ДАН, 2004, т. 397, No.6, с. 727–732

МАТЕМАТИКА

УДК 519.854

МЕТОД РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ БОЛЬШОЙ РАЗМЕРНОСТИ

©2004 г. А.И. Голиков, член-корреспондент РАН Ю.Г. Евтушенко

Для решения задачи линейного программирования (ЛП) предлагается использовать вспомогательную функцию, сходную с модифицированной функцией Лагранжа (см., например, [1, 2]). Эта функция является вогнутой, кусочно-квадратичной, один раз непрерывно дифференцируемой, у нее существует обобщенная матрица Гессе, что позволяет для ее максимизации применить обобщенный метод Ньютона.

Этот подход характеризуется следующим. Начиная с некоторого фиксированного значения коэффициента штрафа, после однократной безусловной максимизации вспомогательной функции вычисляется по простым формулам точная проекция заданной точки на множество решений прямой задачи ЛП (теорема 1). В этой теореме при определенном предположении получена формула для порогового значения коэффициента штрафа. Подставляя найденную проекцию во вспомогательную функцию и максимизируя ее, находим точное решение двойственной задачи ЛП (теорема 2). В теореме 3 утверждается, что для приведенного ниже итеративного процесса, начиная с произвольного положительного коэффициента штрафа и произвольного начального вектора прямой задачи, получаются точные решения прямой и двойственной задач за конечное число шагов. Вспомогательная задача максимизации решается с помощью обобщенного метода Ньютона, который для данной задачи глобально сходится за конечное число шагов.

Вычислительные эксперименты показали высокую эффективность метода при решении задачи ЛП с большим числом неотрицательных переменных (несколько миллионов) и средним числом ограничений-равенств (несколько тысяч). Время решения таких задач на компьютере P-IV с тактовой частотой 2.6 ГГц составляло от нескольких десятков до нескольких тысяч секунд. Это объясняется тем, что основная вычислительная трудность предлагаемого метода приходится на решение вспомогательной задачи безусловной максимизации, которая, во-первых, решается быстросходящимся методом Ньютона и при которой, во-вторых, ее размерность определяется количеством ограничений типа равенства, число которых значительно меньше, чем число переменных в прямой задаче ЛП.

Пусть задана прямая задача ЛП в стандартной форме

$$f_* = \min_{x \in X} c^\top x, \qquad X = \{x \in \mathbb{R}^n : Ax = b, \ \ x \geq 0_n\}. \tag{P}$$

Двойственная к ней имеет вид

$$f_* = \max_{u \in U} b^\top u, \qquad U = \{u \in \mathbb{R}^m : A^\top u \leq c\}. \tag{D}$$

Здесь $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ и $b \in \mathbb{R}^m$ заданы, x — вектор прямых переменных, а u — двойственных, через 0_i обозначен i-мерный нулевой вектор. Предположим, что множество решений X_* прямой задачи (P) непусто, следовательно, множество решений U_* двойственной задачи (D) также непусто. Необходимые и достаточные условия оптимальности (условия

Куна-Таккера) для задач (P) и (D) запишем в виде

$$Ax_* - b = 0_m, \qquad x_* \geq 0_n, \qquad x_*^\top v_* = 0, \qquad v_* = c - A^\top u_* \geq 0_n. \tag{1}$$

Здесь в ограничения двойственной задачи (D) введен неотрицательный вектор дополнительных переменных $v = c - A^\top u \geq 0_n$.

Пусть задан произвольный вектор $\hat{x} \in \mathbb{R}^n$. Рассмотрим задачу нахождения проекции $\hat{x}_*$ этой точки $\hat{x}$ на множество решений X_* прямой задачи (P)

$$\frac{1}{2}\|\hat{x}_* - \hat{x}\|^2 = \min_{x \in X_*} \frac{1}{2}\|x - \hat{x}\|^2, \qquad X_* = \{x \in \mathbb{R}^n : Ax = b,\ c^\top x = f_*,\ x \geq 0_n\}. \tag{2}$$

Здесь и всюду ниже используется евклидова норма векторов.

Введем функцию Лагранжа для задачи (2)

$$L(x, p, \beta, \hat{x}) = \frac{1}{2}\|x - \hat{x}\|^2 + p^\top(b - Ax) + \beta(c^\top x - f_*),$$

где $p \in \mathbb{R}^m$ и $\beta \in \mathbb{R}^1$ — множители Лагранжа, а $\hat{x}$ будем считать фиксированным вектором параметров. Выпишем задачу, двойственную к (2):

$$\max_{p \in \mathbb{R}^m} \max_{\beta \in \mathbb{R}^1} \min_{x \in \mathbb{R}^n_+} L(x, p, \beta, \hat{x}). \tag{3}$$

В ней внутренняя задача минимизации решается аналитически, и двойственная задача (3) приводится к виду

$$\max_{p \in \mathbb{R}^m} \max_{\beta \in \mathbb{R}^1} \tilde{L}(p, \beta, \hat{x}), \tag{4}$$

где $\tilde{L}(p, \beta, \hat{x})$ — двойственная функция

$$\tilde{L}(p, \beta, \hat{x}) = b^\top p - \frac{1}{2}\|(\hat{x} + A^\top p - \beta c)_+\|^2 - \beta f_* + \frac{1}{2}\|\hat{x}\|^2.$$

Решение $\hat{x}^*$ прямой задачи (2) выражается через решение $[p^*, \beta^*]$ двойственной задачи (4) по формуле

$$\hat{x}^* = (\hat{x} + A^\top p^* - \beta^* c)_+.$$

Задача безусловной максимизации (4) содержит априори неизвестную величину f_* — оптимальное значение целевой функции задачи ЛП. Задачу (4) можно упростить, избавившись от этого недостатка. Для этого вместо (4) будем решать задачу

$$I_1 = \max_{p \in \mathbb{R}^m} S(p, \beta, \hat{x}), \tag{5}$$

где вектор $\hat{x}$ и скаляр β фиксированы, а функция $S(p, \beta, \hat{x})$ определена следующим образом:

$$S(p, \beta, \hat{x}) = b^\top p - \frac{1}{2}\|(\hat{x} + A^\top p - \beta c)_+\|^2. \tag{6}$$

Без потери общности предположим, что первые ℓ компонент вектора $\hat{x}_*$ строго больше нуля. В соответствии с этим предположением представим векторы $\hat{x}_*$, $\hat{x}$, v_* и c, а также матрицу A в виде

$$\hat{x}_*^\top = \left[[\hat{x}_*^\ell]^\top, [\hat{x}_*^d]^\top\right],\ \hat{x}^\top = \left[[\hat{x}^\ell]^\top, [\hat{x}^d]^\top\right],\ v_*^\top = \left[[v_*^\ell]^\top, [v_*^d]^\top\right],\ c^\top = \left[[c^\ell]^\top, [c^d]^\top\right],\ A = [A_\ell \mid A_d],$$

где $\hat{x}_*^\ell > 0_\ell$, $\hat{x}_*^d = 0_d$, $d = n - \ell$, и согласно условию дополняющей нежесткости ($\hat{x}_*^\top v_* = 0$, $\hat{x}_* \geq 0_n$, $v_* \geq 0_n$) выражение (1) запишется в виде

$$v_*^\ell = c^\ell - A_\ell^\top u_* = 0_\ell, \qquad v_*^d = c^d - A_d^\top u_* \geq 0_d.$$

Введем индексное множество $\sigma = \{\ell + 1 \leq i \leq n : (v_*^d)^i > 0\}$. Определим число

$$\beta_* = \begin{cases} \max\limits_{i \in \sigma} \dfrac{q^i}{(v_*^d)^i}, & \text{если } \sigma \neq \emptyset, \\ \alpha > -\infty, & \text{если } \sigma = \emptyset, \end{cases} \tag{7}$$

где q^i есть i-я компонента вектора $q = \hat{x}^d + A_d^\top (A_\ell A_\ell^\top)^{-1} A_\ell (\hat{x}_*^\ell - \hat{x}^\ell)$ и α — произвольное число.

Теорема 1. *Пусть множество решений X_* задачи* (P) *непусто, ранг матрицы A_ℓ, соответствующий ненулевым компонентам вектора $\hat{x}_*$, равен m. Тогда при любом $\beta \geq \beta_*$ проекция $\hat{x}_*$ точки $\hat{x}$ на множество решений X_* прямой задачи* (P) *определяется по формуле*

$$\hat{x}_* = (\hat{x} + A^\top p\,(\beta) - \beta c)_+, \tag{8}$$

где $p\,(\beta)$ — решение задачи безусловной максимизации (5).

Эта теорема позволяет заменить задачу (4), содержащую априори неизвестное число f_*, на задачу (5), в которой вместо этого числа фигурирует полуинтервал $[\beta_*, +\infty)$, что существенно упрощает вычисления. Теорема обобщает результаты, полученные в [3] и посвященные нахождению нормального решения прямой задачи ЛП (проекции нуля на множество решений задачи (P)). Очевидно, что значение β_*, найденное из формулы (7), может быть отрицательным. Соответствующий пример нахождения проекции нуля приведен в [3].

Формально задача безусловной максимизации (5) не имеет функции Лагранжа и, следовательно, соответствующая двойственная задача не может быть построена. Но в задачу (5) можно ввести дополнительные переменные и с их помощью сконструировать искусственные ограничения, т.е. получить эквивалентную задачу нелинейного программирования, для которой уже можно построить двойственную задачу. Такое построение двойственной задачи не является общепринятым, оно основано на двухшаговом представлении задачи (5) (см., например, [4]). В этом смысле задача (5) является взаимно двойственной к следующей задаче квадратичного программирования:

$$I_2 = \min_{x \in X} \left\{ \beta c^\top x + \frac{1}{2}\|x - \hat{x}\|^2 - \frac{1}{2}\|\hat{x}\|^2 \right\}, \qquad X = \{x \in \mathbb{R}^n : Ax = b, \;\; x \geq 0_n\}. \tag{9}$$

Задачу (9) можно рассматривать как возмущенную, или регуляризованную, задачу (P). Решение $x(\beta)$ задачи (9) единственно и выражается через решение $p\,(\beta)$ задачи (5) по формуле

$$x(\beta) = (\hat{x} + A^\top p\,(\beta) - \beta c)_+.$$

Согласно теореме 1 при $\beta \geq \beta_*$ проекция $\hat{x}_*$ совпадает с решением задачи (9); кроме того, эта проекция выражается через решение задачи (5) по формуле (8).

Если ранг матрицы A_ℓ равен m, то, согласно теореме 1, при $\|v_*\| = 0$ значение β_* может быть любым. Если $q \leq 0_d$, тогда $\beta_* \leq 0$. Если положить $\beta = 0$, то регуляризованная задача (9) превращается в задачу нахождения проекции $\hat{x}_*$ заданного вектора $\hat{x}$ на допустимое множество X задачи (P). Следовательно, в случае $\beta_* \leq 0$ расстояние от заданного вектора $\hat{x}$ до множества решений X_* задачи (P) совпадает с расстоянием от $\hat{x}$ до допустимого множества X задачи (P).

Положим $\hat{x} = 0_n$ в $S(p, \beta, \hat{x})$. Тогда задача (5) примет вид

$$\max_{p \in \mathbb{R}^m} \left\{ b^\top p - \frac{1}{2} \|(A^\top p - \beta c)_+\|^2 \right\}. \tag{10}$$

Пусть в этой задаче $\beta \geq \beta_*$ и $\beta > 0$. Сделаем замену переменных, положив $p = \beta u$. Тогда задача (10) заменяется эквивалентной

$$\max_{u \in \mathbb{R}^m} \left\{ b^\top u - \frac{\beta}{2} \|(A^\top u - c)_+\|^2 \right\}, \tag{11}$$

т.е. приходим к методу внешнего квадратичного штрафа, примененному к задаче (D). Из теоремы 1 следует, что при $\beta > \beta_*$ решение $u(\beta)$ задачи (11) определяет по формуле

$$\tilde{x}_* = \beta \left(A^\top u(\beta) - c \right)_+ \tag{12}$$

точное нормальное решение задачи (P). Если $\beta_* \leq 0$, то при любом положительном β, и если $\beta_* > 0$, то при любом $\beta \geq \beta_*$ с помощью решения $u(\beta)$ задачи (11) получим нормальное решение прямой задачи (P). Из свойств метода внешнего квадратичного штрафа следует, что $u(\beta) = p(\beta)/\beta$ стремится к точному решению u_* двойственной задачи (D) только асимптотически, при $\beta \to +\infty$. Задача (10) была введена в работах [3, 10]. Так как задача (10) эквивалентна (11), то (10) можно рассматривать как новый вариант метода внешней штрафной функции, примененного к двойственной задаче (D).

Задача (11) является взаимно двойственной к следующей классической задаче регуляризации ЛП по Тихонову (см., например, [5]–[7]):

$$\min_{x \in X} \left\{ c^\top x + \frac{1}{2\beta} \|x\|^2 \right\}. \tag{13}$$

При $\beta \geq \beta_* > 0$ решение задачи (13) и нормальное решение задачи (P) совпадают и выражаются через решения задач безусловной максимизации (11) и (10) соответственно по формулам (12) и (8). Если $\beta_* \leq 0$, то указанное свойство выполняется при любом $\beta > 0$.

Следующая теорема утверждает, что если известна какая-нибудь точка $x_* \in X_*$, то можно получить решение двойственной задачи (D) после однократного решения задачи безусловной максимизации (5).

Теорема 2. *Пусть множество решений X_* задачи ЛП* (P) *непусто. Тогда для любых $\beta > 0$ и $\hat{x} = x_* \in X_*$ точное решение двойственной задачи* (D) *находится по формуле $u_* = p(\beta)/\beta$, где $p(\beta)$ — точка, доставляющая максимум функции $S(p, \beta, x_*)$.*

Функцию (6) можно рассматривать как модифицированную функцию Лагранжа для двойственной задачи (D). Запишем итеративный процесс

$$p_{k+1} \in \arg\max_{p \in \mathbb{R}^m} \left\{ b^\top p - \frac{1}{2} \|(x_k + A^\top p - \beta c)_+\|^2 \right\}, \tag{14}$$

$$x_{k+1} = (x_k + A^\top p_{k+1} - \beta c)_+, \tag{15}$$

где x_0 — произвольная начальная точка.

Этот итеративный процесс является конечным и дает точное решение x_* прямой задачи (P) и точное решение u_* двойственной задачи (D).

Теорема 3. *Пусть множество решений X_* прямой задачи* (P) *непусто. Тогда при любом $\beta > 0$ и любой начальной точке x_0 итеративный процесс* (14)–(15) *сходится к $x_* \in$*

$\in X_*$ *за конечное число шагов* ω. *Формула* $u_* = p_{\omega+1}/\beta$ *дает точное решение двойственной задачи* (D).

Если в (14), (15) сделать замену переменных $p = \beta u$, то приходим к методу модифицированной функции Лагранжа для решения задачи ЛП [1, 2]. Заметим, что $x_\omega = x_* \in X_*$ является проекцией точки $x_{\omega-1}$ на множество решений X_* задачи (P).

Максимизируемые функции в задачах (5) и (14) являются вогнутыми кусочно-квадратичными и дифференцируемыми. Выпишем градиент функции $S(p, \beta, x_k)$ из задачи (14)

$$S_p\left(p, \beta, x_k\right) = b - A(x_k + A^\top p - \beta c)_+$$

и обобщенную матрицу Гессе

$$\partial_p^2 S(p, \beta, x_k) = -AD^\sharp(z)A^\top,$$

где через $D^\sharp(z)$ обозначена диагональная $(n \times n)$-матрица с i-м диагональным элементом z^i, равным 1, если $(x_k + A^\top p - \beta c)^i > 0$, и равным 0, если $(x_k + A^\top p - \beta c)^i \le 0$, $i = 1, \ldots, n$. В этом случае модифицированный метод Ньютона имеет вид

$$p_{s+1} = p_s - \left(\partial_p^2 S(p_s, \beta, x_k) - \delta I_m\right)^{-1} S_p(p_s, \beta, x_k). \tag{16}$$

Так как обобщенная матрица Гессе может быть вырожденной, то в (16) используется ее коррекция на величину δI_m, где δ есть малая положительная величина (обычно при расчетах полагалось $\delta = 10^{-4}$) и I_m — единичная матрица порядка m. Критерий окончания работы метода полагался следующим: $\|p_{s+1} - p_s\| \le tol$.

Доказательство конечной глобальной сходимости обобщенного метода Ньютона с выбором шага по правилу Армихо для безусловной минимизации подобной выпуклой кусочно-квадратичной функции можно найти в [8]–[10].

В вычислительном эксперименте задачи ЛП генерировались следующим образом. Задавались числа m и n, определяющие количество строк и столбцов матрицы A, и ρ — плотность заполнения матрицы A ненулевыми элементами. В частности, значение $\rho = 1$ означает, что случайным образом генерировались все элементы матрицы A, а значение $\rho = 0.01$ указывает, что в матрице A генерировался только 1% элементов, а остальные полагались равными нулю. Элементы матрицы A определялись случайным образом из интервала $[50, +50]$. Решение x_* прямой задачи (P) и решение u_* двойственной задачи (D) генерировались следующим образом. Полагали, что в векторе x_* содержится $n - 3m$ нулевых компонент, а остальные выбирали случайным образом из интервала $[0, 10]$. Половину компонент вектора u_* выбирали случайным образом из интервала $[-10, 10]$, а остальные полагали равными нулю. Решения x_* и u_* использовали для вычисления коэффициентов целевой функции c и правых частей b задачи ЛП (P). Векторы b и c определяли по формулам $b = Ax_*$, $c = A^\top u_* + \xi$. При этом если $x^{*i} > 0$, то $\xi^i = 0$; если $x^{*i} = 0$, то компонента ξ^i выбиралась случайным образом из интервала $0 \le \gamma^i \le \xi^i \le \theta^i$. В приведенных ниже результатах расчетов считалось, что все $\gamma^i = 1$ и $\theta^i = 10$. Заметим, что при близких к нулю γ^i величина $\xi^i = (c - A^\top u_*)^i = (v_*^d)^i$ может также оказаться очень малой величиной. Согласно формуле (7) априори неизвестная величина β_* может быть очень большой. Тогда сгенерированная задача ЛП может оказаться трудно решаемой.

Предлагаемый метод решения прямой и двойственной задач ЛП, сочетающий в себе итеративный процесс (14), (15) и обобщенный метод Ньютона, реализован в системе MATLAB 6.5. Для вычислений был использован компьютер с процессором Pentium-4, тактовой частотой 2.6 ГГц, оперативной памятью 1 Гб. Результаты вычислений случайным образом сгенерированных задач ЛП приведены в таблице 1.

Таблица 1

$m \times n \times \rho$	T, с	It	$\|A\hat{x}_* - b\|$	$\|(A^\top u_* - c)_+\|$	$\lvert c^\top \hat{x}_* - b^\top u_*\rvert$	B
$100 \times 10^6 \times 0.01$	29.3	17	1.7×10^{-11}	2.0×10^{-13}	9.7×10^{-11}	5
$300 \times 10^6 \times 0.01$	42.0	13	1.0×10^{-10}	7.0×10^{-13}	2.6×10^{-10}	5
$600 \times 10^6 \times 0.01$	68.4	12	3.1×10^{-10}	1.7×10^{-12}	2.8×10^{-10}	5
$1000 \times 10^6 \times 0.01$	95.8	10	9.4×10^{-10}	3.5×10^{-12}	6.9×10^{-10}	8
$3000 \times 10^4 \times 0.01$	81.5	7	2.0×10^{-9}	9.1×10^{-12}	3.7×10^{-9}	2
$4000 \times 10^4 \times 0.01$	196.2	8	2.9×10^{-9}	1.2×10^{-11}	2.6×10^{-8}	2
$500 \times (3 \cdot 10^6) \times 0.01$	179.1	12	3.2×10^{-10}	1.4×10^{-12}	1.9×10^{-11}	8
$1000 \times (3 \cdot 10^6) \times 0.01$	309.1	11	1.2×10^{-9}	4.1×10^{-12}	4.9×10^{-9}	10
$500 \times (5 \cdot 10^6) \times 0.01$	300.8	12	3.8×10^{-10}	1.6×10^{-12}	8.4×10^{-11}	10
$1000 \times (5 \cdot 10^6) \times 0.01$	412.8	8	7.3×10^{-9}	7.4×10^{-12}	7.0×10^{-8}	100
$500 \times 10^7 \times 0.01$	387.8	8	7.6×10^{-9}	3.6×10^{-12}	1.1×10^{-7}	400
$1000 \times 10^4 \times 1$	117.2	7	1.3×10^{-7}	1.0×10^{-10}	2.9×10^{-7}	2
$1000 \times 10^5 \times 1$	1496.5	5	5.2×10^{-7}	1.9×10^{-10}	8.2×10^{-7}	200
$100 \times 10^6 \times 1$	376.5	9	4.2×10^{-8}	1.2×10^{-11}	3.0×10^{-7}	500

В таблице 1 указаны размерности m, n и ρ (ρ — плотность ненулевых элементов матрицы A), T — время решения задачи ЛП в секундах, It — количество итераций, сделанных методом Ньютона при первом решении задачи максимизации (14). В четвертом, пятом и шестом столбцах приведены точности выполнения ограничений прямой и двойственной задач и разность оптимальных значений целевых функций прямой и двойственной задач. При вычислении произведения матриц $AD^\sharp(z)A^\top$ для ускорения расчетов, а также в случае задач особо высокой размерности ($n \sim 3 \cdot 10^6$) из-за нехватки оперативной памяти матрица A разбивалась на B блоков и умножение производилось поблочно. Соответствующее количество блоков B приведено в последнем столбце таблицы.

Начальную точку в итерационном процессе (14), (15) во всех примерах брали равной нулю: $x_0 = 0_n$. Всюду полагалось $\beta = 1$, $tol = 10^{-12}$. Оказалось, что во всех примерах имело место $\beta > \beta_*$. Таким образом, нормальное решение $\hat{x}_*$ прямой задачи (P) было получено из процесса (14), (15) за одну итерацию, т.е. $\omega = 1$. Число итераций обобщенного метода Ньютона (16) приведено в третьем столбце. Согласно теореме 2 максимизация функции $S(p, \beta, \hat{x}_*)$ по p дает вектор $p(\beta)$, который равен $u_*\beta$. Во всех примерах для этой максимизации обобщенным методом Ньютона (16) потребовалось всего лишь две итерации.

Приведенные в таблице 1 результаты показывают высокую эффективность предложенного метода. Для задач ЛП, обладающих большим количеством неотрицательных переменных и средним количеством ограничений типа равенства, реализация на компьютере предложенного метода превосходила имеющиеся в распоряжении авторов комплексы программ решения задач ЛП, основанные на симплекс-методе.

Дальнейшее развитие предлагаемого метода авторы видят в использовании параллельных расчетов при реализации обобщенного метода Ньютона, что позволит решать задачи ЛП с большим количеством ограничений типа равенства.

Работа выполнена при финансовой поддержке программы “Ведущие научные школы” (грант НШ–1737.2003.1) и РФФИ (грант 03-01-00465).

СПИСОК ЛИТЕРАТУРЫ

1. *Поляк Б.Т., Третьяков Н.В.* // Экон. и мат. методы. 1972. Т.7. В.5. С. 740–751.
2. *Антипин А.С.* Методы нелинейного программирования, основанные на прямой и двойственной модификации функции Лагранжа. М.: ВНИИСИ, 1979.
3. *Голиков А.И., Евтушенко Ю.Г.* // ЖВМ и МФ. 2000. Т.40. N° 12. С. 1766–1786.
4. *Голиков А.И., Евтушенко Ю.Г.* // ЖВМ и МФ. 2003. Т.43. N° 3. С. 354–375.
5. *А.Н.Тихонов.* Избранные труды. М.: МАКС-Пресс, 2001.
6. *Еремин И.И.* Теория линейной оптимизации. Екатеринбург: УрО РАН, 1998.
7. *Васильев Ф.П., Иваницкий А.Ю.* Линейное программирование. М.: Факториал, 2003.
8. *Mangasarian O.L.* // Optim. Meth. and Software. 2002. V.17. P. 913–930.
9. *Mangasarian O.L.* A Newton Method for Linear Programming // Data Mining Institute. Techn. Rep. 02-02. March 2002. ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/02-02.ps.
10. *Kanzow C., Qi H., Qi L.* // J. Optim. Theory and Appl. 2003. V.116. P.333–345.