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

2. ОСНОВЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Литература:

3. Хачиян Л. Г. Сложность задач линейного программирования. М.: Знание, 1987, N 10.

§5. Понятие о сложности задачи линейного программирования (ЛП)

Определение основной задачи ЛП (озЛП). Принцип граничных решений и геометрическое описание симплекс-метода. Алгебраическая и битовая сложность методов ЛП. Результаты по сложности задач, близких к ЛП. Теорема о границах решений задач ЛП с целыми коэффициентами. Теорема о мере несовместности систем линейных неравенств с целыми коэффициентами.

1. Согласно [3] линейное программирование — это раздел прикладной математики, изучающий теорию, приложения и методы решения конечных систем линейных неравенств с конечным числом вещественных неизвестных $x_1, \dots, x_n$:

$$\left.\begin{array}{rl} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & \le b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & \le b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n & \le b_m \end{array}\right\} \tag{1}$$

или в сокращенной записи $Ax \le b$. Считаем, что матрица A не содержит нулевых строк a_i. *Основная задача ЛП (озЛП)* состоит в нахождении такого решения (1), которое максимизирует заданную линейную функцию $\langle c, x\rangle \doteq c_1x_1 + c_2x_2 + \ldots + c_nx_n$ вектора неизвестных x по всем вещественным x, удовлетворяющим системе (1):

$$\max_{x \in \mathbf{R}^n:\ Ax \le b} \langle c, x\rangle; \tag{2}$$

озЛП (2) с n неизвестными и m ограничениями называется задачей размерности (n, m) и задается числовой таблицей

$$\left|\begin{array}{ccccc} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \\ c_1 & c_2 & \ldots & c_n & 0 \end{array}\right| \tag{3}$$

своих коэффициентов. В частном случае $c = \bar{0}$ задача (2) эквивалентна (1), так что умение решать озЛП предполагает умение решать системы линейных неравенств (ЛН). В §7 будет показано обратное сведение. Вообще говоря, в форме (2) может быть представлена любая задача ЛП с ограничениями равенствами и неравенствами, в том числе *каноническая задача* ЛП

$$\max_{Ax=b,\ x\geq\bar{0}} \langle c, x\rangle.$$

(Здесь и далее черта сверху будет использоваться для выделения вектора в отличие от похожего числа.)

УПРАЖНЕНИЕ 5. Представить каноническую задачу ЛП в форме озЛП.

Несмотря на то, что формально задачи ЛП не являются дискретными ($x \in \mathbf{R^n}$), их решение нетрудно свести к перебору конечного числа угловых точек (вершин полиэдра (1), задающего ограничения) на основании *принципа граничных решений*:

если задача (2) имеет решение, то найдется такая подматрица A_I матрицы A, что любое решение системы уравнений $A_I x = b_I$, т.е.

$$\{a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i \mid i \in I\},$$

реализует максимум в (2).

Отметим, что для невырожденных A_I решение соответствующей системы уравнений $A_I x = b_I$, удовлетворяющее ограничениям (1), является угловой точкой (1). Из принципа граничных решений следует, что если угловая точка (1) существует, то разрешимая задача (2) имеет решение и в угловой точке (1), т.е. она эквивалентна максимизации $\langle c, x\rangle$ на конечном множестве вершин полиэдра (1). Процедура решения системы линейных уравнений методом Гаусса требует не более полинома 3-й степени от m, n (точнее, $\max(m,n)[\min(m,n)]^2$) арифметических операций с элементами A и b. Однако число возможных подматриц матрицы A экспоненциально, и метод полного их перебора не эффективен.

В 1820-х гг. Ж. Фурье и затем в 1947 г. Дж. Данциг предложили метод направленного перебора смежных вершин (1) — в направлении возрастания целевой функции (2) — *симплекс-метод*. Хотя каждый шаг симплекс-метода (представляющий собой определенную процедуру пересчета элементов симплекс-таблицы (3)) ограничен по порядку числом mn арифметических операций, в настоящее время для всех известных вариантов симплекс-метода приведены примеры, экспоненциальные по числу итераций, когда перебирается более $2^{\min(n,m/2)}$ вершин, но доказательство невозможности построить полиномиальный симплекс-метод также отсутствует. Подчеркнем, что на практике симплекс-метод не показывает данной оценки ("плохие" примеры довольно редки). Можно построить алгоритм решения задачи ЛП с оценкой $f(n)m$ арифметических операций (над числами, записанными в (3)), где $f(\cdot)$ растет быстрее экспоненты. Алгоритм с полиномиальной оценкой одновременно по n и m не известен и вряд ли будет построен.

Теперь заметим, что функция, оценивающая число арифметических операций в зависимости от n и m, не учитывает длину кода элементов (3), а только их количество и поэтому не является временно́й сложностью алгоритма. Указанная функция носит название *алгебраической сложности* в отличие от *битовой сложности* — функции, оценивающей число арифметических операций с битами (или с конечными порциями — по размеру машинного регистра) цифровой записи параметров индивидуальной задачи ЛП в зависимости от длины входного слова, т.е. от n, m и длин l кодов чисел в симплекс-таблице. Очевидно, битовая сложность алгоритма соответствует его временно́й сложности (см. §1). Входные коэффициенты задачи ЛП обычно рациональны, поэтому далее условимся считать их целыми, тогда l — длина записи максимального коэффициента в (3) — конечна. Набор (n, m, l) называется битовой размерностью задачи ЛП. Вопрос о существовании алгоритма ЛП с полиномиальной битовой сложностью был решен Л. Г. Хачияном в 1978 г., и тем самым была доказана полиномиальность задач ЛП. Основные моменты этого доказательства излагаются в следующем пункте и §6. Здесь же укажем на отличие классов сложности задачи ЛП и других линейных задач.

Метод Гаусса решения системы линейных алгебраических уравнений имеет полиномиальную алгебраическую сложность, т.е. является *сильнополиномиальным.* Для ЛП вопрос о существовании сильнополиномиального алгоритма открыт. Кроме того, задача решения системы линейных уравнений принадлежит классу **NC**, а аналогичный результат для ЛП означал бы равенство **NC**=**P**, ожидать которое нет оснований.

Из полиномиальности ЛП вытекает полиномиальность задачи **ЛН** (существует ли решение системы ЛН): **ЛН** $\in$ **P**. Аналогичные задачи с дополнительным ограничением целочисленности или булевости решения **NP**-полны: **ЦЛН**, **БЛН** $\in$ **NPC** (см. §2), т.е. полиномиальные алгоритмы для них вряд ли будут построены.

Существует неполиномиальное обобщение ЛП — задача проверки истинности высказываний вида

$$\mathbf{Q_1} x_1 \dots \mathbf{Q_n} x_n \ \mathbf{F}(\langle a_1, x\rangle \le b_1, \dots, \langle a_m, x\rangle \le b_m),$$

где $\mathbf{Q_i} \in \{\forall, \exists\}$, а $\mathbf{F}(\cdot, \dots, \cdot)$ — предложение, составленное из линейных неравенств с помощью связок $\&, \vee, \neg$ (и, или, отрицание). Доказано, что любой алгоритм, решающий эту массовую задачу, имеет не менее чем экспоненциальную сложность. Тот же результат будет и при замене равенствами всех неравенств в постановке задачи.

2. Рассмотрим некоторые свойства задач ЛП с целыми коэффициентами. Для любой целочисленной матрицы D введем параметр

$$\Delta(D) \doteq \max_{\{D' \text{ – квадратная подматрица } D\}} |\det D'|.$$

Будем обозначать через $[A|b]$ матрицу, составленную из A и вектора-столбца $b \in \mathbf{Z^m}$, дописанного справа. Здесь и далее $\mathbf{Z^m}$ — m-мерное пространство целочисленных векторов, $\mathbf{Z^m_+}$ — его неотрицательный ортант.

Теорема 1 *(о границах решений)*. Если озЛП (2) размерности (n, m) с целыми коэффициентами разрешима, то у нее существует рациональное решение x^* в шаре $\|x\| \le n^{1/2}\Delta([A|b])$ и значением озЛП (2) $d^* \doteq \langle c, x^*\rangle$ является рациональное число t/s со знаменателем, ограниченным величиной $\Delta(A)$.

Доказательство. На основании принципа граничных решений $\exists A_I \subseteq A$: по правилу Крамера $|x_j^*| = |\det A_I^j / \det A_I| \le \Delta([A|b])$, ибо $|\det A_I| \ge 1$, а определитель матрицы A_I^j, полученной из A_I заменой

j-го столбца на $\pm b_I$, не превышает по модулю $\Delta([A|b])$. Отсюда для евклидовой нормы x^* получаем требуемую оценку. С учетом целочисленности вектора c знаменатель d^* может быть выбран равным знаменателю x_j^* $\forall j$, и 2-е утверждение теоремы следует из определения $\Delta(A) \geq |\det A_I|$.

Определение 1. Точка x^ε называется ε-приближенным решением системы линейных неравенств (1), если

$\langle a_i, x^\varepsilon \rangle \leq b_i + \varepsilon \;\; \forall i = \overline{1,m}$, где a_i — i-я строка матрицы A,
или в матричной записи, обозначая $\mathbf{e}$ — вектор-столбец из единиц,

$$Ax^\varepsilon \leq b + \varepsilon \mathbf{e}. \qquad (1_\varepsilon)$$

Теорема 2 *(о мере несовместности)*. Если система ЛН (1) имеет ε_1-приближенное решение для $\varepsilon_1 \doteq 1/[(n+2)\Delta(A)]$, то эта система разрешима, т.е. имеет точное решение x^0.

Доказательство. Обозначим через ε^* минимальное ε, при котором система (1_ε) имеет решение (по условию $\varepsilon^* \leq \varepsilon_1$):

$$\varepsilon^* \doteq \min_{(x,\varepsilon):\; Ax \leq b + \varepsilon \mathbf{e}} \varepsilon.$$

Допустим, что утверждение теоремы не верно, тогда $\varepsilon^* > 0$. Задача определения ε^* является (с учетом равенства $\min(\cdot) = -\max(-\cdot)$) озЛП с целевым вектором $c = (0, \ldots, 0, -1)$, $n+1$ переменными (x, ε) и ограничениями $Ax - \varepsilon \mathbf{e} \leq b$. Следовательно, по теореме 1 ε^* может быть представлена в виде дроби со знаменателем, не превышающим $\Delta([A| - \mathbf{e}]) \leq (n+1)\Delta(A)$, т.е. $\varepsilon^* \geq 1/[(n+1)\Delta(A)] > \varepsilon_1$ — пришли к противоречию с определением ε^*.

Аналогичное утверждение справедливо и для озЛП.

Определение 2. Точка x_ε^* называется ε-приближенным решением озЛП (2), если она является ε-приближенным решением системы (1) и реализует максимум в (2) с ε-точностью:

$$\langle a_i, x_\varepsilon^* \rangle \leq b_i + \varepsilon \;\; \forall i = \overline{1,m} \;\text{ и }\; \langle c, x_\varepsilon^* \rangle \geq d^* - \varepsilon.$$

Теорема 2 * *(о мере несовместности)*. Если озЛП (2) имеет ε_2-приближенное решение для $\varepsilon_2 \doteq 1/(2n^2\Delta^3(A))$, то эта задача имеет точное решение x^*.

Доказательство см. в [3, с. 21].

§6. Метод эллипсоидов

Полиномиальный алгоритм округления ε_1-приближенного решения системы линейных неравенств. Метод эллипсоидов ε_2-приближенного решения озЛП. Оценка сложности метода эллипсоидов. Полиномиальность ЛП.

1. Имея ε-приближенное решение (1) с $\varepsilon \le \varepsilon_1$, можно (на основании теоремы 2, §5) быть уверенным в существовании точного решения системы линейных неравенств. Оказыватся, процедура получения x^0 из x^{ε_1} является полиномиальной. Соответствующий *алгоритм округления* ε_1-приближенного решения системы (1) до точного был указан Л. Г. Хачияном и состоит в следующем.

Присвоим $x^1 := x^{\varepsilon_1}$ и подставим x^1 в (1). Разобьем множество $M \doteq \{1, \dots, m\}$ индексов неравенств в системе на два подмножества

$$M(x^1) \doteq \{i: \ \ |\langle a_i, x^1\rangle - b_i| \le \varepsilon_1\},$$
$$M \setminus M(x^1) \doteq \{i: \ \ \langle a_i, x^1\rangle - b_i \le -\varepsilon_1\}.$$

Найдем решение x'^1 системы равенств $A_{M(x^1)}x = b_{M(x^1)}$ (существует по теореме 2). Пусть x'^1 не является точным решением (1), т.е. в x'^1 не выполнилось i-е неравенство для какого-либо $i \notin M(x^1)$. Тогда введем множество индексов невыполненных неравенств $M^+ \doteq \{i|\ \langle a_i, x'^1\rangle > b_i\} \subseteq M \setminus M(x^1)$ и рассмотрим на отрезке $[x^1, x'^1]$ ближайшую к x'^1 точку, в которой еще выполнены все неравенства для $i \in M^+$ (в x^1 они выполнены с ε_1-запасом). А именно определим

$$\tau \doteq \min_{i \in M^+} \frac{b_i - \langle a_i, x^1\rangle}{\langle a_i, x'^1\rangle - \langle a_i, x^1\rangle}, \quad i_1 \doteq \arg\min_{i \in M^+} \frac{b_i - \langle a_i, x^1\rangle}{\langle a_i, x'^1\rangle - \langle a_i, x^1\rangle}$$

и присвоим $x^2 := (1-\tau)x^1 + \tau x'^1$. Имеем $M(x^2) \supseteq M(x^1) \cup \{i_1\}$, ибо неравенства с индексами из $M(x^1)$ ε_1-приближенно выполнялись как равенства на всем отрезке $[x^1, x'^1]$, а неравенство с индексом $i_1 \in M^+$, не выполненное в точке x'^1, выполняется в x^2 как равенство по построению. Таким образом, $M(x^2) \supset M(x^1)$, но $|M(x)| \le m$, поэтому, повторяя указанную процедуру с заменой x^1 на x^2 и т.д., придем не более чем через $\max(n, m)$ шагов к тому, что решение x' соответствующей системы равенств окажется x^0 — решением (1).

С учетом полиномиальности задачи решения систем уравнений предложенный алгоритм округления полиномиален.

Аналогичный алгоритм имеется и для округления ε_2-приближенного решения озЛП $x^*_{\varepsilon_2}$ до точного x^* (см. [3, с. 21]). Поэтому для построения полиномиального алгоритма решения озЛП осталось указать полиномиальный алгоритм поиска ε_2-приближенного решения озЛП в шаре $\|x\| \leq n^{1/2}\Delta$ или удостоверения, что такого решения нет (по теоремам 1,2* из §5). Требуемый алгоритм, основанный на *методе эллипсоидов*, который предложили в 1976–77 гг. Д. Б. Юдин и А. С. Немировский и (независимо) Н. З. Шор, приводится в следующих пунктах.

Здесь и далее $\Delta \doteq \Delta(D)$, где матрица D задается таблицей (3).

2. Пусть E — произвольный эллипсоид в $\mathbf{R^n}$ с центром ξ и ненулевого объема vol$E$. Рассмотрим $(n-1)$-мерную плоскость, заданную вектором g нормали и проходящую через центр ξ эллипсоида E. Обозначим через $E^-(g)$ один из двух полуэллипсоидов, на которые разбивает E данная плоскость, $E^-(g) = E \cap \{x| \ \langle g, x-\xi\rangle \leq 0\}$.

Утверждение 1. Полуэллипсоид $E^-(g)$ эллипсоида E можно целиком заключить в новый эллипсоид E', имеющий объем, строго меньший E,

$$\frac{\mathrm{vol}E'}{\mathrm{vol}E} < e^{-1/(2n+2)}, \qquad (*)$$

и E' можно вычислить по $E^-(g)$ за $O(n^2)$ арифметических операций.

Доказательство. Пусть E — единичный шар с центром в точке $\bar{0}$: $E = \{x \in \mathbf{R^n} : \ \|x\| \leq 1\}$, а $E^-(g) = E \cap \{x_n \geq 0\}$. Поместим центр E' в точку $\xi' = (0,\ldots,0,\frac{1}{n+1})$, тогда

$$E' = \{x| \ (x_1^2 + \ldots + x_{n-1}^2)/\beta^2 + (x_n - \frac{1}{n+1})^2/\alpha^2 \leq 1\},$$

где $\alpha \doteq 1 - 1/(n+1) < e^{-1/(n+1)}$, $\beta^2 \doteq 1 + 1/(n^2-1) < e^{1/(n^2-1)}$. Отношение объемов равно произведению полуосей $\alpha\beta^{n-1} < e^{-1/(2n+2)}$, отсюда получаем $(*)$, ибо любой эллипсоид можно превратить в шар афинным преобразованием координат, сохраняющим объем. Действительно, будем представлять произвольный эллипсоид E с помощью его центра ξ и матрицы Q $(n \times n)$, задающей указанное преобразование: $E = \{x| \ x = \xi + Qy, \ \|y\| \leq 1\}$. Обозначим $\eta \doteq Q^T g/\|Q^T g\|$, где верхний индекс ^{T} — знак транспонирования. Тогда ξ' и Q', представляющие эллипсоид E' минимального объема, описанный вокруг полуэллипсоида $E^-(g)$, пересчитываются по формулам

$$\xi' = \xi - \frac{1}{n+1} Q\eta, \quad Q' = \frac{n}{\sqrt{(n^2-1)}} \{Q + (\sqrt{\frac{n-1}{n+1}} - 1) Q\eta\eta^T\}$$

за $O(n^2)$ арифметических операций.

3. *Метод эллипсоидов получения ε-приближенного решения озЛП.* Положим $\varepsilon := \varepsilon_2 \doteq 1/(2n^2\Delta^3)$. Введем множество ε-приближенных решений озЛП в шаре радиуса $R \doteq n^{1/2}\Delta$ с центром в $\bar{0}$: $X^*_\varepsilon \doteq \{x| \langle a_i, x\rangle \le b_i + \varepsilon \ \ \forall i = \overline{1,m}, \ \ \langle c, x\rangle \ge d^* - \varepsilon, \ \ \|x\| \le R\}$. Выберем указанный выше шар в качестве начальной итерации для эллипсоида $E \supset X^*_\varepsilon$. Рассмотрим произвольную итерацию.

Проверяем, является ли центр ξ эллипсоида E ε-приближенным решением. Если да, то алгоритм заканчивает свою работу, в противном случае строим эллипсоид E' для очередной итерации как минимальный по объему эллипсоид, содержащий полуэллипсоид $E^-(g)$ (см. п.2), где вектор g определяется следующим образом. Так как $\xi \notin X^*_\varepsilon$, то либо

1^0) $\exists i: \ \ \langle a_i, \xi\rangle > b_i + \varepsilon$, и тогда $g := a_i$, либо

2^0) $\langle c, \xi\rangle < d^* - \varepsilon$ и $g := -c$.

Убедимся, что при этом $X^*_\varepsilon \subset E'$. Действительно, для варианта 1^0 $\forall x \in X^*_\varepsilon \ \ \langle a_i, x\rangle \le b_i + \varepsilon < \langle a_i, \xi\rangle$, т.е. $X^*_\varepsilon \subset E \cap \{x| \langle a_i, x - \xi\rangle \le 0\} = E^-(a_i) \subset E'$; и аналогично получим для варианта 2^0

$$X^*_\varepsilon \subset E \cap \{x| \langle c, x - \xi\rangle \ge 0\} = E^-(-c) \subset E'.$$

Теперь с $E := E'$ возвращаемся к началу итерации (на новый шаг).

Оценим число итераций метода эллипсоидов. Покажем, что X^*_ε содержит шар радиуса $r/2$, где $r \doteq \varepsilon/(hn^{1/2}) < R$, $\ h \ge |a_{ij}|, |c_j|$ (h *высота задачи*). Пусть x^* — точное решение в X^*_ε. Из $\|x^* - x\| \le r$ следует $|\langle a_i, x\rangle - \langle a_i, x^*\rangle| \le \|a_i\| \ \|x^* - x\| \le hn^{1/2}r = \varepsilon \ \ \forall i \in M$ и $|\langle c, x\rangle - \langle c, x^*\rangle| \le \|c\| \ \|x^* - x\| \le hn^{1/2}r$, т.е. указанный выбор r гарантирует, что все такие x будут ε-приближенными решениями. Поскольку $\|x^*\| \le R$, то множество тех из рассматриваемых x, для которых $\|x\| \le R$ (т.е. пересечение шаров радиуса r и R, включающее центр первого), содержит шар радиуса $r/2$. Этот шар и принадлежит X^*_ε. Таким образом, объем X^*_ε больше объема n-мерного шара радиуса $r/2$. Значит, объем эллипсоида, построенного последним, например E^k для k-й итерации, не должен оказаться меньше объема этого шара. Отсюда и из утверждения 1 получаем для k соотношение

$$\left(\frac{r}{2R}\right)^n \leq \frac{\mathrm{vol} X_\varepsilon^*}{\mathrm{vol} E^1} \leq \frac{\mathrm{vol} E^k}{\mathrm{vol} E^1} < e^{-k/(2n+2)},$$

из которого k (по определению r, R, ε, h и Δ) не превосходит

$$2n^2 \ln(Rnh/\varepsilon) < 2n^2 \ln(2n^{3.5}\Delta^5) < 10n^2 \ln(n\Delta).$$

УПРАЖНЕНИЕ 6. Оценить по порядку битовую длину L входа озЛП: доказать, что $L > O(\ln(n\Delta))$.

Следовательно, число итераций метода эллипсоидов $k < O(n^2)L$, и с учетом $O(n^2 + nm)$ арифметических операций для каждой итерации получим оценку $O(n^3(n+m)L)$ для числа арифметических операций, достаточного методу эллипсоидов при поиске ε_2-приближенного решения озЛП. Алгоритм округления ε_2-приближенного решения до точного этой оценки не портит (см.[3, с. 21]). Можно также показать, что при реализации метода эллипсоидов и алгоритма округления все арифметические операции достаточно проводить с числами двоичной длины, ограниченной $O(L)$. При этом ошибки, возникающие за счет конечности числа разрядов (ошибки округлений), поглощаются путем некоторого дополнительного увеличения (“раздутия”) описанного эллипсоида E' на каждой итерации [3, с. 24], что не влияет на порядок оценки для общего числа итераций. В результате временна́я сложность такой процедуры решения озЛП оказывается полиномом от длины входа и справедлива

Теорема 3. Задача ЛП с целыми коэффициентами разрешима за полиномиальное от длины входа время.

Следствием данной теоремы является

Утверждение 2. **ЛН $\in$ P**.

Подчеркнем, что несмотря на доказанную полиномиальность, метод эллипсоидов не может конкурировать с симплекс-методом при практическом решении задач ЛП (реально он применяется в выпуклом квадратичном программировании), поскольку полученная оценка числа его итераций достигается на любых индивидуальных задачах, даже если в качестве начального приближения взять решение. Тогда как симплекс-метод для "хороших" (невырожденных) задач дает оценку $O(n^3)$, на порядок меньшую, чем метод эллипсоидов, и за одну итерацию может подтвердить, что начальное приближение является решением. Тем не менее сам факт полиномиальности ЛП инициировал поиск новых методов ЛП, что привело к созданию целого класса эффективных методов математического программирования — *методы внутренней точки* — и позволило построить конкурентоспособные полиномиальные алгоритмы ЛП. Идея их построения будет изложена в следующем параграфе, где также приводятся необходимые сведения по теории ЛП, начиная с ЛН.

§7. Теория двойственности ЛП. Идея метода Кармаркара

Следствия систем ЛН. Афинная лемма Фáркаша /без доказательства/. Лемма Фаркаша о неразрешимости. Теорема двойственности ЛП. Сведение озЛП к однородной системе уравнений с ограничением положительности. Идея метода Кармаркара и его отличие от симплекс-метода.

1. Система ЛН (1) называется *разрешимой*, если $\exists x$: $Ax \leq b$, и *неразрешимой* — в противном случае. ОзЛП (2) разрешима, когда разрешима система (1) и максимум в (2) достигается.

Определение 3. Линейное неравенство

$$\langle c, x\rangle \leq d \tag{4}$$

является *следствием* разрешимой системы линейных неравенств (1), если для любого x, удовлетворяющего (1), выполнено (4).

Способ получения неравенств-следствий довольно прост: выберем произвольные $\lambda_i \geq 0 \;\; \forall i \in M$, домножим на λ_i каждое i-е неравенство системы (1) и сложим; получим для вектора

$$c = \sum_{i\in M} \lambda_i a_i \;\; \text{и любого числа} \;\; d \geq \sum_{i\in M} \lambda_i b_i,$$

что (4) будет следствием (1). Оказывается, других следствий у ЛН не бывает. А именно справедлива

Лемма *Фа́ркаша (афинная).* Линейное неравенство (4) является следствием разрешимой в вещественных переменных системы ЛН (1) тогда и **только тогда**, когда существует вектор $\lambda \in \mathbf{R^m}$:

$$c = \sum_{i\in M} \lambda_i a_i, \quad d \geq \sum_{i\in M} \lambda_i b_i, \quad \lambda_i \geq 0 \;\; \forall i \in M. \qquad (5)$$

(Схему доказательства см. в [3, с. 18].)

Для неразрешимой системы ЛН (1) можно формально считать следствием (1) заведомо неверное неравенство $\langle \bar{0}, x\rangle \leq -1$ и далее пользоваться афинной леммой Фаркаша, как показывает

Лемма Фаркаша *о неразрешимости.* Система ЛН (1) неразрешима тогда и только тогда, когда разрешима система

$$\sum_{i\in M} \lambda_i a_i = \bar{0}, \quad \sum_{i\in M} \lambda_i b_i \leq -1, \quad \lambda_i \geq 0 \;\; \forall i \in M. \qquad (6)$$

Доказательство. Пусть (1) неразрешима, тогда из разрешимости системы $\langle a_i, x\rangle + x_{n+1} \leq b_i \;\; \forall i \in M$ должно следовать, что $x_{n+1} \leq -\varepsilon < 0$, т.е. следствием этой системы является неравенство $\langle (0,\ldots,0,1/\varepsilon),(x,x_{n+1})\rangle \leq -1$ и из афинной леммы Фаркаша получаем (6) (а также в дополнение $\sum \lambda_i = 1/\varepsilon$). Если же (6) разрешима и (1) разрешима, то указанное выше неравенство $\langle \bar{0}, x\rangle \leq -1$ оказывается следствием (1) и должно выполняться для всех x, удовлетворяющих (1), значит, таких не существует.

Теперь мы можем доказать основной теоретический результат ЛП — теорему двойственности, на которой базируются как методы решения задач ЛП, так и способы анализа решения, и которая фактически дает необходимые и достаточные условия оптимальности в ЛП. Наличие двойственности, обусловив хорошую характеризацию задачи **ЛН**, предопределило полиномиальность ЛП.

Определение 4. *Двойственной* к задаче ЛП на максимум с ограничениями неравенствами в форме озЛП (2) называется следующая задача ЛП на минимум с ограничениями в канонической форме:

$$\min \{ \sum_{i\in M} \lambda_i b_i | \sum_{i\in M} \lambda_i a_i = c, \; \lambda_i \geq 0 \; \forall i \in M \}, \quad \text{или в краткой записи}$$

$$\min_{\lambda A=c,\ \lambda\geq\bar{0}}\langle\lambda, b\rangle. \tag{7}$$

Для того, чтобы построить двойственную к произвольной задаче ЛП, надо представить ее в форме озЛП, применить формулу (7), а затем вернуться к обозначениям исходной задачи.

УПРАЖНЕНИЕ 7. Показать, что двойственная задача к двойственной задаче ЛП совпадает с прямой задачей ЛП: представить (7) в форме озЛП (аналогично упражнению 5), выписать двойственную к полученной задаче и свести ее к (2).

ТЕОРЕМА 4 (*двойственности ЛП*). Задача ЛП разрешима тогда и только тогда, когда разрешима двойственная к ней. В случае разрешимости оптимальные значения целевых функций в обеих задачах совпадают, т.е. $d^*=d^{**}$, где d^* — значение (2), d^{**} — значение (7).

ДОКАЗАТЕЛЬСТВО проведем для случая озЛП, поскольку любая задача ЛП адекватно представляется в такой форме.

Пусть задача (2) разрешима, тогда (4) является следствием (1) $\forall d\geq d^*$ и не является $\forall d<d^*$, что по афинной лемме Фаркаша эквивалентно разрешимости (5) при $d\geq d^*$ и неразрешимости (5) при $d<d^*$, т.е. $d^*=\min\{d|\,(5)\}$, а это и есть значение (7).

И наоборот, из разрешимости (7) следует неразрешимость (6), ибо в противном случае min в (7) обращался бы в $-\infty$ (так как прибавление решения (6) к решению (7) дает допустимую точку и уменьшает значение целевой функции (7)). Отсюда получаем разрешимость (1) по лемме Фаркаша о неразрешимости. Кроме того, разрешимость (7) означает разрешимость (5) для любого $d\geq d^{**}$, так что (4) оказывается следствием (1) для $d\geq d^{**}$, и поэтому d^{**} ограничивает сверху значение (2), т.е. максимум в (2) достигается. Таким образом получили разрешимость задачи (2) и можем вернуться к началу доказательства для установления равенства $d^*=d^{**}$.

Из теоремы 4 непосредственно получаем

УТВЕРЖДЕНИЕ 3. Задача ЛП оптимизации эквивалентна решению системы линейных неравенств.

Действительно, озЛП (2) эквивалентна задаче ЛП (7) и обе они эквивалентны системе ЛН относительно неизвестных (x,λ):

$$Ax\leq b,\ \langle c,x\rangle=\langle b,\lambda\rangle,\ \lambda A=c,\ \lambda\geq\bar{0}. \tag{8}$$

Утверждение 4. Задача ЛП оптимизации эквивалентна решению системы линейных уравнений в неотрицательных переменных.

Доказательство. От системы ЛН (8) переходим к ограничениям в канонической форме аналогично упражнениям 5,7.

Утверждение 5. Задача ЛП эквивалентна поиску неотрицательного ненулевого решения однородной системы линейных уравнений.

Доказательство. На основании утверждения 4 озЛП сводится к некоторой системе ЛН (с целыми коэффициентами) относительно вектора вещественных неизвестных y:

$$\hat{P}y = \hat{q}, \quad y \geq \bar{0}, \tag{9}$$

пусть $\hat{P}$ — матрица $(K \times (N-1))$. Введем параметр Δ, мажорирующий координаты некоторого решения (9) (по теореме о границах решений), если система (9) разрешима. Добавим к (9) неравенство

$$\langle y, \mathbf{e} \rangle = y_1 + \ldots + y_{N-1} \leq N\Delta,$$

которое превратим в равенство с помощью дополнительной переменной y_N: $\langle \hat{y}, \mathbf{e} \rangle = y_1 + \ldots + y_{N-1} + y_N = N\Delta$, а (9) перепишется как $[\hat{P}|\bar{0}]\hat{y} = \hat{q}, \quad \hat{y} \geq \bar{0}$. Теперь сделаем замену переменных $x := \hat{y}/\Delta$ и обозначим $P \doteq N\Delta[\hat{P}|\bar{0}] - [\hat{q}|\hat{q}| \ldots |\hat{q}]$. Придем к однородной системе $Px = \bar{0}$ с дополнительными ограничениями $x = (x_1, \ldots, x_N) \geq \bar{0}$, $\langle x, \mathbf{e} \rangle = N$, что соответствует системе $Px = \bar{0}, \quad x \geq \bar{0}, \quad \langle x, \mathbf{e} \rangle > 0$ с решениями-лучами $tx^0 \quad \forall t > 0$, любое из которых пересчитывается в решение исходной системы.

2. *Метод Кармаркара* (N. Karmarkar, 1984 г.). Воспользуемся утверждением 5 и обозначениями, введенными при его доказательстве. Пусть $p(x) \doteq (\langle p_1, x \rangle)^2 + \ldots + (\langle p_K, x \rangle)^2$, где p_i — строки P. Тогда $p(x) = 0$ эквивалентно $Px = \bar{0}$. Введем *функцию Кармаркара*

$$k(x) \doteq \frac{[p(x)]^{N/2}}{x_1 x_2 \cdot \ldots \cdot x_N}.$$

Применяя теорему 2 и алгоритм округления к задаче (9), можно показать, что для точного ее решения достаточно найти такой $\hat{x} > \bar{0}$, для которого $k(\hat{x}) \leq 1/[3(\Delta(\hat{P}))^N]$ [3, с. 25–26].

Полиномиальный алгоритм поиска нужного приближенного $\hat{x}$ приводится в [3, с. 26–28], и мы не будем его описывать. Отметим только, что аналогичный алгоритм может быть построен на основании применения метода Ньютона (см. в разд.3) к задаче минимизации функции Кармаркара или ей подобных. В результате получаем целый класс полиномиальных алгоритмов ЛП, которые на практике оказываются сравнимыми с симплекс-методом, не имея теоретических недостатков последнего. Предложенные алгоритмы строятся на принципиально новой идее: не дискретной, а непрерывной трактовки задачи ЛП, когда вместо перебора конечного числа угловых точек осуществляют поиск решения в исходном пространстве вещественных переменных, и траектории алгоритмов не проходят через угловые точки. Напомним, что метод эллипсоидов также не ориентируется на угловые точки многогранника ограничений. Характерно, что именно такой уход от дискретного программирования позволил построить полиномиальные алгоритмы ЛП. Поэтому далее будет дан некоторый обзор основных подходов к решению непрерывных задач оптимизации.

Замечание. Если бы речь шла о непосредственном поиске точного решения задачи ЛП указанными методами, то нельзя было бы гарантировать конечношаговость (не то, что полиномиальность) соответствующих алгоритмов. Для их применения существенной является возможность остановки в приближенном решении благодаря наличию полиномиального алгоритма округления. Но поскольку для его работы требуется начальное приближение из определенной окрестности решения, зависящей от длины l или высоты h, или длины входа L конкретной задачи ЛП, то и число итераций алгоритмов, базирующихся на рассматриваемом принципе, зависит от числа цифр в записи элементов матрицы ограничений. Так что не удается использовать данную идею для построения сильнополиномиальных алгоритмов ЛП, кроме как в частных случаях ограниченности элементов матрицы (например, в задачах на графах и сетях, где $a_{ij} = 0, \pm 1$).

Содержание

1. ВВЕДЕНИЕ В ТЕОРИЮ СЛОЖНОСТИ
§1. Понятие о сложности решения задач 3
§2. NP-полные (универсальные) задачи 10
§3. Классы сложности. Сильная NP-полнота
и псевдополиномиальность 15
§4. Приближенное решение задач комбинаторной
оптимизации 21
2. ОСНОВЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ
§5. Понятие о сложности задачи
линейного программирования (ЛП) 24
§6. Метод эллипсоидов 29
§7. Теория двойственности ЛП. Идея метода Кармаркара 33
3. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОГО
ПРОГРАММИРОВАНИЯ
§8. Обзор идей математического программирования (МП) 38
§9. Двойственность в МП 45
4. СПОСОБЫ РЕШЕНИЯ ПЕРЕБОРНЫХ ЗАДАЧ
§10. Глобальная оптимизация.
Метод ветвей и границ (МВГ) 51
§11. Целочисленное линейное программирование (ЦЛП) 54
§12. Метод динамического программирования (ДП) 58