УДК 519.85

П.И. СТЕЦЮК, В.А. СТОВБА, И.С. МАРТЫНЮК

АЛГОРИТМЫ МЕТОДА ЭЛЛИПСОИДОВ ДЛЯ НАХОЖДЕНИЯ L_p-РЕШЕНИЯ СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ

Предложены два алгоритма метода эллипсоидов для нахождения L_p-решения системы линейных уравнений при двусторонних ограничениях на компоненты решения. Первый алгоритм использует метод Шора, а второй – метод Юдина – Немировского. Показано, что оба алгоритма требуют количества итераций, которое зависит только от числа неизвестных компонент в L_p-решении.

© П.И. Стецюк, В.А. Стовба, И.С. Мартынюк, 2017

Введение. Метод эллипсоидов независимо предложили Д.Б. Юдин и А.С. Немировский [1], исходя из методов последовательных отсечений, Н.З. Шор [2], исходя из субградиентных методов с растяжением пространства. Метод эллипсоидов является частным случаем субградиентных методов с растяжением пространства в направлении субградиента, которые предложены Н.З. Шором в 1969 году, т. е. за 7 лет до появления метода эллипсоидов. Замечательной чертой метода эллипсоидов есть то, что его скорость сходимости зависит только от размерности пространства переменных n и не зависит от свойств задачи. И хотя первоначально метод эллипсоидов был изобретен для минимизации выпуклых функций – он применим для более широкого класса задач, таких как задача выпуклого программирования, задача отыскания седловых точек выпукло-вогнутых функций, частные случаи задач решения вариационных неравенств, специальные классы задач линейной и нелинейной дополнительности.

В статье рассмотрим две алгоритмические реализации метода эллипсоидов для задачи выпуклого программирования, которая связана с нахождением решения системы линейных уравнений при двухсторонних ограничениях на переменные. Первая реализация использует метод эллипсоидов Шора, требующий коррекции несимметричной матрицы, а вторая реализация – метод эллипсоидов Юдина и Немировского, требующий коррекции симметричной матрицы.

1. Постановка задачи. Пусть имеется система линейных алгебраических уравнений:

$$Ax \approx b \tag{1}$$

при условиях

$$l \le x \le u. \tag{2}$$

Здесь $A - m \times n$-матрица, $b \in R^m$ – m-мерный вектор; $l \in R^n$, $u \in R^n$ – n-мерные векторы, такие, что для всех $i = 1, \ldots, n$, $u_i > l_i$; $x \in R^n$ – n-мерный вектор неизвестных параметров.

Требуется найти такой вектор $x_p^* \in R^n$, который является «наилучшим» решением системы (1) – (2) в так называемой L_p-норме, т. е. когда норма вектора невязок $y = Ax - b = (y_1, \ldots, y_m)^T$ определена следующим образом: $\|y\|_p = (\sum_{i=1}^{m} |y_i|^p)^{1/p}$, где $p \ge 1$. Случай $p = \infty$ определяется как $\|y\|_\infty = \max_{i=1,\ldots,m} |y_i|$. Случай $p = 2$ соответствует стандартной евклидовой норме $\|y\|$ для вектора невязок.

Нахождению «наилучшего» решения системы (1) – (2) поставим в соответствие следующую задачу выпуклого программирования:
найти

$$f_p^* = f_p(x_p^*) = \min_{x \in R^n} \left\{ f_p(x) = \|Ax - b\|_p \right\} \tag{3}$$

при ограничениях

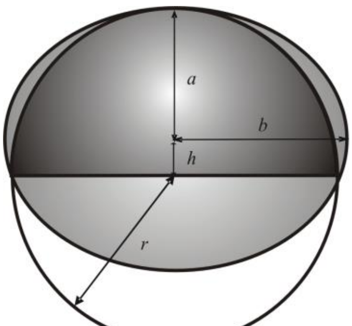
$$l \le x \le u, \tag{4}$$

где $p \in R$ – скалярный параметр, такой, что $p \ge 1$, который гарантирует выпуклость функции $f_p(x)$. Здесь x_p^* – решение задачи (2) – (3), для удобства будем считать, что оно единственное. Ограничения (4) называют двусторонними ограничениями на переменные.

Подобные задачи часто встречаются в самых разных областях прикладной математики, например, при обработке результатов наблюдений, построении и анализе различного рода моделей (физических, биологических, экономических, социальных и других), при поиске компромиссных решений в моделях с противоречивыми данными и т. д. Для линейной регрессии задачи (3) – (4) соответствует метод наименьших квадратов ($p = 2$), метод наименьших модулей ($p = 1$) и минимаксный (чебышевский) метод ($p = \infty$).

Задача (3)– (4) всегда имеет решение. Если $m > n$, то система линейных уравнений является переопределенной, и если для нее ранг матрицы A равен n, то задача (3) – (4) имеет единственное решение. В общем случае решение задачи (3) – (4) не обязательно будет единственным. В случае его неоднозначности будем находить x_p^* – одно из множества решений, которым соответствует оптимальное значение функции f_p^*.

2. Метод эллипсоидов и задача (3) – (4). Метод эллипсоидов базируется на использовании в R^n эллипсоида минимального объема, который содержит полушар, полученный в результате пересечения n-мерного шара и полупространства, проходящего через его центр. Этот эллипсоид имеет сплюснутую форму в направлении нормали к гиперплоскости, которая проходит через центр шара радиуса r. Его параметры (даны на рис. 1) следующие: a – длина меньшей полуоси в направлении нормали, определяющей полушар; b – длина большей полуоси (количество таких полуосей равно $n-1$); h – расстояние от центра шара до центра эллипсоида в направлении меньшей из его полуосей.



$$a = r\frac{n}{n+1},$$

$$b = r\frac{n}{\sqrt{n^2-1}},$$

$$h = r\frac{n}{n+1}.$$

РИСУНОК. Эллипсоид минимального объема, содержащий полушар в R^n

Итерация метода эллипсоидов состоит в переходе от текущего эллипсоида к следующему с постоянным коэффициентом уменьшения их объемов. Этот коэффициент определяется отношением объема эллипсоида с длинами полуосей a и b к объему шара радиуса r в R^n и может быть записан в виде

$$q_n = \left(\frac{a}{r}\right)\left(\frac{b}{r}\right)^{n-1} = \frac{n}{n+1}\left(\frac{n}{\sqrt{n^2-1}}\right)^{n-1} < 1. \tag{5}$$

В работе [3] показано, что

$$q_n < \exp\left\{-\frac{1}{2n}\right\} < 1, \tag{6}$$

следовательно, при больших n коэффициент уменьшения объема хорошо аппроксимируется асимптотической формулой

$$q_n \approx 1 - \frac{1}{2n}. \tag{7}$$

Поэтому, для уменьшения объема эллипсоида, локализующего решение задачи, в 10 раз, требуется сделать K итераций, где

$$K = -\frac{\ln 10}{\ln q_n} \approx (2\ln 10)n \approx 4.6n. \tag{8}$$

Чтобы для решения задачи (3) – (4) применить метод эллипсоидов требуется: во-первых – определить градиентное поле $g(x)$ (способ построения в точке $x \in R^n$ гиперплоскости, локализирующей точку x_p^* в одном из полупространств пространства R^n); во-вторых – выбрать начальный радиус области локализации оптимального решения x_p^*. Удовлетворить эти требования для задачи (3) – (4) не представляет особых проблем. Так, первую часть этих требований можно удовлетворить, используя лемму.

Лемма 1 [4]. Пусть $\partial f_p(x)$ – субградиент функции $f_p(x)$ в точке x; $t^* = \max\{t_{i^*}, t_{j^*}\}$, $t_{i^*} = \max\limits_{i=1,\dots,n}\{x_i - u_i\}$, $t_{j^*} = \max\limits_{j=1,\dots,n}\{l_j - x_j\}$; i^*, j^* – значения i, j $(1 \le i, j \le n)$, на которых достигаются t_{i^*}, t_{j^*}; e_k – k-й орт в E^n, $1 \le k \le n$. Тогда вектор

$$g_p(x) = \begin{cases} \partial f_p(x), & \text{если} \quad t^* \le 0, \\ e_{i^*}, & \text{если} \quad t^* > 0 \quad \text{и} \quad t^* > t_{i^*}, \\ -e_{j^*}, & \text{если} \quad t^* > 0 \quad \text{и} \quad t^* \le t_{j^*}, \end{cases}$$

удовлетворяет свойству

$$(g_p(x), x - x_p^*) \ge 0 \quad \text{для всех} \quad x \in R^n. \tag{9}$$

Лемма 1 означает следующее. Если точка x находится внутри допустимой области, заданной ограничениями (4), то в качестве $g_p(x)$ выбирается $\partial f_p(x)$ – субградиент функции $f_p(x)$ в этой точке, который вычисляется по формуле:

$$\partial f_p(x) = \|Ax - b\|_p^{1-p} \sum_{j=1}^{m} \left(\mathrm{sgn}(a_j x - b_j) \cdot \left| a_j x - b_j \right|^{p-1} a_j^T \right),$$

где a_j – вектор-строка матрицы A с номером j, $j = 1, \dots, m$. Если же точка находится вне допустимой области, то выбирается субградиент к максимально нарушенному ограничению вида (4). Выпуклость функции $f_p(x)$ и ограничений (4) для векторного поля $g_p(x)$ гарантирует выполнение свойства (9).

Априорную информацию о локализации точки x_p^* в шаре легко обеспечить, если за центр шара выбрать центр параллелепипеда, заданного двусторонними ограничениями на переменные (4), и установить радиус шара таким, чтобы он содержал параллелепипед и имел минимальный объем. Это обеспечивает лемма.

Лемма 2 [4]. Если $x_0 = \frac{1}{2}(u + l)$ и $r_0 = \frac{1}{2}\|u - l\|$, то параллелепипед $P(x) = \{x : l \le x \le u\}$ содержится в n-мерном шаре $S(x_0, r_0) = \{x : \|x - x_0\| \le r_0\}$.

3. Алгоритм Шора для нахождения x_p^*. В соответствии с правилом вычисления $g_p(x)$ леммы 1 построим формулу для вычисления «обобщенного» значения функции в задаче (3) – (4):

$$F_p(x) = \begin{cases} +\infty, & \text{если } t^* > 0; \\ f_p(x), & \text{если } t^* \le 0. \end{cases}$$

Значение $F_p(x)$ будем использовать при построении критерия останова в алгоритме нахождения x_p^*. Входными параметрами алгоритма будут величина p ($p \ge 1$), с помощью которой определена L_p-норма в (3), и величина ε_f – точность, с которой требуется найти значение $f_p^* = f_p(x_p^*)$.

Учитывая вышеизложенное, алгоритм Шора для нахождения x_p^* примет следующий вид.

Инициализация. Положим стартовую точку $x_0 = (u+l)/2$ и начальный радиус $r_0 = \|u-l\|/2$. Введем в рассмотрение $n \times n$-матрицу B и положим $B_0 := I_n$, где I_n – единичная $n \times n$-матрица. Перейдем к первой итерации со значениями x_0, r_0 и B_0.

Пусть на k-й итерации найдены значения $x_k \in E^n$, r_k, B_k. Переход к $(k+1)$-й итерации состоит в выполнении такой последовательности действий.

Шаг 1. Вычислим $F_p(x_k)$. Если $F_p(x_k) = 0$, то "Останов: $k^* = k$ и $x_p^* = x_k$". Иначе вычислим $g_p(x_k)$. Если $F_p(x_k) < +\infty$ и $\|B_k^T g_p(x_k)\| r_k \le \varepsilon_f$, то "Останов: $k^* = k$ и $x_p^* = x_k$". Иначе переходим к шагу 2.

Шаг 2. Положим $\xi_k := \dfrac{B_k^T g_p(x_k)}{\|B_k^T g_p(x_k)\|}$.

Шаг 3. Вычислим очередную точку $x_{k+1} := x_k - h_k B_k \xi_k$, где $h_k = \dfrac{1}{n+1} r_k$.

Шаг 4. Вычислим $B_{k+1} := B_k + \left(\sqrt{\dfrac{n-1}{n+1}} - 1\right)(B_k \xi_k)\xi_k^T$ и $r_{k+1} := r_k \dfrac{n}{\sqrt{n^2-1}}$.

Шаг 5. Переходим к $(k+1)$-й итерации со значениями x_{k+1}, r_{k+1}, B_{k+1}.

Алгоритм базируется на методе эллипсоидов, разработанном Н.З. Шором в [2]. Его сходимость обеспечивает теорема.

Теорема 1 [5]. Последовательность точек $\{x_k\}_{k=0}^{k^*}$ удовлетворяет неравенству

$$\|B_k^{-1}(x_k - x_p^*)\| \le r_k, \quad k = 0,1,2,\ldots,k^*.$$

На каждой итерации $k>0$ величина уменьшения объема эллипсоида $E_k = \left\{ x \in R^n : \left\| B_k^{-1}(x_k - x) \right\| \le r_k \right\}$, локализующего x_p^*, есть величина постоянна и равна

$$q = \frac{vol(E_k)}{vol(E_{k-1})} = \frac{n}{n+1}\left(\frac{n}{\sqrt{n^2-1}}\right)^{n-1} < \exp\left\{-\frac{1}{2n}\right\} < 1.$$

Из теоремы следует, что для уменьшения в 10 раз объема эллипсоида, локализующего x_p^*, требуется $K = 4.6n$ итераций (см. формулы (5) – (8)). Для задачи (3) – (4) это означает следующее: чтобы на порядок улучшить отклонение найденного рекордного значения функции $f_p(x)$ от ее оптимального значения f_p^*, потребуется сделать $4.6n^2$ итераций. Если в задаче переменных не более двадцати, то для нахождения f_p^* с относительной точностью 10^{-10} максимальное количество итераций легко определить из таблицы. Здесь приведены необходимые количества итераций алгоритма для $n = 2 \div 19$ и относительной точности $\varepsilon = 10^{-10}$, где $\varepsilon = (f_p(x_k^*) - f_p^*)/(f_p(x_0) - f_p^*)$.

ТАБЛИЦА. Количество итераций алгоритма (относительная точность 10^{-10})

n	itn	n	itn	n	itn
2	177	8	2940	14	9019
3	407	9	3723	15	10355
4	730	10	4598	16	11782
5	1144	11	5565	17	13302
6	1651	12	6624	18	14914
7	2249	13	7776	19	16617

3. Алгоритм Юдина – Немировского для нахождения x_p^*. Входными параметрами алгоритма будут величины p ($p \ge 1$) и ε_f – точность, с которой требуется найти $f_p(x_p^*) = f_p^*$. Алгоритм Юдина – Немировского для нахождения x_p^* имеет следующий вид.

Инициализация. Положим стартовую точку $x_0 = (u+l)/2$ и начальный радиус $r_0 = \|u-l\|/2$. Введем в рассмотрение симметричную $n \times n$-матрицу H и положим $H_0 := I_n$, где I_n – единичная $n \times n$-матрица. Перейдем к первой итерации со значениями x_0, r_0 и H_0.

Пусть на k-й итерации найдены значения $x_k \in E^n$, r_k, H_k. Переход к $(k+1)$-й итерации состоит в выполнении такой последовательности действий.

Шаг 1. Вычислим $F_p(x_k)$. Если $F_p(x_k)=0$, то "Останов: $k^*=k$ и $x_p^*=x_k$ ". Иначе вычислим $g_p(x_k)$. Если $F_p(x_k)<+\infty$ и $r_k\sqrt{g_p^T(x_k)H_k g_p(x_k)}\le\varepsilon_f$, то "Останов: $k^*=k$ и $x_p^*=x_k$ ". Иначе переходим к шагу 2.

Шаг 2. Вычислим очередную точку

$$x_{k+1}:=x_k-h_k\frac{H_k g_p(x_k)}{\sqrt{g_p^T(x_k)H_k g_p(x_k)}},\quad \text{где}\quad h_k=\frac{1}{n+1}r_k.$$

Шаг 3. Вычислим

$$H_{k+1}:=H_k-\frac{2}{n+1}\frac{H_k g_p(x_k)g_p^T(x_k)H_k}{g_p^T(x_k)H_k g_p(x_k)}\quad \text{и}\quad r_{k+1}:=r_k\frac{n}{\sqrt{n^2-1}}.$$

Шаг 4. Переходим к $(k+1)$-й итерации со значениями x_{k+1}, r_{k+1}, H_{k+1}.

Приведенный алгоритм использует метод эллипсоидов, разработанный Д.Б. Юдиным и А.С. Немировским в [1]. Он получен как вариант методов последовательных отсечений и работает с симметричной $n\times n$-матрицей $H=BB^T$, где B – $n\times n$-матрица в алгоритме Шора. Сходимость алгоритма Юдина – Немировского обеспечивает следующая теорема.

Теорема 2. Последовательность точек $\{x_k\}_{k=0}^{k^*}$ удовлетворяет неравенству

$$\left(x_k-x_p^*\right)^T H_k^{-1}\left(x_k-x_p^*\right)\le r_k^2,\quad k=0,1,2,\dots,k^*.$$

На каждой итерации $k>0$ величина уменьшения объема эллипсоида $\mathrm{E}_k=\left\{x\in R^n:\left(x_k-x\right)^T H_k^{-1}\left(x_k-x\right)\le r_k^2\right\}$, локализующего x_p^*, есть величина постоянна и равна

$$q=\frac{vol(\mathrm{E}_k)}{vol(\mathrm{E}_{k-1})}=\frac{n}{n+1}\left(\frac{n}{\sqrt{n^2-1}}\right)^{n-1}<\exp\left\{-\frac{1}{2n}\right\}<1.$$

Выводы. Предложенные алгоритмы можно успешно применять для нахождения x_p^*, если количество переменных небольшое. Алгоритмы будут устойчивыми для решения плохо-обусловленных систем линейных уравнений. Чтобы найти точку минимума выпуклой функции с относительной точностью по значению функции, равной 10^{-10}, методу эллипсоидов при $n=10$ достаточно осуществить 4600 итераций. Для современных персональных компьютеров это требует не меньше секунды процессорного времени.

Величина m не влияет на скорость сходимости алгоритмов, от нее зависит трудоемкость вычисления значения функции $f_p(x)$ и ее субградиента $\partial f_p(x)$. При m: 1000 это будет вносить в трудоемкость обоих алгоритмов более весомый вклад, чем алгоритмические операции – (шаги 2 – 4) в алгоритме Шора и (шаги 2 – 3) в алгоритме Юдина – Немировского.

Работа выполнена при поддержке НАН Украины, проекты № 0117U000327 и № 0116U004558.

П.І. Стецюк, В.О. Стовба, І.С. Мартинюк

АЛГОРИТМИ МЕТОДУ ЕЛІПСОЇДІВ ДЛЯ ЗНАХОДЖЕННЯ L_p-РОЗВ'ЯЗКУ СИСТЕМИ ЛІНІЙНИХ РІВНЯНЬ

Запропоновано два алгоритми методу еліпсоїдів для знаходження L_p-розв'язку системи лінійних рівнянь з двосторонніми обмеженнями на компоненти розв'язку. У першому алгоритмі використовується метод Шора, в другому – метод Юдіна – Немировського. Показано, що кількість ітерацій, яку потребують обидва алгоритми, залежить лише від кількості невідомих компонент у L_p-розв'язку.

P.I. Stetsyuk, V.O. Stovba, I.S. Martynyuk

ALGORITHMS OF ELLIPSOID METHOD FOR FINDING L_p-SOLUTION OF LINEAR EQUATIONS SYSTEM

We propose two algorithms of ellipsoid method to find L_p-solution of linear equations system with two-sided constraints on solution components. The first and the second algorithms use Shor's and Yudin-Nemirovskii methods accordingly. It is shown, that number of iterations required by each algorithm depends merely on the number of unknown components in L_p-solution.

1. *Юдин Д.Б., Немировский А.С.* Информационная сложность и эффективные методы решения выпуклых экстремальных задач. *Экономика и мат. методы*. 1976. Вып. 2. С. 357 –359.
2. *Шор Н.З.* Метод отсечения с растяжением пространства для решения задач выпуклого программирования. *Кибернетика*. 1977. № 1. С. 94 – 95.
3. *Grőtschel M., Lovàsz L., Schrijver A*. Geometric Algorithms and Combinatorial Optimization. Berlin: Springer-Verlag, 1988. 362 p.
4. *Стецюк П.И., Колесник Ю.С., Березовский О.А*. Об одном методе нахождения L_p-решения системы линейных уравнений. *Теория оптимальных решений*. Киев: Ин-т кибернетики имени В.М. Глушкова НАН Украины, 2003. С. 83 – 90.
5. *Стецюк П.И., Била Г.Д., Стовба В.А.* Метод эллипсоидов для нахождения L_p-решения системы линейных уравнений. Інформатика та системні науки (ІСН-2017): матеріали VIII Всеукр. наук.-практ. конф. за міжнародною участю (м. Полтава, 16 – 18 березня 2017 року) /за ред. Ємця О.О. – Полтава: ПУЕТ, 2017.

Получено 27.03.2017