МОДИФИЦИРОВАННЫЙ СИМПЛЕКС-МЕТОД*

В. Н. Малозёмов

malv@math.spbu.ru

20 ноября 2010 г.

Симплекс-метод решения задач линейного программирования является одним из выдающихся математических достижений 20-го столетия. В докладе излагается вариант симплекс-метода с обратной матрицей (модифицированный симплекс-метод) в том виде, в каком он читается мною в течение многих лет в курсе «Экстремальные задачи». Этим докладом я хотел бы обратить внимание участников семинара на замечательную, но уже забытую, книгу [1].

1°. Рассмотрим задачу линейного программирования в канонической форме

$$\begin{aligned} f(x) := c[N] \times x[N] &\to \inf, \\ A[M,N] \times x[N] &= b[M], \\ x[N] &\geqslant \mathbb{O}[N]. \end{aligned} \tag{1}$$

Вектор $x = x[N]$, удовлетворяющий ограничениям задачи (1), называется *планом*. Требуется найти план, доставляющий минимум целевой функции $f(x)$.

С планом x связан его *носитель*

$$N_+(x) = \bigl\{ j \in N \mid x[j] > 0 \bigr\}.$$

План x называется *базисным*, если столбцы $A[M,j]$ матрицы $A[M,N]$ при $j \in$ $\in N_+(x)$ линейно независимы. Базисный план x называется *невырожденным*, если $|N_+(x)| = |M|$.

Невырожденному базисному плану x соответствует квадратная *базисная матрица* $A[M, N_+(x)]$ с линейно независимыми столбцами. Она обратима. Матрица

$$B[N_+(x), M] = \bigl(A[M, N_+(x)]\bigr)^{-1}$$

*Семинар по дискретному гармоническому анализу и геометрическому моделированию «DHA & CAGD»: `http://www.dha.spb.ru/`

называется *обратной базисной матрицей*. По определению

$$B[N_+(x), M] \times A[M, N_+(x)] = E[N_+(x), N_+(x)],$$
$$A[M, N_+(x) \times B[N_+(x), M] = E[M, M].$$

Задачу (1) будем решать с помощью симплекс-метода, который позволяет от базисного плана x_0 перейти к «соседнему» базисному плану x_1 с меньшим значением целевой функции, $f(x_1) < f(x_0)$.

Предполагается, что выполнено *условие невырожденности*: все базисные планы задачи (1) невырождены. Роль этого условия существенна для доказательства конечной сходимости симплекс-метода.

2°. Перейдём к описанию метода. Возьмём начальный базисный план x_0 с носителем $N_+(x_0) =: N_+^{(0)}$ (вопрос о построении начального базисного плана рассмотрим позже). Проверим план x_0 на оптимальность.

Запишем двойственную задачу

$$\begin{gathered} b[M] \times u[M] \to \sup, \\ u[M] \times A[M, N] \leqslant c[N]. \end{gathered} \tag{2}$$

Условия дополнительности для плана x_0 имеют вид

$$u[M] \times A[M, N_+^{(0)}] = c[N_+^{(0)}].$$

Отсюда находим двойственный вектор

$$u_0[M] = c[N_+^{(0)}] \times B_0[N_+^{(0)}, M]. \tag{3}$$

Проверим, является ли вектор u_0 планом двойственной задачи. Для этого вычислим *оценки*

$$\Delta_0[j] = u_0[M] \times A[M, j] - c[j], \qquad j \in N \setminus N_+^{(0)}.$$

Если при всех $j \in N \setminus N_+^{(0)}$ выполняется неравенство $\Delta_0[j] \leqslant 0$, то u_0 — план двойственной задачи. По второй теореме двойственности x_0 — решение задачи (1) (и u_0 — решение двойственной задачи (2)).

3°. Предположим, что $\Delta_0[j_0] > 0$ при некотором $j_0 \in N \setminus N_+^{(0)}$. С учётом (3) перепишем это условие в виде

$$\Delta_0[j_0] = c[N_+^{(0)}] \times B_0[N_+^{(0)}, M] \times A[M, j_0] - c[j_0] > 0. \tag{4}$$

Обозначим

$$z_0[N_+^{(0)}] = B_0[N_+^{(0)}, M] \times A[M, j_0]. \tag{5}$$

Получим

$$\Delta_0[j_0] = c[N_+^{(0)}] \times z[N_+^{(0)}] - c[j_0] > 0. \tag{6}$$

Отметим, что в силу определения (5)

$$A[M, N_+^{(0)}] \times z_0[N_+^{(0)}] = A[M, j_0], \tag{7}$$

так что $z_0[N_+^{(0)}]$ есть вектор коэффициентов разложения столбца $A[M, j_0]$ по столбцам базисной матрицы $A[M, N_+^{(0)}]$. Доопределим вектор $z_0[N_+^{(0)}]$, положив

$$z_0[j] = \begin{cases} -1 & \text{при } \ j = j_0, \\ 0 & \text{при остальных } \ j \in N \setminus N_+^{(0)}. \end{cases}$$

Тогда соотношение (6) примет вид

$$\Delta_0[j_0] = c[N] \times z_0[N] > 0. \tag{8}$$

Соотношение (7) перепишется так:

$$A[M, N] \times z_0[N] = \mathbb{O}[M]. \tag{9}$$

4°. Введём луч

$$x(t) = x_0 - t\, z_0\,, \quad t > 0,$$

с направляющим вектором $-z_0$. Имеем

$$f\big(x(t)\big) = f(x_0) - t\,\Delta_0[j_0]. \tag{10}$$

Согласно (8) целевая функция при увеличении t убывает.

Отметим также, что согласно (9) при всех $t > 0$

$$A\,x(t) = b. \tag{11}$$

Предположим, что

$$z_0[N_+^{(0)}] \leqslant \mathbb{O}[N_+^{(0)}]. \tag{12}$$

Тогда и $z_0[N] \leqslant \mathbb{O}[N]$. Как следствие, $x(t) \geqslant \mathbb{O}$ при всех $t > 0$. Получаем, что вектор $x(t)$ является планом задачи (1) при всех $t > 0$. При этом в силу (10) $f\big(x(t)\big) \to -\infty$ при $t \to +\infty$.

Приходим к следующему заключению: при выполнении условия (12) задача (1) не имеет решения (целевая функция не ограничена снизу на множестве планов).

5°. Допустим, что найдётся индекс $s \in N_+^{(0)}$, на котором $z_0[s] > 0$. Обозначим через Γ_0 множество всех таких индексов:

$$\Gamma_0 = \bigl\{ s \in N_+^{(0)} \mid z_0[s] > 0 \bigr\}.$$

Очевидно, что $z_0[s] \leqslant 0$ при $s \in N_+^{(0)} \setminus \Gamma_0$.

При $t > 0$ имеем

$$x_0[s] - t\, z_0[s] > 0, \qquad s \in N_+^{(0)} \setminus \Gamma_0\,.$$

Кроме того,

$$x_0[j] - t\, z_0[j] = \begin{cases} 0 & \text{при } \ j \in \bigl(N \setminus N_+^{(0)}\bigr) \setminus \{j_0\}, \\ t & \text{при } \ j = j_0. \end{cases}$$

При $s \in \Gamma_0$ неравенство $x_0[s] - t\, z_0[s] \geqslant 0$ эквивалентно следующему

$$t \leqslant \frac{x_0[s]}{z_0[s]}\,.$$

Положим

$$t_0 = \min\biggl\{ \frac{x_0[s]}{z_0[s]} \;\Big|\; s \in \Gamma_0 \biggr\}.$$

Обозначим s_0 индекс, на котором достигается этот минимум (ниже будет показано, что при выполнении условия невырожденности такой индекс единствен). Очевидно, что $t_0 > 0$.

Введём вектор

$$x_1 = x_0 - t_0\, z_0\,. \tag{13}$$

В силу выбора t_0 имеем $x_1 \geqslant \mathbb{O}$, причём $x_1[s_0] = 0$. К этому нужно добавить, что согласно (11) $Ax_1 = b$. Значит, вектор x_1 является планом задачи (1).

Перепишем (13) в координатной форме:

$$\begin{aligned} x_1[s] &= x_0[s] - t_0\, z_0[s], \quad s \in N_+^{(0)} \setminus \{s_0\}; \\ x_1[j_0] &= t_0; \\ x_1[j] &= 0 \quad \text{при остальных } j \in N \text{ (включая } j = s_0). \end{aligned} \tag{14}$$

Покажем, что x_1 — *базисный* план*.

Обозначим $N_+^{(1)} = \bigl(N_+^{(0)} \setminus \{s_0\}\bigr) \cup \{j_0\}$. Согласно (14), $N_+(x_1) \subset N_+^{(1)}$. Базисность плана x_1 будет установлена, если выяснится, что столбцы матрицы $A[M, N_+^{(1)}]$ линейно независимы.

*Идея приводимого доказательства предложена И. В. Агафоновой.

Запишем

$$\sum_{j\in N_+^{(0)}\setminus\{s_0\}} \alpha_j\, A[M,j] \,+\, \beta\, A[M,j_0] = \mathbb{O}[M] \tag{15}$$

и покажем, что это равенство возможно только тогда, когда все коэффициенты α_j, β равны нулю. Умножим обе части (15) слева на матрицу $B_0[N_+^{(0)},M]$. С учётом (5) получим

$$\sum_{j\in N_+^{(0)}\setminus\{s_0\}} \alpha_j\, e_j[N_+^{(0)}] \,+\, \beta\, z_0[N_+^{(0)}] = \mathbb{O}[N_+^{(0)}], \tag{16}$$

где $e_j[N_+^{(0)}]$ — единичные орты. В частности,

$$\sum_{j\in N_+^{(0)}\setminus\{s_0\}} \alpha_j\, e_j[s_0] \,+\, \beta\, z_0[s_0] = 0.$$

Сумма по $j \in N_+^{(0)} \setminus \{s_0\}$ равна нулю, а величина $z_0[s_0]$ по определению s_0 положительна. Значит, $\beta = 0$. Теперь из (16) следует, что и все коэффициенты α_j равны нулю. Линейная независимость столбцов матрицы $A[M,N_+^{(1)}]$, а с нею и базисность плана x_1, установлены.

В силу условия невырожденности $N_+(x_1) = N_+^{(1)}$. Действительно, нужно принять во внимание, что $N_+(x_1) \subset N_+^{(1)}$ и

$$\bigl|N_+(x_1)\bigr| = \bigl|M\bigr| = \bigl|N_+^{(0)}\bigr| = \bigl|N_+^{(1)}\bigr|.$$

Теперь понятно, почему минимум в определении t_0 достигается на единственном индексе. Иначе у плана x_1 появились бы лишние нулевые компоненты.

Отметим также, что согласно (10)

$$f(x_1) = f(x_0) - t_0\, \Delta_0[j_0].$$

Как следствие, $f(x_1) < f(x_0)$.

6°. Обратная базисная матрица

$$B_1[N_+^{(1)},M] = \bigl(A[M,N_+^{(1)}]\bigr)^{-1}$$

порождена базисной матрицей $A[M,N_+^{(1)}]$, которая отличается от базисной матрицы $A[M,N_+^{(0)}]$ только одним столбцом. Естественно, что обратные базисные матрицы $B_1[N_+^{(1)},M]$ и $B_0[N_+^{(0)},M]$ должны быть связаны между собой. Чтобы разобраться в этом, потребуется некоторая подготовка.

Пусть $C = C[1:m, 1:m]$ — обратимая матрица со столбцами $C_1, \dots, C_m$. Заменим в ней s-й столбец C_s столбцом P, разложение которого по базису $C_1, \dots, C_m$ имеет вид

$$P = \sum_{j=1}^{m} z_j\, C_j\,.$$

Полученную матрицу обозначим D. Нетрудно понять, что

$$D = C\,U, \tag{17}$$

где матрица U отличается от единичной только s-м столбцом, который у U равен $(z_1, \dots, z_m)^T$. Таким образом,

$$U = \begin{pmatrix} 1 & & z_1 & & \\ & \ddots & \vdots & & \\ & & z_s & & \\ & & \vdots & \ddots & \\ & & z_m & & 1 \end{pmatrix}.$$

ЛЕММА 1. *Если $z_s \neq 0$, то матрица D обратима и*

$$D^{-1} = V\,C^{-1}, \tag{18}$$

где матрица V отличается от единичной только s-м столбцом, который у V равен

$$\Bigl(-\frac{z_1}{z_s}, \dots, -\frac{z_{s-1}}{z_s}, \frac{1}{z_s}, -\frac{z_{s+1}}{z_s}, \dots, -\frac{z_m}{z_s}\Bigr)^T.$$

Доказательство. Достаточно проверить, что $U^{-1} = V$, то есть что

$$\begin{pmatrix} 1 & & z_1 & & \\ & \ddots & \vdots & & \\ & & z_s & & \\ & & \vdots & \ddots & \\ & & z_m & & 1 \end{pmatrix} \begin{pmatrix} 1 & & -z_1/z_s & & \\ & \ddots & \vdots & & \\ & & 1/z_s & & \\ & & \vdots & \ddots & \\ & & -z_m/z_s & & 1 \end{pmatrix} = E. \tag{19}$$

После этого справедливость леммы будет следовать из (17).

У произведения матриц UV из левой части (19) j-й столбец при $j \neq s$ равен единичному орту e_j. Запишем представление для s-го столбца:

$$\bigl(UV\bigr)_s = \sum_{j\neq s} \Bigl(-\frac{z_j}{z_s}\Bigr)\, e_j + \frac{1}{z_s} \sum_{j=1}^{m} z_j\, e_j\,.$$

Очевидно, что $(UV)_s = e_s$. Лемма доказана. □

Матрица V называется *мультипликатором.*

Распишем равенство (18) по строкам:

$$\begin{aligned} D^{-1}[s,\cdot] &= \frac{1}{z_s}\,C^{-1}[s,\cdot]; \\ D^{-1}[j,\cdot] &= C^{-1}[j,\cdot] - z_j\,D^{-1}[s,\cdot] \quad \text{при } j \neq s. \end{aligned} \tag{20}$$

Строка $D^{-1}[s,\cdot]$ называется *рабочей строкой.*

Видим, что матрица D^{-1} легко пересчитывается по матрице C^{-1}.

7°. Обратимся к матрице $A[M, N_+^{(1)}]$. Она получается из обратимой базисной матрицы $A[M, N_+^{(0)}]$ заменой столбца с индексом s_0 на столбец с индексом j_0. При этом известно разложение вводимого столбца $A[M, j_0]$ по столбцам матрицы $A[M, N_+^{(0)}]$ (см. (7)). В указанном разложении коэффициент $z_0[s_0]$ у выводимого столбца по определению s_0 положителен. Мы находимся в условиях леммы, согласно которой матрица $A[M, N_+^{(1)}]$ обратима. Более того, на основании (20) для строк обратной матрицы $B_1[N_+^{(1)}, M]$ справедливы формулы пересчёта:

$$\begin{aligned} B_1[j_0, M] &= \frac{1}{z_0[s_0]}\,B_0[s_0, M] \quad \text{(рабочая строка)}; \\ B_1[j, M] &= B_0[j, M] - z_0[j]\,B_1[j_0, M] \quad \text{при } j \in N_+^{(0)} \setminus \{s_0\}. \end{aligned} \tag{21}$$

Строка $B_1[j_0, M]$ обратной матрицы соответствует столбцу $A[M, j_0]$, заменившему в матрице $A[M, N_+^{(0)}]$ столбец $A[M, s_0]$.

8°. Выведем формулу пересчёта для двойственного вектора. Аналогично (3) имеем

$$u_1[M] = c[N_+^{(1)}] \times B_1[N_+^{(1)}, M].$$

ЛЕММА 2. *Справедлива формула*

$$u_1[M] = u_0[M] - \Delta_0[j_0]\,B_1[j_0, M].$$

Доказательство. В силу (21)

$$
\begin{aligned}
u_1[M] &= \sum_{j\in N_+^{(0)}\setminus\{s_0\}} c[j]\,B_1[j,M] + c[j_0]\,B_1[j_0,M] = \\
&= \sum_{j\in N_+^{(0)}\setminus\{s_0\}} c[j]\,\big(B_0[j,M] - z_0[j]\,B_1[j_0,M]\big) + c[j_0]\,B_1[j_0,M] = \\
&= \sum_{j\in N_+^{(0)}} c[j]\,B_0[j,M] - c[s_0]\,B_0[s_0,M] - \\
&\qquad - \left(\sum_{j\in N_+^{(0)}\setminus\{s_0\}} c[j]\,z_0[j] \;-\; c[j_0]\right) B_1[j_0,M].
\end{aligned}
$$

Воспользуемся равенством $B_0[s_0,M] = z_0[s_0]\,B_1[j_0,M]$ и формулой (6). Получим

$$
\begin{aligned}
u_1[M] = u_0[M] - &\left(\sum_{j\in N_+^{(0)}} c[j]\,z_0[j] - c[j_0]\right) B_1[j_0,M] = \\
&= u_0[M] - \Delta_0[j_0]\,B_1[j_0,M].
\end{aligned}
$$

Лемма доказана. $\square$

9°. Опишем общий шаг модифицированного симплекс-метода — переход от базисного плана x_k к базисному плану x_{k+1} с меньшим значением целевой функции. Считаем, что известны

$$
x_k\,,\; N_+^{(k)},\; f(x_k),\; B_k[N_+^{(k)},M],\; u_k[M]. \tag{22}
$$

1) Последовательно вычисляем оценки

$$
\Delta_k[j] = u_k[M] \times A[M,j] - c[j], \qquad j \in N \setminus N_+^{(k)}.
$$

Если все $\Delta_k[j]$ неположительны, то x_k — оптимальный план. Для проверки правильности вычислений можно использовать соотношение двойственности

$$
b[M] \times u_k[M] = f(x_k).
$$

Процесс завершён. Иначе переходим к следующему пункту.

2) Берём индекс $j_k \in N \setminus N_+^{(k)}$, на котором $\Delta_k[j_k] > 0$. Вычисляем

$$
z_k[N_+^{(k)}] = B_k[N_+^{(k)},M] \times A[M,j_k].
$$

Если $z_k[s] \leqslant 0$ при всех $s \in N_+^{(k)}$, то задача (1) не имеет решения (целевая функция не ограничена снизу на множестве планов). Процесс закончен. Иначе переходим к следующему пункту.

3) Вычисляем t_k по формуле

$$t_k = \min\left\{ \frac{x_k[s]}{z_k[s]} \;\middle|\; s \in N_+^{(k)},\ z_k[s] > 0 \right\}.$$

Обозначим s_k индекс, на котором достигается минимум.

4) Находим очередной базисный план x_{k+1}:

$$\begin{aligned} x_{k+1}[s] &= x_k[s] - t_k\, z_k[s], \quad s \in N_+^{(k)} \setminus \{s_k\}; \\ x_{k+1}[j_k] &= t_k\,, \\ x_{k+1}[j] &= 0 \quad \text{при остальных } \ j \in N. \end{aligned}$$

Обозначим $N_+^{(k+1)} = \big(N_+^{(k)} \setminus \{s_k\}\big) \cup \{j_k\}$.

5) Пересчитываем обратную базисную матрицу

$$\begin{aligned} B_{k+1}[j_k, M] &= \frac{1}{z_k[s_k]}\, B_k[s_k, M], \\ B_{k+1}[j, M] &= B_k[j, M] - z_k[j]\, B_{k+1}[j_k, M], \quad j \in N_+^{(k)} \setminus \{s_k\}. \end{aligned}$$

6) Пересчитываем значение целевой функции и двойственный вектор

$$\begin{aligned} f(x_{k+1}) &= f(x_k) - t_k\, \Delta_k[j_k], \\ u_{k+1}[M] &= u_k[M] - \Delta_k[j_k]\, B_{k+1}[j_k, M]. \end{aligned}$$

Проделав указанные действия, получим информацию вида (22):

$$x_{k+1}\,,\ N_+^{(k+1)},\ f(x_{k+1}),\ B_k[N_+^{(k+1)}, M],\ u_{k+1}[M].$$

Теперь можно переходить к очередной итерации.

Описанный метод сходится за конечное число шагов. Действительно, метод прекращает работу в двух случаях: либо когда очередной базисный план оптимален, либо когда выясняется, что задача не имеет решения. Иначе производится переход к следующему базисному плану с меньшим значением целевой функции. Конечность множества базисных планов и строгое уменьшение целевой функции гарантируют сходимость симплекс-метода за конечное число шагов.

10°. Остаётся разобраться с начальным базисным планом. Для его нахождения существует универсальный приём.

Будем считать, что в ограничениях задачи (1) все компоненты вектора b положительны. Рассмотрим вспомогательную задачу линейного программирования

$$\begin{gathered}\sum_{i\in M} y[i] \to \inf,\\ A[M,N]\times x[N] + E[M,M]\times y[M] = b[M],\\ x[N]\geqslant \mathbb{O}[N],\quad y[M]\geqslant \mathbb{O}[M].\end{gathered}\tag{23}$$

Множество её планов непусто (содержит $x = \mathbb{O}$, $y = b$) и целевая функция ограничена снизу (неотрицательна) на множестве планов. Значит, задача (23) имеет оптимальный базисный план. Его можно найти с помощью модифицированного симплекс-метода, взяв в качестве начального базисного плана $x = \mathbb{O}$, $y = b$. Базисной матрицей, равно как и обратной базисной матрицей, будет $E[M, M]$.

Значение целевой функции задачи (23) на оптимальном базисном плане будет либо положительным, либо равным нулю. Первый случай возможен только тогда, когда множество планов исходной задачи (1) пусто, то есть когда задача (1) не имеет решения. Во втором случае оптимальный базисный план имеет вид $(x_0, \mathbb{O})$, где x_0 — план задачи (1). Для носителя $N_+(x_0) =: N_+^{(0)}$ этого плана, вообще говоря, выполняется условие $|N_+^{(0)}| = |M|$. При этом по ходу реализации симплекс-метода найдена обратная базисная матрица $B_0[N_+^{(0)}, M]$.

План x_0 можно взять в качестве начального базисного плана для решения задачи (1). Вычислив

$$\begin{gathered}f(x_0) = c[N_+^{(0)}]\times x_0[N_+^{(0)}],\\ u_0[M] = c[N_+^{(0)}]\times B_0[N_+^{(0)}, M],\end{gathered}$$

получим информацию вида (22) при $k = 0$. После этого можно приступить к решению задачи (1) с помощью модифицированного симплекс-метода.

11°. Конечная сходимость симплекс-метода доказана в предположении, что выполняется условие невырожденности. Однако симплекс-метод прекрасно работает и в вырожденном случае. Зацикливание не исключено, но оно возникает в редких случаях. Один из первых примеров такого рода описан в [2, с. 146–151]. Впрочем, симплекс-метод можно немного усовершенствовать, чтобы избежать зацикливания ([1, с. 71–81]). Усовершенствование состоит в том, что неоднозначный выбор индекса с положительной оценкой заменяется специальным образом организованным однозначным выбором.

ЛИТЕРАТУРА

1. Булавский В. А., Звягина Р. А., Яковлева М. А. *Численные методы линейного программирования.* М.: Наука, 1977. 368 с.

2. Гасс С. *Линейное программирование.* М.: Физматгиз, 1961. 303 с.