗМП минимум в (1) достигается в условиях теоремы Вейерштрасса (f непрерывна, X компактно или для некоторого $\hat{x} \in X$ ограничено *множество Лебега* функции f — $\{x \in X | f(x) \leq f(\hat{x})\}$).

Кроме разделения на условные и безусловные, ЗМП классифицируются по свойствам целевой функции и множества ограничений соответственно на задачи ЛП, выпуклого программирования, гладкие или негладкие и др. Для каждого из классов ЗМП разрабатываются свои численные методы их решения. С точки зрения численных методов существенно также разделение на *локальную* и *глобальную* оптимизацию. В определении (2) речь идет о глобальном минимуме, который, однако, найти не просто, и поэтому задачу стараются свести к дискретной оптимизации на множестве локальных минимумов.

Определение 1. Точка $x^0 \in X$ называется точкой *локального* минимума в ЗМП (1), если $\exists \varepsilon > 0 : \quad f(x^0) \leq f(x) \quad \forall x \in X \cap \mathbf{O}_\varepsilon(x^0)$. Здесь и далее $\mathbf{O}_\varepsilon(x)$ обозначает ε-окрестность точки x.

Для поиска локального минимума применяются специальные методы, которые при определенных предположениях оказываются эффективными. Тогда как общая задача глобальной оптимизации является **NP**-трудной. Действительно к ней сводится **NP**-полная.

Утверждение 1. **ЦЛН**$\propto$ЗМП.

Доказательство. Поскольку задача ЛН является частным случаем задачи ЛП, то для сведения **ЦЛН** к ЗМП достаточно представить условие целочисленности переменных в виде ограничений (неравенств) на вещественные переменные, что нетрудно сделать, например, так: $\{x_j \in \mathbf{Z}\}$ эквивалентно $\{x_j \in \mathbf{R} | \sin^2(\pi x_j) \leq 0\}$.

Поэтому методы глобальной оптимизации будут рассмотрены в разд.4, а в данном параграфе остановимся на поиске локального экстремума. Отметим, что для ряда экстремальных постановок задач физики точки локального экстремума имеют самостоятельное значение. Кроме того, существует целый класс ЗМП, для которого локальный экстремум совпадает с глобальным минимумом, это — задачи выпуклого программирования.

Определение 2. Функция f называется *выпуклой на* X, если ее надграфик $\mathrm{epigraf}_X f \doteq \{(x, y) | \ y \geq f(x), \ x \in X\}$ — выпуклое множество. Функция, выпуклая на всей области определения, называется выпуклой. Множество называется выпуклым, если вместе с любыми двумя своими точками оно содержит отрезок, их соединяющий.

Утверждение 2. Любая точка локального минимума выпуклой функции является точкой ее глобального минимума.

Доказательство. Пусть $f(x^0) > f(x^*)$. Тогда $f(x^0) > f(x)$ для всех точек x полуинтервала $(x^0, x^*]$ (по определению 2), а значит, и в некоторой окрестности x^0 — противоречие с определением 1.

Для решения задач выпуклого программирования приме́ним метод эллипсоидов, причем в гладком случае отсечение полуэллипсоида проводится на основе градиента невыполненного ограничения в полной аналогии с алгоритмом из §6. Поэтому задача поиска ε-приближенного решения задачи выпуклого программирования оказывается полиномиально разрешимой. Для *острых* задач выпуклого программирования — когда функция цели убывает в окрестности минимума не медленнее некоторой линейной функции — можно получить и точное решение.

2. Общими методами локальной оптимизации (для произвольного, не обязательно выпуклого, случая) являются *методы локального спуска.*

Определение 3. Вектор $h \in \mathbf{R^n}$ называется *направлением убывания* функции f в точке x, если $f(x+\alpha h) < f(x)$ для всех достаточно малых $\alpha > 0$.

Утверждение 3. Пусть f дифференцируема в точке x. Тогда, если $\langle \mathrm{grad} f(x), h\rangle < 0$, то h — направление убывания функции f в точке x, а если h — направление убывания функции f в точке x, то $\langle \mathrm{grad} f(x), h\rangle \le 0$.

Доказательство. Из условия дифференцируемости f имеем для достаточно малых $\alpha > 0$: $f(x+\alpha h) - f(x) = \langle \mathrm{grad} f(x), \alpha h\rangle + o(\alpha) = \alpha\{\langle \mathrm{grad} f(x), h\rangle + o(\alpha)/\alpha\}$. Очевидно, последняя добавка не изменит знака выражения в фигурных скобках, если скалярное произведение строго отрицательно или строго положительно. Отсюда автоматически вытекает требуемое утверждение.

Таким образом, направление локального убывания дифференцируемой функции должно составлять острый угол с ее антиградиентом, который является в смысле линейного приближения наилучшим направлением убывания. Для мнемоники приведем эпиграф к главе, посвященной градиентным методам минимизации, из 1-го издания книги Ф. П. Васильева *Численные методы решения экстремальных задач*: "Вот кто-то с горочки спустился — антиградиент!"

Если $\mathrm{grad} f(x) = 0$, то x будет *стационарной точкой*. Отметим, что в условной оптимизации равенство нулю градиента уже не является необходимым условием минимума (соответствующие условия будут рассмотрены в §9). Но в более простом случае $X = \mathbf{R^n}$ можно, двигаясь небольшими шагами в направлении антиградиента функции f в текущей точке, прийти в стационарную точку, как правило, локального минимума. Так мы получаем идею *градиентного метода безусловной минимизации*, задаваемого итеративной процедурой

$$x^{t+1} = x^t - \alpha_t \mathrm{grad} f(x^t), \quad t = 1, 2, \ldots, \quad \forall x^1 \in \mathbf{R^n}.$$

Параметр α_t называется шаговым множителем и может выбираться, исходя из различных соображений, разными способами.

1) *Пассивные способы* — $\{\alpha_t\}$ выбирается заранее.
Постоянный шаг — $\alpha_t = \alpha_0$ для достаточно малых α_0.
Убывающий шаг (если α_0 не известно или при наличии помех) — $\alpha_t \downarrow 0, \ \sum \alpha_t = \infty, \ \sum \alpha_t^2 < \infty$, например $\alpha_t = 1/t$.

2) *Адаптивные способы* — $\{\alpha_t\}$ зависит от реализующейся $\{x^t\}$.
Метод скорейшего спуска — $\alpha_t \in \mathrm{Arg}\min_{\alpha > 0} f(x^t - \alpha\, \mathrm{grad} f(x^t))$.
Метод дробления шага (деления пополам) — если $f(x^{t+1}) > f(x^t)$, то возврат к t-й итерации с $\alpha_t := \alpha_t/2$. (Возможно и увеличение шага при стабильном убывании f, т.е. приближенный скорейший спуск.)
Правило Армихо — путем дробления шага добиваемся для α_t выполнения условия $f(x^t - \alpha_t \mathrm{grad} f(x^t)) - f(x^t) \le -\varepsilon \alpha_t \|\mathrm{grad} f(x^t)\|^2$.

В общем случае дифференцируемой, ограниченной снизу f можно получить сходимость градиентного метода к множеству стационарных точек, а при дополнительных предположениях доказывается (за исключением варианта с убывающим шагом) *линейная скорость сходимости*, которая в выпуклых задачах означает $\|x^{t+1} - x^*\| \le q\|x^t - x^*\|$ для некоторого $0 < q < 1$. Указанная линейная оценка объясняется тем, что в процессе минимизации градиентным методом используется линейная аппроксимация целевой функции на каждом шаге. Более высокую скорость сходимости получают для методов, основанных на квадратичной аппроксимации, в предположении дважды дифференцируемости f. Типичным примером здесь является *метод Ньютона*.

Пусть $f \in \mathbf{C^2}(\mathbf{R^n})$, разложим функцию f в ряд Тейлора в окрестности текущей точки x^t:

$$f(x)-f(x^t) = \langle \mathrm{grad} f(x^t), x-x^t\rangle+\frac{1}{2}\langle f''(x^t)(x-x^t), x-x^t\rangle+o(\|x-x^t\|^2).$$

Выберем x^{t+1} из условия минимизации квадратичной аппроксимации $f(x)$ в точке x^t, т.е. квадратичной части приращения $f(x)-f(x^t)$, получим метод Ньютона:

$$x^{t+1} = x^t - (f''(x^t))^{-1}\mathrm{grad} f(x^t), \quad t = 1, 2, \ldots,$$

где начальное приближение x^1 должно находиться достаточно близко к точке оптимума x^*. В таком случае (и при дополнительных предположениях, более сильных, чем для приведенной ранее оценки скорости сходимости градиентного метода) для метода Ньютона будет справедлива *квадратичная скорость сходимости*

$$\|x^{t+1} - x^*\| \leq Q\|x^t - x^*\|^2, \quad \text{т.е.} \quad \|x^{t+1} - x^*\| \leq \frac{1}{Q}(Q\|x^1 - x^*\|)^{2^t},$$

что предполагает $\|x^1 - x^*\| < 1/Q$ (оценку для Q см., например, в [5, с. 192]). Еще раз подчеркнем, что градиентный метод в отличие от ньютоновского сходится при любом начальном приближении. Из определения метода Ньютона также следует требование невырожденности матрицы вторых производных (гессиана) функции f.

Нетрудно видеть, что полученная формула метода Ньютона решения задач безусловной минимизации совпадает с формулой метода Ньютона решения системы уравнений $\mathrm{grad} f(x) = 0$, соответствующей необходимым условиям экстремума.

3. Для задач условной минимизации, например $\min\limits_{x\in[1,2]} x^2$,

предложенные методы нуждаются в модификации. В частности, для приведенного примера, когда множество X имеет достаточно простую структуру, указанные выше формулы совмещаются с процедурой проектирования на X на каждом шаге метода. Так приходим к методу *проекции градиента*

$$x^{t+1} = \mathrm{Pr}_X\{x^t - \alpha_t \mathrm{grad} f(x^t)\}, \quad t = 1, 2, \ldots, \quad \forall x^1 \in \mathbf{R^n}.$$

Для более сложных множеств X, допустим, задаваемых ограничениями неравенствами

$$X \doteq \{x \in \mathbf{R^n} |\ g_i(x) \leq 0 \ \ \forall i \in M\}, \tag{3}$$

универсальным способом освобождения от ограничений является их штрафование. А именно для достаточно большой константы $C > 0$ вместо задачи условной минимизации (1),(3) рассматривают задачу безусловной минимизации оштрафованной целевой функции

$$\min_{x \in \mathbf{R^n}} \{f(x) + C \sum_{i \in M} [g_i^+(x)]^p\}, \quad \text{где} \quad \sum_{i \in M} [g_i^+(x)]^p \ -$$

это *функция штрафа (штрафная функция)* для ограничений неравенств, $g^+(\cdot) \doteq \max[0, g(\cdot)]$ — *срезка* g, параметр штрафа $p \geq 1$. (Другие виды функций штрафа см. в [4,5].) В условиях непрерывности функций $f,\ g_i$, непустоты X и ограниченности множества Лебега функции f можно доказать, что с ростом константы штрафа

$$\lim_{C \uparrow \infty} \min_{x \in \mathbf{R^n}} \{f(x) + C \sum_{i \in M} [g_i^+(x)]^p\} = f^*. \tag{4}$$

Если $p = 1$ (функция-срезка и, следовательно, штрафная функция является острой), то $\exists C^* :\ \ \min\{f(x) + C^* \sum_{i \in M} g_i^+(x)\} = f^*$ (существует *точный штраф*). Однако при $p > 1$ — *гладкий штраф* подобное равенство означало бы несущественность ограничений $x \in X$ (точка безусловного минимума и так находится в X).

Утверждение 4. Пусть $f, g_i \in \mathbf{C^1(R^n)}$, выпуклы, $p > 1$ и $\exists C^*$: $x^C \doteq \arg\min\{f(x) + C^* \sum [g_i^+(x)]^p\} \in X$, тогда

$$x^C \in \mathrm{Arg} \min_{x \in \mathbf{R^n}} f(x), \ \text{т.е.} \ \ \min_{x \in \mathbf{R^n}} f(x) = \min_{x \in X} f(x).$$

Доказательство. Так как x^C — точка безусловного экстремума дифференцируемой функции, то градиент оштрафованной функции цели в ней равен нулю: $\mathrm{grad} f(x^C) + C^* p \sum [g_i^+(x^C)]^{p-1} \mathrm{grad} g_i(x^C) = 0$. Но из условия $x^C \in X$ все выражения в квадратных скобках, а значит, и второе слагаемое равны нулю. Отсюда следует $\mathrm{grad} f(x^C) = 0$, т.е. необходимое условие экстремальности точки x^C для задачи безусловной оптимизации, которое в выпуклом случае оказывается и достаточным (см. утверждение 2). Поэтому x^C — точка безусловного минимума f. Но $x^C \in X$, так что x^C — и точка условного минимума f на X, ибо безусловный минимум не превышает условного. Утверждение доказано.

Таким образом, для гладкого штрафа не удается свести задачу условной минимизации к безусловной, тем не менее формула (4) позволяет итеративно комбинировать метод штрафов и градиентный метод в следующей процедуре: $\forall x^1 \in \mathbf{R^n}$

$$x^{t+1} = x^t - \alpha_t\{\mathrm{grad} f(x^t) + C_t p \sum_{i \in M} [g_i^+(x^t)]^{p-1} \mathrm{grad} g_i(x^t)\}, \ \ t = 1, 2, \dots,$$

которая сходится при определенных соотношениях между $\{\alpha_t\}$ и $\{C_t\}$, в частности для убывающего шага при $\sum \alpha_t^2 C_t^2 < \infty$ (например, $\alpha_t = 1/t, \ \ C_t < \sqrt{t}$).

Утверждение 4 показывает, что траектории метода штрафа проходят, вообще говоря, вне множества ограничений X, хотя и сходятся к данному множеству. Из-за этого рассмотренный метод иногда также называют методом внешних штрафов в отличие от методов *внутренней точки*, или *барьеров*. Типичным примером применения метода барьеров является описанный в §7 метод Кармаркара, когда задача (9), эквивалентная задаче условной минимизации

$$\min_{x \geq \bar{0}, \ \sum x_j = N} p(x),$$

сводится к безусловной минимизации специальной барьерной функции $k(x)$, не позволяющей методу Ньютона выйти за ограничения $x > 0$, если в этих ограничениях выбрано начальное приближение. Различные виды барьерных функций см. в [4,5] — для них характерно быстрое возрастание при приближении изнутри к границе множества ограничений (тогда как штрафная функция стремится к нулю при приближении к множеству ограничений — извне). Для решения общей задачи МП (1),(3) с ограничениями неравенствами методу Кармаркара соответствует использование вместо рассмотренной выше штрафной функции, основанной на срезке, *логарифмической* барьерной функции, равной

$$-\frac{1}{C} \sum_{i \in M} \ln[-g_i(x)]$$

при $g_i(x) < 0 \; \forall i \in M$ и $+\infty$ в противном случае. Эта функция также прибавляется к целевой, и справедливо соотношение, аналогичное (4).

Другие способы сведения задач условной оптимизации к безусловной, основанные на методе *множителей Лагранжа*, будут вытекать из результатов следующего параграфа.

§9. Двойственность в МП

Необходимые условия локального минимума обобщенно дифференцируемых функций при ограничениях неравенствах. Теорема Куна-Таккера. Понятие о регулярности ограничений неравенств в задаче МП. Метод множителей Лагранжа.

1. В этом параграфе будем рассматривать задачу условной оптимизации (1) с $X \subset \mathbf{R^n}$, $\quad X \neq \emptyset$, по преимуществу, с ограничениями неравенствами (3). Как уже отмечалось, условие равенства нулю градиента для таких задач может не иметь никакого отношения к точкам условного экстремума. Поэтому выведем соответствующие необходимые условия для рассматриваемого случая. Вначале они будут даны в достаточно общей форме, допускающей применение для широкого класса задач МП (кусочно-гладких и при произвольным образом заданных ограничениях, а также не обязательно конечномерных). Затем проведем конкретизацию для ограничений (3). Для обычных задач МП (конечномерных, с непрерывно дифференцируемыми функциями) справедливы все дальнейшие построения и выводы при замене знака ∇ обычным градиентом. Таким образом, основой обобщения является следующее

Определение 4. Функция f называется *дифференцируемой по Адамару* в точке $x \in \mathbf{R^n}$, если существует вектор $\nabla f(x) \in \mathbf{R^n}$, такой что $\forall y \in \mathbf{R^n}$ выполнено:

$$\lim_{(\tau, y') \to (+0, y)} \frac{f(x + \tau y') - f(x)}{\tau} = \langle \nabla f(x), y \rangle.$$

Для бесконечномерных задач, когда f — функционал: $E \to \mathbf{R^1}$, где E некоторое функциональное пространство, требуется: $\nabla f(x) \in E'$ для пространства E', сопряженного к E, и $x, y \in E$. В гладком случае $\nabla f(x) = \mathrm{grad} f(x)$ и можно положить y' тождественно равным y.

В безусловной оптимизации существенную роль играли направления спуска (убывания целевой функции). В условной оптимизации, кроме убывания целевой функции, требуется отслеживать еще и невыход за ограничения. Поэтому вводится понятие *возможного* или *допустимого* направления в точке $x \in X$ для множества ограничений X как такого вектора y, для которого $\exists \tau^0 > 0: \quad x + \tau y \in X \ \ \forall \tau \in [0, \tau^0]$. Замыкание множества всех допустимых направлений в точке x для X дает следующее

Определение 5. *Контингентным конусом* к множеству X в точке x называется следующее множество $K(X, x)$ векторов $y \in \mathbf{R}^n$:

$$K(X, x) \doteq \{y \mid \exists \ \{(\tau_t, y^t)\}_{t=1}^{\infty} \colon (\tau_t, y^t) \to (+0, y), \quad x + \tau_t y^t \in X \ \ \forall t\}.$$

Очевидно, для $\hat{x} \notin X \quad K(X, \hat{x}) = \emptyset$, а для $x' \in \mathrm{int}X \quad K(X, x') = \mathbf{R^n}$. Для $x \in \partial X$ в случае гладкой границы конус $K(X, x)$ называется также *конусом касательных* и соответствует касательным направлениям для ограничений-равенств.

Теорема 1 (*общий вид необходимых условий локального минимума в задаче (1)*). Пусть функция f дифференцируема по Адамару, $X \subset \mathbf{R^n}, \quad X \neq \emptyset, \quad x^0$ — точка локального минимума f в задаче (1), тогда $\forall y \in K(X, x^0) \ \ \langle \nabla f(x^0), y \rangle \geq 0$.

Доказательство. Выберем $y \in K(X, x^0)$. Для соответствующих ему по определению 5 $\{\tau_t, y^t\}$ выполнено $x^0 + \tau_t y^t \in X$, и, начиная с достаточно большого $t, \quad x^0 + \tau_t y^t \in X \cap \mathbf{O}_\varepsilon(x^0)$ (ибо $\tau_t \to 0$), следовательно, по определению 1 $f(x^0 + \tau_t y^t) \geq f(x^0)$. В пределе получим

$$\lim_{(\tau, y') \to (+0, y)} \frac{f(x^0 + \tau y') - f(x^0)}{\tau} = \lim_{(\tau_t, y^t) \to (+0, y)} \frac{f(x^0 + \tau_t y^t) - f(x^0)}{\tau_t} \geq 0,$$

и требуемое соотношение вытекает из определения 4.

Содержательно данные условия означают, что среди допустимых направлений в точке локального минимума не должно быть направлений убывания целевой функции (см. утверждение 3 §8). Однако в таком общем виде этими условиями неудобно пользоваться.

Конкретизируем полученные условия для ограничений неравенств, когда X задается формулой (3). Введем $\forall x \in X$ множество индексов $J(x)=\{i\in M|\ g_i(x)=0\}$ — *активных ограничений* в точке x, т.е. таких неравенств из (3), которые в этой точке выполнены как равенства. И определим множество (конус)

$$G(x)\doteq\{y\in\mathbf{R^n}|\ \langle\nabla g_j(x),y\rangle\le 0\ \ \forall j\in J(x)\}.$$

Определение 6. Множество X для ограничений неравенств (3) называется *регулярным в точке* $x\in X$, если $G(x)\subseteq K(X,x)$.

Теорема 2 *(необходимые условия локального минимума с ограничениями неравенствами).* Пусть функции $f,\ g_i\ \forall i\in M$ дифференцируемы по Адамару, $X\neq\emptyset,\ \ x^0$ — точка локального минимума f в задаче (1),(3) и множество X регулярно в точке x^0. Тогда

$$\exists\lambda_j\ge 0:\quad \nabla\{f(x^0)+\sum_{j\in J(x^0)}\lambda_j g_j(x^0)\}=0. \qquad (5)$$

Доказательство. По теореме 1 и из определения регулярности X в x^0 следует, что $\langle\nabla f(x^0),y\rangle\ge 0$ для всех y, удовлетворяющих условию $\langle\nabla g_j(x^0),y\rangle\le 0\ \ \forall j\in J(x^0)$. Значит, по определению 3 §7, линейное неравенство $\langle\nabla f(x^0),y\rangle\ge 0$ является следствием системы линейных неравенств $\{\langle\nabla g_j(x^0),y\rangle\le 0\ \ \forall j\in J(x^0)\}$. Приведя это неравенство к стандартному виду $\langle-\nabla f(x^0),y\rangle\le 0$ и применив афинную лемму Фаркаша (§7), получим, что

$$\exists\lambda_j\ge 0:\quad -\nabla f(x^0)=\sum_{j\in J(x^0)}\lambda_j\nabla g_j(x^0).$$

Таким образом, для регулярных ограничений необходимым условием локального минимума в гладкой задаче (1),(3) является равенство нулю дифференциала функции в фигурных скобках в (5) для хоть каких-нибудь $\lambda_j\ge 0$. Чтобы не записывать в явном виде множество активных ограничений, вводят *функцию Лагранжа*

$$L(\lambda,x)\doteq f(x)+\sum_{j\in M}\lambda_j g_j(x)\doteq f(x)+\langle\lambda,\bar g(x^0)\rangle$$

(регулярной) задачи (1),(3), где вектор-функция $\bar g(\cdot)\doteq(g_j(\cdot)|\ j\in M)$. Из теоремы 2 следует, что равенство нулю дифференциала функции Лагранжа для $\lambda_j\ge 0$ также является необходимым условием локального минимума в регулярной задаче (1),(3), ибо *множители Лагранжа* λ_j, соответствующие неактивным ограничениям, можно взять равными нулю. Последнее условие записывается как

$$\langle \lambda, \bar{g}(x^0) \rangle = 0 \tag{6}$$

и называется *условием дополняющей нежесткости.* Итак, доказана

Теорема 3 *(принцип оптимальности Лагранжа).* В предположениях теоремы 2 для задачи (1),(3) существует неотрицательный вектор множителей Лагранжа $\lambda \geq \bar{0}$, такой, что для x^0 выполнены *условия оптимальности*: $\nabla_x L(x^0, \lambda) = \bar{0}$ и (6).

Для выпуклых задач (1),(3) данные необходимые условия являются в регулярном случае и достаточными, и может быть доказана

Теорема 4 (*Куна, Таккера*). Если в задаче (1),(3) функции $f, g_j \in \mathbf{C^1(R^n)}$ выпуклы и множество X регулярно (в любой точке), то x^* — точка оптимума в этой задаче тогда и только тогда, когда в ней выполнены условия оптимальности для $\lambda \geq \bar{0}$.

Доказательство. Необходимость следует из предыдущих теорем, покажем достаточность. Для данного λ в точке x^* выполнено условие экстремальности x^* для функции $L(\cdot, \lambda)$. С учетом неотрицательности λ эта функция выпукла по x, значит, x^* является точкой ее минимума (см. утверждение 2 §8). Отсюда и из условия дополняющей нежесткости получим, что $f(x^*) = f(x^*) + \langle \lambda, \bar{g}(x^*) \rangle = L(x^*, \lambda) \leq L(x, \lambda) \doteq f(x) + \langle \lambda, \bar{g}(x) \rangle \leq f(x)$ $\forall x \in X$ (ибо $g_j(x) \leq 0$ для x, удовлетворяющих ограничениям), что и требуется в определении (2).

Аналогичные теоремам 2–4 утверждения справедливы и для случая, когда X задается ограничениями-равенствами, и для смешанных систем ограничений равенств и неравенств: $g_j(x) \leq 0$, $g_i(x) = 0$. Только на соответствующие ограничениям-равенствам множители Лагранжа λ_i не надо накладывать условия неотрицательности, а на условие дополняющей нежесткости эти ограничения не влияют (в случае ограничений-равенств вообще опускаем (6) и приходим к классическому *правилу множителей Лагранжа*).

2. Теперь вспомним, что полученные условия являются значимыми лишь в предположении регулярности ограничений, для которого определение 6 не дает конструктивного способа проверки. В данном пункте будут рассмотрены некоторые достаточные условия регулярности ограничений неравенств (3) для гладких задач.

Кроме $G(x)$, определенного в п.**1**, введем также множество

$$G^0(x) \doteq \{y \in \mathbf{R^n} | \ \langle \nabla g_j(x), y \rangle < 0 \ \ \forall j \in J(x)\},$$

отличающееся заменой нестрогого неравенства строгим. Но это множество уже включается в контингентный конус.

Утверждение 5. В предположении дифференцируемости по Адамару (или непрерывной дифференцируемости) функций g_j, задающих ограничения (3), $G^0(x) \subset K(X, x) \ \ \forall x \in X$.

Доказательство (от противного). Пусть существует направление $y \in G^0(x)$, не входящее в $K(X, x)$, т.е. для любой последовательности, фигурирующей в определении 5, найдется подпоследовательность $(\tau_t, y^t) \to (+0, y)$: $x + \tau_t y^t \notin X$, следовательно, $\forall t \ \exists$ индекс j, такой что $g_j(x + \tau_t y^t) > 0$. Возможных индексов — конечное число, а различных t бесконечно много, значит, найдется ограничение, пусть i-е, которое нарушается бесконечное число раз. Рассмотрим соответствующую подпоследовательность $\{t_k\}: \ \ g_i(x + \tau_{t_k} y_{t_k}) > 0$ и, устремляя $k \to \infty$, получим, что $g_i(x) \geq 0$. Но из условия $x \in X$ справедливо обратное неравенство, откуда следует равенство, т.е. $i \in J(x)$. Однако для этого i по определению 4 будем иметь $\langle \nabla g_i(x), y \rangle \doteq$

$$\doteq \lim_{(\tau, y') \to (+0, y)} \frac{g_i(x + \tau y') - g_i(x)}{\tau} = \lim_{k \to \infty} \frac{g_i(x + \tau_{t_k} y_{t_k}) - g_i(x)}{\tau_{t_k}} \geq 0.$$

Пришли к противоречию с $y \in G^0(x)$.

Отсюда получаем следующее *условие регулярности*:

$$G(x) = \overline{G^0(x)}. \tag{7}$$

Здесь и далее черта над множеством обозначает его замыкание.

Утверждение 6. В сделанных предположениях условие (7) обеспечивает регулярность X в точке x.

Для доказательства достаточно заметить, что множество $K(X, x)$ является замкнутым, а включение $G^0(x) \subset K(X, x)$ приводит к $\overline{G^0(x)} \subseteq K(X, x)$ после взятия операции замыкания.

Утверждение 7. Достаточным для (7) является

$$G^0(x) \neq \emptyset. \tag{8}$$

Доказательство. Из (8) для алгебраической суммы G и G^0 следует: $G + G^0 \subseteq G^0$, т.е. $\overline{G + G^0} \subseteq \overline{G^0}$, а $\overline{G^0} \supseteq \bar{0}$ дает $G + \overline{G^0} \supseteq G$. И из линейности оператора замыкания и замкнутости G получаем (7).

Для выпуклых X выполнение (8) и, следовательно, регулярность (в любой точке) ограничений (3) гарантируется *условием Слэйтера* ($\exists x' \in X : \quad g_i(x') < 0 \quad \forall i \in M$). Линейные ограничения всегда регулярны (множество G совпадает с контингентным конусом), хотя условие Слэйтера или (8) для них может не выполняться.

Другие типы условий регулярности, а также условия регулярности для смешанных систем ограничений равенств и неравенств см. в [4–6]. В частности, классическим условием регулярности для ограничений-равенств является линейная независимость градиентов ограничений в экстремальной точке.

УПРАЖНЕНИЕ 8. Получить теорему двойственности ЛП как следствие теоремы Куна-Таккера (для случая озЛП).

Условия оптимальности служат основным инструментом теоретического исследования задач условной оптимизации. Чтобы численно (приближенно) найти условный экстремум с их помощью, применяют методы безусловной оптимизации для поиска седловой точки функции Лагранжа или комбинируют штрафную функцию с функцией Лагранжа для получения точного гладкого штрафа. К сожалению, все эти методы останавливаются в первом попавшемся локальном экстремуме. Глобальный оптимум можно искать, перебирая локальные оптимумы, но для задач неодномерной минимизации не понятно, как находить все локальные оптимумы. Некоторые из существующих подходов к решению задач глобальной оптимизации приводятся в следующем параграфе.

Содержание