МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
ПРОБЛЕМЫ И РЕЗУЛЬТАТЫ

Москва, “Наука”, 2003 с. 36–42

ОПРЕДЕЛЕНИЕ НАПРАВЛЕНИЯ НАИСКОРЕЙШЕГО СПУСКА С ПОМОЩЬЮ ТЕОРЕМ ОБ АЛЬТЕРНАТИВАХ[1]

А.И. Голиков, Ю.Г. Евтушенко
Вычислительный центр им. А.А. Дородницына РАН

Теоремы об альтернативах линейных систем равенств и неравенств являются важным инструментом исследования в теории оптимизации. Они позволяют создавать новые эффективные численные методы нахождения решений систем равенств и неравенств, задач линейного программирования, проводить коррекцию несобственных систем, строить разделяющие гиперплоскости для многогранников [1, 2]. В данной работе показано, что новые теоремы об альтернативах существенно упрощают реализацию метода наискорейшего спуска, используемого в нелинейном программировании. С помощью такого подхода на каждом шаге решается лишь задача минимизации квадратичной функции на неотрицательном ортанте и из ее решения легко определяется направление наискорейшего спуска.

1. Следуя [1, 2], приведем некоторые вспомогательные результаты о теореме об альтернативах. Пусть A — матрица $m \times n$ — задана в следующем виде:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Здесь прямоугольные матрицы A_{11}, A_{12}, A_{21}, A_{22} соответственно имеют размерности: $m_1 \times n_1$, $m_1 \times n_2$, $m_2 \times n_1$, $m_2 \times n_2$. Пусть векторы $y \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $b \in \mathbb{R}^m$ имеют разбиение $y^\top = [y_1^\top, y_2^\top]$, $u^\top = [u_1^\top, u_2^\top]$, $b^\top = [b_1^\top, b_2^\top]$, где $y_1 \in \mathbb{R}^{n_1}$, $y_2 \in \mathbb{R}^{n_2}$, $n = n_1 + n_2$, $u_1 \in \mathbb{R}^{m_1}$, $u_2 \in \mathbb{R}^{m_2}$, $b_1 \in \mathbb{R}^{m_1}$ и $b_2 \in \mathbb{R}^{m_2}$, $m = m_1 + m_2$. Введем два вспомогательных множества $\Pi_y = \{[y_1, y_2] : y_1 \in \mathbb{R}^{n_1}_+,\ y_2 \in \mathbb{R}^{n_2}\}$, $\Pi_u = \{[u_1, u_2] : u_1 \in \mathbb{R}^{m_1}_+,\ u_2 \in \mathbb{R}^{m_2}\}$.

Рассмотрим систему линейных равенств и неравенств

$$A_{11}y_1 + A_{12}y_2 \geq b_1, \quad A_{21}y_1 + A_{22}y_2 = b_2, \qquad y_1 \geq 0_{n_1}. \tag{I}$$

Определим *сопряженную систему* к (I)

$$A_{11}^\top z_1 + A_{21}^\top z_2 \leq 0_{n_1}, \quad A_{12}^\top z_1 + A_{22}^\top z_2 = 0_{n_2}, \qquad z_1 \geq 0_{m_1}. \tag{I$'$}$$

Введем *альтернативную* к (I) систему

$$A_{11}^\top u_1 + A_{21}^\top u_2 \leq 0_{n_1}, \quad A_{12}^\top u_1 + A_{22}^\top u_2 = 0_{n_2}, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho, \ \ u_1 \geq 0_{m_1}. \tag{II}$$

Здесь $\rho > 0$ — произвольное фиксированное положительное число и $\|b\| \neq 0$.

Введем вектор $w \in \mathbb{R}^{n+1}$, представимый в виде $w^\top = [w_1^\top, w_2^\top, w_3]$, где $w_1 \in \mathbb{R}^{n_1}$, $w_2 \in \mathbb{R}^{n_2}$, $w_3 \in \mathbb{R}^1$, и вспомогательное множество $\Pi_w = \{[w_1, w_2, w_3] : w_1 \in \mathbb{R}^{n_1}_+,\ w_2 \in \mathbb{R}^{n_2},\ w_3 \in \mathbb{R}^1\}$.

[1]Работа выполнена при финансовой поддержке Программы Президиума РАН "Математическое моделирование".

Для системы (II) сопряженная система имеет вид

$$A_{11}w_1 + A_{12}w_2 - b_1w_3 \geq 0_{m_1}, \quad A_{21}w_1 + A_{22}w_2 - b_2w_3 = 0_{m_2}, \qquad w_1 \geq 0_{n_1}. \qquad (\mathrm{II}')$$

Множества решений систем (I), (I′), (II) и (II′) обозначим соответственно через Y, Z, U и W. В отличие от (I) и (II) системы (I′) и (II′) всегда имеют решения, так как $0_m \in Z$ и $0_{n+1} \in W$.

Лемма 1. *Системы* (I) *и* (II) *одновременно неразрешимы.*

Ниже из теоремы 3 будет следовать, что всегда имеет решение одна и только одна из систем: либо (I), либо (II). Поэтому эти системы являются альтернативными. Система, альтернативная к (II), сводится к исходной системе (I).

Через $\mathrm{pen}\,(y, Y)$ обозначим штраф в точке $y \in \Pi_y$ за нарушение условия $y \in Y$. В качестве штрафа будем использовать евклидову норму вектора невязок

$$\mathrm{pen}(y, Y) = [\|(b_1 - A_{11}y_1 - A_{12}y_2)_+\|^2 + \|b_2 - A_{21}y_1 - A_{22}y_2\|^2]^{1/2}.$$

Аналогично определим

$$\mathrm{pen}\,(u, U) = \left[\|(A_{11}^\top u_1 + A_{21}^\top u_2)_+\|^2 + \|A_{12}^\top u_1 + A_{22}^\top u_2\|^2 + (\rho - b_1^\top u_1 - b_2^\top u_2)^2\right]^{1/2}.$$

Здесь и ниже a_+ есть неотрицательная часть вектора a, т.е. i-я компонента вектора a_+ совпадает с i-й компонентой вектора a, если она неотрицательна, и равна нулю в противном случае.

Введем следующие четыре задачи:

$$I_1 = \min_{y \in \Pi_y} [\mathrm{pen}\,(y, Y)]^2/2, \qquad (1)$$

$$I_2 = \min_{u \in \Pi_u} [\mathrm{pen}\,(u, U)]^2/2, \qquad (2)$$

$$I_1^d = \max_{z \in Z} \{b^\top z - \|z\|^2/2\}, \qquad (3)$$

$$I_2^d = \max_{w \in W} \{\rho w_3 - \|w\|^2/2\}. \qquad (4)$$

В отличие от систем (I), (II), которые могут быть разрешимы или неразрешимы, задачи (1) – (4) всегда имеют решения, причем задачи (3) и (4) всегда имеют единственные решения, так как в них допустимые множества Z и W непусты и строго вогнутые квадратичные целевые функции ограничены сверху. Задачи (1) и (2) являются двойственными к задачам (3) и (4) соответственно.

Проекцией точки $\bar{y}$ на непустое замкнутое множество Y назовем точку $y^* \in Y$, ближайшую к точке $\bar{y}$, т.е. y^* является решением задачи $\min\limits_{y \in Y} \|\bar{y} - y\|^2/2 = \|\bar{y} - y^*\|^2/2$. Введем обозначение $y^* = \mathrm{pr}\,(\bar{y}, Y)$, расстояние от точки $\bar{y}$ до множества Y обозначим $\mathrm{dist}\,(\bar{y}, Y) = \|y^* - \bar{y}\|$. Следующие две теоремы характеризуют двойственность задач соответственно (1), (3) и (2), (4).

Теорема 1. *Всякое решение y^* задачи* (1) *определяет единственное решение $z^{*\top} = [z_1^{*\top}, z_2^{*\top}]$ задачи* (3) *по формулам:*

$$z_1^* = (b_1 - A_{11}y_1^* - A_{12}y_2^*)_+, \qquad z_2^* = b_2 - A_{21}y_1^* - A_{22}y_2^* \qquad (5)$$

и справедливы следующие утверждения:

$$\|z^*\|^2 = b^\top z^*, \qquad (6)$$

$$z^*\perp Ay^*, \qquad z^*\perp(b-z^*),$$
$$z^*=\mathrm{pr}\,(b,Z), \quad \|z^*\|=\mathrm{pen}\,(y^*,Y), \quad \|b-z^*\|=\mathrm{dist}\,(b,Z),$$
$$[\mathrm{pen}\,(y^*,Y)]^2+[\mathrm{dist}\,(b,Z)]^2=\|b\|^2.$$

Соотношение (6) следует из равенства оптимальных значений целевых функций прямой (3) и двойственной (1) задач. Это равенство в силу (5) выражено только через z^* — решение задачи (3).

Введем матрицу $\bar{A}=[-A,b]$ размера $m\times(n+1)$ и вектор $r\in\mathbb{R}^{n+1}$, представимый в виде $r^\top=[0_n^\top,\rho]$.

Теорема 2. *Пусть* $u^{*\top}=[u_1^{*\top},u_2^{*\top}]$ *— произвольное решение задачи* (2)*. Тогда решение* $w^{*\top}=[w_1^{*\top},w_2^{*\top},w_3^*]$ *задачи* (4) *выражается через решение* u^* *задачи* (2) *по формулам*

$$w_1^*=(A_{11}^\top u_1^*+A_{21}^\top u_2^*)_+, \quad w_2^*=A_{12}^\top u_1^*+A_{22}^\top u_2^*, \quad w_3^*=\rho-b_1^\top u_1^*-b_2^\top u_2^* \tag{7}$$

и справедливы следующие утверждения:

$$\|w^*\|^2=\rho w_3^*; \tag{8}$$

$$w^*\perp\bar{A}^\top u^*, \qquad w^*\perp(r-w^*),$$
$$w^*=\mathrm{pr}\,(r,W), \quad \|w^*\|=\mathrm{pen}\,(u^*,U), \quad \|r-w^*\|=\mathrm{dist}\,(r,W),$$
$$[\mathrm{pen}\,(u^*,U)]^2+[\mathrm{dist}\,(r,W)]^2=\|r\|^2,$$
$$\|w^*\|\le\rho, \quad 0\le w_3^*\le\rho, \quad \|w_1^*\|^2+\|w_2^*\|^2\le\rho^2/4.$$

Соотношение (8) следует из равенства оптимальных значений целевых функций прямой (4) и двойственной (2) задач. В силу (7) это равенство выражено только через w^* — решение задачи (4).

Теорема 3 (об альтернативах). *Пусть* y^* *и* u^* *— произвольные решения задач* (1) *и* (2) *соответственно, векторы минимальных невязок* z^* *и* w^* *определяются по формулам* (5) *и* (7)*. Тогда*

1. *системы* (I) *и* (II) *— альтернативны, т.е. разрешима одна и только одна из них;*
2. *если система* (I) *несовместна, то* $\|z^*\|\ne 0$*, нормальное решение* $\tilde{u}^*$ *системы* (II) *и вектор минимальных невязок* z^* *системы* (I) *коллинеарны и*
$$\tilde{u}^*=\rho z^*/\|z^*\|^2, \qquad z^*=\rho\tilde{u}^*/\|\tilde{u}^*\|^2;$$
3. *если система* (II) *несовместна, то вектор* $\tilde{y}^*=\mathrm{pr}\,(0_n,Y)$ *— нормальное решение системы* (I)*, его составляющие имеют вид*
$$\tilde{y}_1^*=w_1^*/w_3^*, \qquad \tilde{y}_2^*=w_2^*/w_3^*,$$
где w^* *— единственное решение задачи* (4)*, определяемое через произвольное решение* u^* *задачи* (2) *по формуле* (7)*.*

2. Перейдем теперь к нахождению направления наискорейшего спуска. Пусть не равен нулю вектор z^* — решение задачи (3). Введем нормированные векторы $z_n=z/\|z^*\|$ и

$z_n^* = z^*/\|z^*\|$. Определим допустимое множество нормированных векторов $Z_n = \{z_n \in \in \mathbb{R}^m : z_n \in Z,\ \|z_n\| = 1\}$, где множество Z есть сопряженная система (I$'$).

Рассмотрим следующую вспомогательную задачу:

$$I_3 = \max_{z_n \in Z_n} b^\top z_n. \tag{9}$$

Теорема 4. *Пусть y^* — произвольное решение задачи* (1), *z^* — решение задачи* (3), *определяемое по формуле* (5), *кроме того $\|z^*\| \neq 0$. Тогда решением задачи* (9) *является вектор $z_n^* = z^*/\|z^*\|$, причем $I_3 = b^\top z_n^* = \|z^*\|$.*

Доказательство. Из равенства оптимальных значений целевых функций двойственных задач (1) и (3) следует

$$\frac{1}{2} = \max_{z_n \in Z} \left[\frac{b^\top z_n}{\|z^*\|} - \frac{\|z_n\|^2}{2} \right].$$

Так как $Z_n \subset Z$, то имеем

$$\frac{1}{2} = \max_{z_n \in Z} \left[\frac{b^\top z_n}{\|z^*\|} - \frac{\|z_n\|^2}{2} \right] \geq \max_{z_n \in Z_n} \left[\frac{b^\top z_n}{\|z^*\|} - \frac{\|z_n\|^2}{2} \right] = \max_{z_n \in Z_n} \left[\frac{b^\top z_n}{\|z^*\|} - \frac{1}{2} \right].$$

Отсюда следует

$$\|z^*\| \geq \max_{z_n \in Z_n} b^\top z_n. \tag{10}$$

Если в качестве z_n взять вектор $z^*/\|z^*\|$, который принадлежит допустимому множеству Z_n, то согласно (6) неравенство (10) переходит в равенство. □

Задача вида (9) возникает в методе возможных направлений при решении задачи нелинейного программирования (НЛП)

$$\min_{x \in X} f(x), \qquad X = \{x \in \mathbb{R}^m : h(x) \leq 0_{n_1},\ g(x) = 0_{n_2}\}. \tag{11}$$

Здесь $f : \mathbb{R}^m \to \mathbb{R}^1$, $h : \mathbb{R}^m \to \mathbb{R}^{n_1}$, $g : \mathbb{R}^m \to \mathbb{R}^{n_2}$, функции $f(x)$, $h(x)$, $g(x)$ непрерывно дифференцируемы, множество X не пусто, задача (11) имеет решение.

Предположим, что фиксирована произвольная допустимая точка $x \in X$. Введем вектор множителей Лагранжа $y \in \mathbb{R}^n$, $y^\top = [y_1^\top, y_2^\top]$, где $y_1 \in \mathbb{R}^{n_1}_+$, $y_2 \in \mathbb{R}^{n_2}$, $n = n_1 + n_2$, и определим функцию Лагранжа $L(x, y) = f(x) + h^\top(x) y_1 + g^\top(x) y_2$.

Введем условия дополняющей нежесткости

$$y_1^i h^i(x) = 0, \qquad 1 \leq i \leq n_1. \tag{12}$$

Компонента $h^i(x)$ вектора $h(x)$ называется *активной* в точке $x \in X$, если $h^i(x) = 0$. В силу (12) все компоненты вектора y_1, соответствующие неактивным компонентам вектора $h(x)$, равны нулю. Для простоты будем считать, что в функцию Лагранжа входит вектор $h(x)$, все компоненты которого активны. Приведем условия Куна–Таккера для задачи (11), вычисленные в точке $[x, y]$, где $x \in X$

$$L_p(x, y) = f_p(x) + h_p(x) y_1 + g_p(x) y_2 = 0_m, \qquad y_1 \geq 0_{n_1}. \tag{13}$$

При фиксированном $x \in X$ эти условия будем рассматривать как частный случай системы (I) относительно переменных y.

Введем вектор $x' = x + \tau z$, где τ — длина шага, направление спуска $z \in \mathbb{R}^m$, $\|z\| = 1$. В задаче (11) линеаризуем целевую функцию и функции, задающие ограничения. Величину τ считаем малой; отбрасывая члены высшего порядка малости, приходим к следующей задаче отыскания направления наискорейшего спуска:

$$I_4 = \min_{z \in \hat{Z}_n} z^\top f_x(x), \qquad \hat{Z}_n = \{z \in \mathbb{R}^m : h_x^\top(x) z \le 0_{n_1},\ g_x^\top(x) z = 0_{n_2},\ \|z\| = 1\}. \tag{14}$$

Если задача (14) имеет решение z_n^* и при этом $I_4 < 0$, то такое направление назовем *направлением наискорейшего спуска*. Это значит, что по крайней мере в линейном приближении можно “улучшить” точку x, взяв новый вектор x'. Тогда при достаточно малом шаге τ вектор x' остается в допустимом множестве X и значение целевой функции $f(x') < f(x)$. Если задача (14) не имеет такого решения, тогда точку x невозможно локально улучшить.

Чтобы воспользоваться ранее введенными результатами, будем считать, что $h_x^\top(x) = A_{21}^\top$, $g_x^\top(x) = A_{22}^\top$, $-f_x(x) = b_2$, остальные подматрицы матрицы A и вектор b_1 — нулевые. Тогда альтернативная к (13) система запишется в виде

$$u^\top h_x(x) \le 0_{n_1}^\top, \quad u^\top g_x(x) = 0_{n_2}^\top, \quad -u^\top f_x(x) = \rho > 0.$$

Если эта система разрешима, то согласно теореме 3 ее нормальное решение имеет вид $\tilde{u}^* = \rho z^* / \|z^*\|^2$, где $z^* = -L_x(x, y^*)$ и вектор y^* находится из решения задачи безусловной минимизации (1), которая в данном случае записывается следующим образом:

$$I_1 = \min_{y_1 \in \mathbb{R}_+^{n_1}} \min_{y_2 \in \mathbb{R}^{n_2}} \|L_x(x, y)\|^2 / 2. \tag{15}$$

Нормируем вектор $\tilde{u}^*$, получим $\tilde{u}_n^* = z^* / \|z^*\| = z_n^*$. Вектор z_n^* принадлежит $\hat{Z}_n$ и согласно теореме 4 имеем $I_4 = -I_3 = -\|z^*\|$, т.е. направление z_n^* является направлением наискорейшего спуска в линеаризованной задаче (14). Это направление существует тогда и только тогда, когда система (13) при фиксированном $x \in X$ не может быть разрешима относительно множителей Лагранжа $y_1 \in \mathbb{R}_+^{n_1}$, $y_2 \in \mathbb{R}^{n_2}$ задачи (11), т.е. $I_1 > 0$.

Таким образом, для определения наискорейшего спуска нет необходимости решать задачу условной минимизации (14). Это направление находится из решения задачи безусловной минимизации (15). Отметим, что такой подход особенно эффективен, когда n — количество активных ограничений в точке $x \in X$ — много меньше размерности вектора x, так как минимизация в задаче (15) ведется в n-мерном пространстве.

ЛИТЕРАТУРА

1. Голиков А.И., Евтушенко Ю.Г. *Новый метод решения систем линейных равенств и неравенств* // ДАН. 2001. Т. 381. N° 4. С. 1–4.

2. Голиков А.И., Евтушенко Ю.Г. *Применение теорем об альтернативах к нахождению нормальных решений линейных систем* // Изв. ВУЗов: Математика. 2001. N° 12 (475). С. 21–31.