ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2004, том 44, № 9, с. 1564–1573

УДК 519.658.4

ПРИМЕНЕНИЕ МЕТОДА НЬЮТОНА К РЕШЕНИЮ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ БОЛЬШОЙ РАЗМЕРНОСТИ[1]

© 2004 г. А.И. Голиков, Ю.Г. Евтушенко, Н. Моллаверди

(119991 Москва, ул. Вавилова, 40, ВЦ РАН)
e-mail: gol@ccas.ru, evt@ccas.ru

В статье устранены некоторые опечатки

Для одновременного решения прямой и двойственной задач линейного программирования (ЛП) предлагается использовать новую вспомогательную функцию, близкую к модифицированной функции Лагранжа, и применить обобщенный метод Ньютона для безусловной максимизации этой функции. Предлагаемый подход применим для решения задач ЛП с большим числом (несколько миллионов) неотрицательных переменных и средним числом (несколько тысяч) ограничений типа равенств. Приводятся результаты тестовых расчетов на компьютере P-IV, которые показали, что задачи указанных размерностей решаются за время от нескольких десятков до нескольких тысяч секунд. Библ. 15. Табл. 1.

Ключевые слова: задачи линейного программирования большой размерности, метод Ньютона, функция Лагранжа.

1. ВВЕДЕНИЕ

Известно, что для нахождения точного решения прямой задачи линейного программирования (ЛП) можно использовать гладкую штрафную функцию, примененную к двойственной задаче ЛП, с конечным коэффициентом штрафа (см., например, [1] – [3]). Такая функция является выпуклой, кусочно-квадратичной, непрерывно дифференцируемой, но у нее не существует матрицы Гессе. Однако для такой штрафной функции можно построить обобщенный метод Ньютона, введя обобщенную матрицу Гессе. В работах [4] – [6] доказана конечная глобальная сходимость обобщенного метода Ньютона для минимизации выпуклой кусочно-квадратичной функции. Минимизация этой штрафной функции, примененной к двойственной задаче ЛП, дает возможность получить точное нормальное (с минимальной евклидовой нормой) решение прямой задачи, начиная с некоторого конечного значения коэффициента штрафа.

В данной работе вместо обычно применяемой кусочно-квадратичной штрафной функции предлагается использовать вспомогательную функцию, сходную с модифицированной функцией Лагранжа (см., например, [7] – [9]). Этот подход характеризуется тем, что, начиная с некоторого фиксированного значения коэффициента штрафа после однократной безусловной максимизации вспомогательной функции по простым формулам, вычисляется точная проекция заданной точки на множество решений прямой задачи ЛП (см. ниже теорему 1, где 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).

[1]Работа выполнена при финансовой поддержке РФФИ (код проекта 03-01-00465), по Программе поддержки ведущих научных школ (код проекта НШ-1737.2003.1) и Программе № 17 Президиума РАН.

В теореме 3 утверждается, что для приведенного ниже итеративного процесса, начиная с произвольного 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.

Предлагаемый метод реализован в системе MATLAB 6.5 для компьютера P-IV с оперативной памятью 1 Гб. Численные эксперименты со случайно сгенерированными задачами ЛП показали высокую эффективность метода при решении задач ЛП с большим числом неотрицательных переменных (несколько миллионов) и средним числом ограничений-равенств (несколько тысяч). Время решения таких задач составляло от нескольких десятков до нескольких тысяч секунд. Эти вычислительные результаты объясняются тем, что основная вычислительная трудность предлагаемого метода приходится на решение вспомогательной задачи безусловной максимизации. Ее размерность определяется количеством ограничений типа равенств, число которых существенно меньше, чем число переменных в исходной задаче ЛП.

2. ОСНОВНЫЕ ТЕОРЕМЫ

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

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

$$v_* = c - A^\top u_* \geq 0_n. \tag{2}$$

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):

$$\frac{1}{2}\|\hat{x}_* - \hat{x}\|^2 = \min_{x \in X_*} \frac{1}{2}\|x - \hat{x}\|^2, \qquad X_* = \{x \in \mathbb{R}^n : Ax = b, \;\; c^\top x = f_*, \;\; x \geq 0_n\}. \tag{3}$$

Здесь и всюду ниже используется евклидова норма векторов.

Введем функцию Лагранжа для задачи (3):

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

где $p \in \mathbb{R}^m$ и $\beta \in \mathbb{R}^1$ суть множители Лагранжа, а $\hat{x}$ будем считать фиксированным вектором параметров. Двойственная к (3) задача имеет вид

$$\max_{p \in \mathbb{R}^m} \; \max_{\beta \in \mathbb{R}^1} \; \min_{x \in \mathbb{R}^n_+} \; L(x, p, \beta, \hat{x}). \tag{4}$$

Запишем условия Куна–Таккера для задачи (3):

$$x - \hat{x} - A^\top p + \beta c \geq 0_n, \qquad D(x)(x - \hat{x} - A^\top p + \beta c) = 0_n, \qquad x \geq 0_n, \tag{5}$$

$$Ax = b, \qquad c^\top x = f_*, \tag{6}$$

где через $D(z)$ обозначена диагональная матрица, у которой i-й диагональный элемент есть i-я компонента вектора z. Легко проверить, что формулы (5) эквивалентны выражению

$$x = (\hat{x} + A^\top p - \beta c)_+, \tag{7}$$

где a_+ обозначает вектор a, у которого все отрицательные компоненты заменены на нули.

Формула (7) дает решение внутренней задачи минимизации в задаче (4). Подставляя (7) в функцию Лагранжа $L(x, p, \beta, \hat{x})$, получаем двойственную функцию

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

Функция $\tilde{L}(p, \beta, x)$ вогнутая, кусочно-квадратичная и непрерывно дифференцируемая. Двойственная задача (4) сводится к решению внешней задачи максимизации

$$\max_{p \,\in\, \mathbb{R}^m} \ \max_{\beta \in \mathbb{R}^1} \tilde{L}(p, \beta, \hat{x}). \tag{8}$$

Решив задачу (8), найдем оптимальные p и β, после их подстановки в (7) получаем проекцию $\hat{x}_*$, т.е. решение задачи (3). Необходимые и достаточные условия оптимальности для задачи (8) имеют вид

$$\begin{aligned} \tilde{L}_p(p, \beta, \hat{x}) &= b - A(\hat{x} + A^\top p - \beta c)_+ = b - Ax = 0_m, \\ \tilde{L}_\beta(p, \beta, \hat{x}) &= c^\top(\hat{x} + A^\top p - \beta c)_+ - f_* = c^\top x - f_* = 0, \end{aligned}$$

где x определен формулой (7). Эти условия выполнены тогда и только тогда, когда $x \in X_*$ и $x = \hat{x}_*$.

К сожалению, задача безусловной оптимизации (8) содержит неизвестную априори величину f_* — оптимальное значение целевой функции задачи ЛП. Однако задачу (8) можно упростить, избавившись от этого недостатка. Для этого вместо (8) предлагается решать следующую упрощенную задачу безусловной максимизации:

$$I_1 = \max_{p \in \mathbb{R}^m} S(p, \beta, \hat{x}), \tag{9}$$

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

$$S(p, \beta, \hat{x}) = b^\top p - \frac{1}{2}\|(\hat{x} + A^\top p - \beta c)_+\|^2. \tag{10}$$

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

$$\hat{x}_*^\top = \left[[\hat{x}_*^\ell]^\top, [\hat{x}_*^d]^\top\right], \quad \hat{x}^\top = \left[[\hat{x}^\ell]^\top, [\hat{x}^d]^\top\right], \quad c^\top = \left[[c^\ell]^\top, [c^d]^\top\right], \quad A = [A_\ell \mid A_d], \tag{11}$$

где $\hat{x}_*^\ell > 0_\ell$, $\hat{x}_*^d = 0_d$, $d = n - \ell$.

В этих обозначениях необходимые и достаточные условия оптимальности (5), (6) для задачи (3) можно переписать в развернутом виде:

$$\hat{x}_*^\ell = \hat{x}^\ell + A_\ell^\top p - \beta c^\ell > 0_\ell, \tag{12}$$

$$\begin{gathered} \hat{x}_*^d = 0_d, \qquad \hat{x}^d + A_d^\top p - \beta c^d \leq 0_d, \\ A_\ell \hat{x}_*^\ell = b, \qquad c^{\ell\top} \hat{x}_*^\ell = f_*. \end{gathered} \tag{13}$$

В (12) линейная система уравнений относительно неизвестных p совместна, поэтому если предположить, что матрица A_ℓ имеет полный ранг m и $\ell \geq m$, то единственное решение p этой системы дается формулой

$$p = (A_\ell A_\ell^\top)^{-1} A_\ell (\hat{x}_*^\ell - \hat{x}^\ell + \beta c^\ell). \tag{14}$$

Подставляя эту формулу в (13), получаем неравенство

$$q \leq \beta z, \tag{15}$$

где введены обозначения $q = \hat{x}^d + A_d^\top (A_\ell A_\ell^\top)^{-1} A_\ell (\hat{x}_*^\ell - \hat{x}^\ell)$ и $z = c^d - A_d (A_\ell A_\ell^\top)^{-1} A_\ell c^\ell$.

Если p определено согласно (14) и β удовлетворяет неравенству (15), то пара $[p, \beta]$ является решением двойственной задачи (8). Найдем минимальное значение β, при котором выполнено неравенство (15).

В соответствии с разбиением (11), оптимальный вектор дополнительных переменных v_* из условий Куна–Таккера (1), (2) для задач (P) и (D) представим в виде $v_*^\top = [v_*^{\ell\top}, v_*^{d\top}]$. Тогда, согласно условию дополняющей нежесткости $x_*^\top v_* = 0$, $x_* \geq 0_n$, $v_* \geq 0_n$, выражение (2) запишется в виде

$$v_*^\ell \;=\; c^\ell - A_\ell^\top u_* = 0_\ell, \tag{16}$$

$$v_*^d \;=\; c^d - A_d^\top u_* = 0_d. \tag{17}$$

Из (16) получаем $u_* = (A_\ell A_\ell^\top)^{-1} A_\ell c^\ell$. Подставляя это выражение в (17), получаем $v_*^d = z \geq 0_d$. Определим следующее индексное множество: $\sigma = \{\ell + 1 \leq i \leq n : (v_*^d)^i > 0\}$. Если $\sigma = \emptyset$, то (15) выполняется при любом β. Определим

$$\beta_* = \begin{cases} \max\limits_{i \in \sigma} \dfrac{q^i}{(v_*^d)^i}, & \sigma \neq \emptyset, \\ \alpha > -\infty, & \sigma = \emptyset, \end{cases} \tag{18}$$

где α — произвольное число. Тогда неравенство (15) справедливо при любом $\beta \geq \beta_*$ и можно решать упрощенную задачу безусловной максимизации (9). Ее решение одновременно дает решение двойственной задачи (8). Далее, используя формулу (7), получаем проекцию $\hat{x}_*$. Итак, справедлива следующая

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*

$$\hat{x}_* = [\hat{x} + A^\top p(\beta) - \beta c]_+, \tag{19}$$

где $p(\beta)$ — решение задачи безусловной максимизации (9).

Эта теорема позволяет заменить задачу (8), содержащую априори неизвестное число f_*, на задачу (9), в которой вместо этого числа фигурирует полуинтервал $[\beta_*, +\infty)$, что существенно проще с вычислительной точки зрения. Теорема обобщает результаты, полученные в [10] и посвященные нахождению нормального решения прямой задачи ЛП (проекции нуля на множество решений задачи (P)). Очевидно, что значение β_*, найденное по формуле (18), может быть отрицательным. Соответствующий пример нахождения проекции начала координат приведен в [10].

Формально задача безусловной максимизации (9) не имеет функции Лагранжа и, следовательно, соответствующая двойственная задача не может быть построена. Но можно ввести в задачу (9) 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 (9) (см., например, [11, 12]).

Введем вектор дополнительных переменных $y = \hat{x} + A^\top p - \beta c$. Тогда задача (9) сводится к эквивалентной задаче минимизации при наличии ограничений:

$$I_1 = \max_{[p,y] \in G} \left\{ b^\top p - \frac{1}{2} \|y_+\|^2 \right\}, \qquad G = \{[p, y] \in \mathbb{R}^{m+n} : y = \hat{x} + A^\top p - \beta c\}. \tag{20}$$

Функция Лагранжа для задачи квадратичного программирования (20) имеет вид

$$L(p, y, x) = b^\top p - \frac{1}{2} \|y_+\|^2 - x^\top (\hat{x} + A^\top p - \beta c - y),$$

где $x \in \mathbb{R}^n$ — вектор множителей Лагранжа. Рассмотрим следующую двойственную к (20) минимаксную задачу:

$$I_2 = \min_{x \in \mathbb{R}^n} \max_{p \in \mathbb{R}^m} \max_{y \in \mathbb{R}^n} L(p, y, x). \tag{21}$$

Из условий оптимальности внутренней задачи максимизации

$$L_p(p, y, x) = b - Ax = 0_m, \qquad L_y(p, y, x) = -y_+ + x = 0_n$$

имеем $x = y_+$, $Ax = b$. Подставляя $y_+ = x$ в $L(p, y, x)$ и учитывая условия $Ax = b$, $x \geq 0_n$, получаем двойственную функцию

$$\tilde{L}(x) = \beta c^\top x + \frac{1}{2} \|x - \hat{x}\|^2 - \frac{1}{2} \|\hat{x}\|^2.$$

Итак, двойственная задача (21) свелась к следующей задаче квадратичного программирования:

$$I_2 = \min_{x \in X} \left\{ \beta c^\top x + \frac{1}{2} \|x - \hat{x}\|^2 - \frac{1}{2} \|\hat{x}\|^2 \right\}, \qquad X = \{x \in \mathbb{R}^n : Ax = b, \ x \geq 0_n\}. \tag{22}$$

Так как целевая функция в (22) строго выпукла, то ее решение $x(\beta)$ единственно. Задача (22) двойственна к (20) и, следовательно, к (9). Теорема двойственности утверждает, что оптимальные значения целевых функций этих задач равны: $I_1 = I_2$. Это условие можно записать в виде $b^\top p(\beta) = \beta f_* + x(\beta)^\top [x(\beta) - \hat{x}]$, где $p(\beta)$ — произвольное решение задачи (9). Стандартным образом можно показать, что двойственной к задаче квадратичного программирования (22) является задача безусловной максимизации (9) кусочно-квадратичной функции $S(p, \beta, \hat{x})$, определяемой формулой (10). Итак, задача безусловной максимизации (9) и задача квадратичного программирования (22) могут рассматриваться как взаимно двойственные. Задачу (22) можно рассматривать как возмущенную, или регуляризованную, задачу (P). Решение $x(\beta)$ задачи (22) выражается через произвольное решение $p(\beta)$ задачи (9) по формуле

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

которая, согласно теореме 1, переходит в формулу (19) при $\beta \geq \beta_*$, т.е. в этом случае имеем $x(\beta) = \hat{x}_*$.

Если ранг матрицы A_ℓ равен m, то, согласно теореме 1, при $\|v_*\| = 0$ значение β_* может быть любым, в том числе и отрицательным. Если $q \leq 0_d$, тогда $\beta_* \leq 0$. Если положить $\beta = 0$, то регуляризованная задача (22) превращается в задачу нахождения проекции $\hat{x}_*$ заданного вектора $\hat{x}$ на допустимое множество X задачи (P). Вектор $\hat{x}_*$ одновременно является решением задачи (22) при любом $\beta \geq \beta_*$ и решением задачи (3). 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).

Рассмотрим частный случай теоремы 1, когда $\hat{x} = 0_n$. Из (19) получаем формулу для нахождения нормального решения $\tilde{x}_*$ прямой задачи (P):

$$\tilde{x}_* = [A^\top p(\beta) - \beta c]_+,$$

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

$$\max_{p \in \mathbb{R}^m} \left\{ b^\top p - \frac{1}{2} \|(A^\top p - \beta c)_+\|^2 \right\} \tag{23}$$

при любом $\beta \geq \beta_*$. Задача (23) и задача квадратичного программирования

$$\min_{x \in X} \left\{ \beta c^\top x + \frac{1}{2} \|x\|^2 \right\}, \qquad X = \{x \in \mathbb{R}^n : Ax = b, \ \ x \geq 0_n\} \tag{24}$$

являются взаимно двойственными. При любом $\beta \geq \beta_*$ единственное решение $\tilde{x}_*$ задачи (24) есть проекция нуля на множество решений X_* прямой задачи (P), т.е. является нормальным решением задачи (P) и определяется из следующей задачи:

$$\min_{x \in X_*} \frac{1}{2} \|x\|^2, \qquad X_* = \{x \in \mathbb{R}^n : Ax = b, \ \ x \geq 0_n, \ \ c^\top x = f_*\}. \tag{25}$$

Приведем задачу, которая взаимно двойственна к (25):

$$\max_{p \in \mathbb{R}^m} \ \max_{\beta \in \mathbb{R}^1} \left\{ b^\top p - \frac{1}{2} \|(A^\top p - \beta c)_+\|^2 - \beta f_* \right\}. \tag{26}$$

Решение $\tilde{x}_*$ задачи квадратичного программирования (25) определяется через любое решение $[p, \beta]$ задачи безусловной максимизации (26) по формуле

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

Рассмотрим случай, когда $\beta \geq \beta_*$ и $\beta > 0$. В задаче (23) сделаем замену переменных, положив $p = \beta u$. Тогда задача (23) заменяется эквивалентной задачей

$$\max_{u \in \mathbb{R}^m} \left\{ b^\top u - \frac{\beta}{2} \|(A^\top u - c)_+\|^2 \right\}, \tag{27}$$

т.е. приходим к методу внешнего квадратичного штрафа, примененному к задаче (D). В этом случае из теоремы 1 следует, что вектор $u(\beta)$, полученный в результате максимизации дифференцируемой внешней штрафной функции в задаче (26), определяет по формуле $\tilde{x}_* = \beta[A^\top u(\beta) - c]_+$ нормальное решение задачи (P) при $\beta > \beta_*$. Из известных свойств метода внешнего квадратичного штрафа (см. [13]) следует, что $u(\beta) = p(\beta)/\beta$ стремится к u_* асимптотически при $\beta \to +\infty$. Задача (23) была введена в работах [10, 6]. Так как задача (23) эквивалентна (27), то (23) может рассматриваться как новый вариант внешней штрафной функции, примененной к двойственной задаче (D).

Таким образом, приведенное в (18) соотношение для β_* дает оценки коэффициента штрафа для классического метода внешнего квадратичного штрафа, примененного к задаче (D), т.е. к задаче ЛП с ограничениями типа неравенств. Если $\beta_* \leq 0$, то при любом положительном β, а если $\beta_* > 0$ — то при любом $\beta \geq \beta_*$ с помощью решения $u(\beta)$ задачи (27) получаем нормальное решение прямой задачи (P).

Задача (27) является взаимно двойственной к следующей задаче регуляризации ЛП (см., например, [14, 15, 1]):

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

Если $\beta_* > 0$, то положим $\varepsilon_* = 1/\beta_*$, если $\beta_* \leq 0$, то ε_* можно полагать любым положительным числом, где β_* определяется согласно теореме 1. Итак, приходим к оценке параметра регуляризации ε в классической задаче регуляризации по Тихонову задачи ЛП:

$$\min_{x \in X} \left\{ c^\top x + \frac{\varepsilon}{2}\|x\|^2 \right\}. \tag{28}$$

При $0 < \varepsilon \leq \varepsilon_*$ решение задачи (28) и нормальное решение задачи (P) совпадают.

Точка $p(\beta)$, которая доставляет максимум функции $S(p, \beta, \hat{x})$ при $\hat{x} \notin X_*$, не дает решения двойственной задачи (D), но если $\beta \geq \beta_*$, то по формуле (19) получаем точное решение $\hat{x}_*$ задачи (3) (проекцию $\hat{x}$ на множество решений X_* прямой задачи (P)) и точное нормальное решение задачи (P) при $\hat{x} = 0_n$.

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 (9).

Теорема 2. *Пусть множество решений X_* задачи* (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 $S(p, \beta, x_*)$.*

Доказательство. Необходимые и достаточные условия оптимальности задачи (9) имеют вид

$$b - A(x_* + A^\top p_* - \beta c)_+ = 0_m. \tag{29}$$

Обозначим $x = (x_* + A^\top p_* - \beta c)_+$, это выражение эквивалентно следующим трем векторным условиям:

$$x - (x_* + A^\top p_* - \beta c) \geq 0_n, \qquad x \geq 0_n, \tag{30}$$

$$D(x)[x - (x_* + A^\top p_* - \beta c)] = 0_n. \tag{31}$$

Возьмем в качестве x вектор x_*, в качестве p_* — вектор βu_*. Тогда из (29) – (31) получим, соответственно,

$$Ax_* = b, \qquad c - A^\top u_* \geq 0_n, \qquad x_* \geq 0_n, \qquad c^\top x_* = b^\top u_*.$$

Отсюда следует, что $[x_*, u_*]$ — точка Куна–Таккера для задачи ЛП и u_* — решение двойственной задачи (D). Теорема доказана. $\square$

Функция (10) может рассматриваться как модифицированная функция Лагранжа для двойственной задачи (D). Введем следующий итеративный процесс:

$$p_{k+1} \in \arg\max_{p \in \mathbb{R}^m} \left\{ b^\top p - \frac{1}{2}\|(x_k + A^\top p - \beta c)_+\|^2 \right\}, \quad (32)$$

$$x_{k+1} = (x_k + A^\top p_{k+1} - \beta c)_+, \quad (33)$$

где вектор 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 любых $\beta > 0$ и начальной точке x_0 итеративный процесс* (32), (33) *сходится к $x_* \in X_*$ за конечное число шагов ω. Формула $u_* = p_{\omega+1}/\beta$ дает точное решение двойственной задачи* (D).

Если в (32), (33) сделать замену переменных $p = \beta u$, то придем к методу модифицированной функции Лагранжа для решения задачи ЛП, предложенному в [7, 8]:

$$u_{k+1} \in \arg\max_{u \in \mathbb{R}^m} \left\{ \beta b^\top u - \frac{1}{2}\left\|[x_k + \beta(A^\top u - c)]_+\right\|^2 \right\}, \quad (34)$$

$$x_{k+1} = [x_k + \beta(A^\top u_{k+1} - c)]_+. \quad (35)$$

Доказательство конечной сходимости метода (34), (35) дано в [8]. Перенос доказательства этого свойства для метода (32), (33) очевиден.

Заметим, что $x_\omega = x_* \in X_*$ является проекцией точки $x_{\omega-1}$ на множество решений X_* задачи (P).

3. ОБОБЩЕННЫЙ МЕТОД НЬЮТОНА

Безусловная максимизация в (9) и (32) может выполняться любым методом, например методом сопряженного градиента. Но, как показал О.Л. Мангасарьян, для безусловной оптимизации кусочно-квадратичной функции особенно эффективен обобщенный метод Ньютона (см. [4, 5]). Приведем краткое описание этого метода и результаты численного эксперимента.

Максимизируемые функции $S(p, \beta, x_k)$ в задаче (32) и $S(p, \beta, \hat{x})$ в задаче (9) являются вогнутыми кусочно-квадратичными и дифференцируемыми. Для этих функций обычная матрица Гессе не существует. Действительно, градиент

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

функции $S(p, \beta, x_k)$ не дифференцируем. Но для этой функции можно определить обобщенную матрицу Гессе, которая является $m \times m$-симметричной отрицательно-полуопределенной матрицей следующего вида:

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

где через $D^\sharp(z)$ обозначена $n \times n$-диагональная матрица с i-м диагональным элементом z^i, равным 1, если $(x_k + A^\top p - \beta c)^i > 0$, и равным 0, если $(x_k + A^\top p - \beta c)^i \leq 0$, $i = 1, 2, \ldots, n$. Так как обобщенная матрица Гессе может быть вырожденной, используется следующее модифицированное ньютоновское направление:

$$-[\partial_p^2 S(p, \beta, x_k) - \delta I_m]^{-1} S_p(p, \beta, x_k),$$

где δ — малая положительная величина (обычно при расчетах полагалось $\delta = 10^{-4}$) и I_m — единичная матрица порядка m.

В этом случае модифицированный метод Ньютона имеет вид

$$p_{s+1} = p_s - [\partial_p^2 S(p_s, \beta, x_k) - \delta I_m]^{-1} S_p(p_s, \beta, x_k), \tag{36}$$

Критерий окончания его работы полагался следующим:

$$\|p_{s+1} - p_s\| \leq \text{tol}.$$

Мангасарьян исследовал сходимость обобщенного метода Ньютона для безусловной минимизации подобной выпуклой кусочно-квадратичной функции с выбором шага по правилу Армихо. Доказательство конечной глобальной сходимости обобщенного метода Ньютона можно найти в [4] – [6].

4. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ РАСЧЕТОВ

Решались сгенерированные случайным образом задачи ЛП с большим числом неотрицательных переменных (до нескольких миллионов) и средним числом ограничений-равенств (до нескольких тысяч), т.е. имело место $n \gg m$.

Итак, задавались числа m и n, определяющие количество строк и столбцов матрицы A, и ρ — плотность заполнения матрицы A ненулевыми элементами. В частности, $\rho = 1$ означает, что случайным образом генерировались все элементы матрицы A, а значение $\rho = 0.01$ указывает, что в матрице А 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. Полагалось, что в векторе x_* содержится $n-3m$ нулевых компонент, а остальные компоненты выбирались случайным образом из интервала $[0, 10]$. Половина компонент вектора u_* полагалась равной нулю, а остальные выбирались случайным образом из интервала $[-10, 10]$. Решения x_* и u_* использовались для вычисления коэффициентов целевой функции c и правых частей b задачи (P). Векторы b и c определялись по формулам

$$b = Ax_*, \qquad c = A^\top u_* + \xi;$$

если $x^{*i} > 0$, то $\xi^i = 0$, если $x^{*i} = 0$, то компонента ξ^i выбиралась случайным образом из интервала

$$0 \leq \gamma^i \leq \xi^i \leq \theta^i.$$

В приведенных ниже результатах расчетов полагалось, 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 (18), 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 итеративный процесс (32), (33) и обобщенный метод Ньютона, реализован в системе MATLAB 6.5. Для вычислений использовался компьютер с процессором Pentium-4, тактовой частотой 2.6 ГГц, оперативной памятью 1 Гб. Результаты вычислений случайным образом сгенерированных задач ЛП приведены в таблице, где указаны размерности m, n и ρ — плотность ненулевых элементов матрицы A, T — время решения задачи ЛП в секундах, количество итераций It, сделанных методом Ньютона при первом решении задачи максимизации (32).

В четвертом, пятом и шестом столбцах приведены точности выполнения ограничений прямой, 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 из-за нехватки оперативной памяти матрица A разбивалась на B блоков и умножение производилось поблочно. Соответствующее количество блоков B приведено в последнем столбце таблицы.

Таблица

$m \times n \times p$	T, c	It	$\\|A\hat{x}_* - b\\|$	$\\|(A^\top u_* - c)_+\\|$	$\|c^\top \hat{x}_* - b^\top u_*\|$	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
$500 \times (3 \times 10^6) \times 0.01$	179.1	12	3.2×10^{-10}	1.4×10^{-12}	1.9×10^{-11}	8
$1000 \times (3 \times 10^6) \times 0.01$	309.1	11	1.2×10^{-9}	4.1×10^{-12}	4.9×10^{-9}	10
$500 \times (5 \times 10^6) \times 0.01$	300.8	12	3.8×10^{-10}	1.6×10^{-12}	8.4×10^{-11}	10
$1000 \times (5 \times 10^6) \times 0.01$	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

Начальная точка в итеративном процессе (32), (33) во всех примерах бралась равной нулю: $x_0 = 0_n$. Всюду полагалось $\beta = 1$, $\mathrm{tol} = 10^{-12}$. Оказалось, что во всех примерах имело место $\beta > \beta_*$. Таким образом, нормальное решение $\hat{x}_*$ прямой задачи (P) было получено из процесса (32), (33) за одну итерацию, т.е. при $\omega = 1$. Число итераций обобщенного метода Ньютона (36) 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 (36) требовалось всего лишь две итерации.

В таблице представлены в основном результаты решения задач со слабо заполненной матрицей A. Исключение составляют три последние строки, где даны результаты решения задач с матрицами A, у которых случайным образом выбирались все элементы.

Приведенные в таблице результаты показывают высокую эффективность предложенного метода. Например, задача ЛП с пятью миллионами неотрицательных переменных и тысячью ограничений-равенств была решена с хорошей точностью менее чем за 7 мин (см. десятую строку таблицы). При этом для реализации метода на компьютере использовались только средства системы MATLAB.

Для задач ЛП, обладающих большим количеством неотрицательных переменных и средним количеством ограничений типа равенств, 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 равенств.

Авторы признательны С.С. Арсеньеву–Образцову, А.А. Станевичюсу и А.В. Тюленеву за полезные обсуждения и помощь в проведении численных расчетов.

СПИСОК ЛИТЕРАТУРЫ

1. *Еремин И.И.* Теория линейной оптимизации. Екатеринбург: УрО РАН, 1998.

2. *Поляк Б.Т.* Введение в оптимизацию. М.: Наука, 1983.

3. *Разумихин Б.С.* Физические модели и методы теории равновесия в программировании и экономике. М.: Наука, 1975.

4. *Mangasarian O.L.* A finite Newton method for classification // Optimizat. Meth. Software. 2002. V. 17. P. 913—-930.

5. *Mangasarian O.L.* A Newton method for linear programming: Data Mining Inst. Techn. Rept. 02-02. March 2002.

6. *Kanzow C., Qi H., Qi L.* On the minimum norm solution of linear program //J. Optimizat. Theory and Appl. 2003. V. 116. P. 333–345.

7. *Поляк Б.Т., Третьяков Н.В.* Об одном итерационном методе линейного программирования и его экономической интерпретации // Экономика и матем. методы. 1972. Т. 8. Вып. 5. С. 740–751.

8. *Антипин А.С.* Методы нелинейного программирования, основанные на прямой и двойственной модификации функции Лагранжа. М.: ВНИИСИ, 1979.

9. *Evtushenko Yu.G., Golikov A.I.* The augmented Lagrangian function for the linear programming problems // Dynamic. Non-homogeneous Systems. M.: Rus. Acad. Sci. Inst. System Analys. 1999. V. 2. P. 63–67.

10. *Голиков А.И., Евтушенко Ю.Г.* Отыскание нормальных решений в задачах линейного программирования // Ж. вычисл. матем. и матем. физ. 2000. Т. 40. № 12. С. 1766–1786.

11. *Еремин И.И.* О квадратичных задачах и полноквадратичных задачах выпуклого программирования // Изв. вузов. Математика. 1998. № 12. С. 22–28.

12. *Голиков А.И., Евтушенко Ю.Г.* Теоремы об альтернативах и их применение в численных методах // Ж. вычисл. матем. и матем. физ. 2003. Т. 43. № 3. С. 354–375.

13. *Фиакко А., Мак-Кормик Г.* Нелинейное программирование. Методы последовательной безусловной минимизации. М.: Мир, 1972.

14. *Тихонов А.Н.* Избранные труды. М.: МАКС Пресс, 2001.

15. *Васильев Ф.П., Иваницкий А.Ю.* Линейное программирование. М.: Факториал, 2003.