Изв. АН Эстонии. Физ. Матем., 1991, 40, № 1, 17—24
https://doi.org/10.3176/phys.math.1991.1.03

УДК 519.8

Эвальд ЮБИ *

НАЧАЛЬНЫЙ БАЗИС МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ. ПРЯМОЙ И ДВОЙСТВЕННЫЙ СИМПЛЕКС-МЕТОД

(Представил Г. Вайникко)

Для определения начального базиса в задаче линейного программирования дважды используется ранее предложенный метод нахождения неотрицательного решения недоопределенной системы линейных уравнений. На первом этапе целевая функция (c, x) приравнивается ожидаемому значению минимума, на втором этапе ее значения не ограничены.

Определяемый таким образом начальный базис зависит и от коэффициентов целевой функции и нередко оказывается оптимальным.

1. Введение

Пусть заданы вещественная $(m\times n)$-матрица A, m-вектор b и n-вектор c. Нужно найти n-вектор x, решение задачи

$$\begin{gathered}(c, x)=z\to\min,\\ Ax=b,\\ x\geqslant 0.\end{gathered}\tag{1}$$

Обозначим через x^* оптимальное решение этой задачи, $z^*=(c, x^*)$.

При решении задачи (1) методом симплексного типа нередко наиболее трудоемким является нахождение начального допустимого базиса. Предлагаемый алгоритм может дать начальное решение, близкое к оптимальному, если минимум z^* незначительно отличается от нуля. Если это предположение не выполняется, составляется эквивалентная задача

$$\begin{gathered}(c, x)-M=z\to\min,\\ Ax=b,\\ x\geqslant 0,\end{gathered}\tag{2}$$

где M задается перед решением задачи как ожидаемое значение минимума z^*. (В случае отсутствия какой-либо информации о величине z^* следует положить $M=0$.) Отметим, что целевую функцию можно записать и в виде $(c, x)-x_{n+1}=z$, если использовать дополнительное ограничение $x_{n+1}=M$.

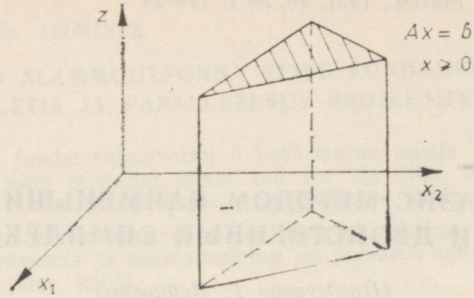
Рассмотрим еще вспомогательную задачу наименьших квадратов (НК)

$$\varphi(x, z)=\frac{1}{2}\|Ax-b\|^2+[(c, x)-z-M]^2\to\min,$$

где $\varphi(x, z)$ минимизируется по всем n-мерным $x\geqslant 0$ и z.

На каждом шаге предлагаемого алгоритма VRMA выполняется такое ортогональное преобразование для системы (2), при котором функция $\varphi(x, z)$ убывает. На первом этапе алгоритма VRMA минимизируется $\varphi(x, z)$ по $x\geqslant 0$ при $z=0$ (т. е. в системе (2) $(c, x)=M$), на втором этапе по $x\geqslant 0$ при произвольном z.

* Tallinna Tehnikaülikool (Таллиннский технический университет). 200026 Tallinn, Ehitajate tee 5. Estonia.



Для решения задачи (2) расширенная система

$$\begin{aligned} Ax&=b, \\ (c,x)-z&=M \end{aligned} \tag{3}$$

приводится ортогональными преобразованиями к треугольному виду. В отличие от обычного QR-разложения расширенной матрицы системы (3) в этом методе порядок выбора столбцов зависит от правых частей уравнений. На k-м шаге для столбцов системы выполняется такое преобразование Хаусхолдера, которое аннулирует элементы ведущего столбца начиная с $k+1$-го. Согласно критерию (4) ведущим выбирается столбец, образующий минимальный угол с невязкой расширенной системы. На каждом шаге множество IJ индексов активных переменных меняется на один элемент.

При добавлении элемента очередная активная переменная $x_{j_{k+1}}$ определяется по решению задачи

$$\frac{\partial\varphi(x^k,z^k)}{\partial x_j}:\|a_j\|\to\min_{j=1,\dots,n}, \tag{4}$$

где $x^0=0$, $z^0=0$, (x^k,z^k) — текущая точка, определенная методом НК из переопределенной системы с k переменными и $m+1$ уравнениями, см. [1]. Для нахождения очередной точки (x^{k+1},z^{k+1}) к подматрице матрицы A, составленной из активных столбцов и приведенной на предыдущих шагах к треугольному виду, присоединяется очередной активный столбец $a_{j_{k+1}}$. Затем для всех столбцов и правой части системы выполняется такое преобразование Хаусхолдера, которое аннулирует в столбце $a_{j_{k+1}}$ нужное количество элементов для получения треугольной системы, из которой можно определить $(x^{k+1},\ z^{k+1})$.

При замене элемента j множества IJ в случае $x_j^h<0$ очередной шаг заключается в исключении x_j и активизации новой переменной. Сначала преобразуется система активных столбцов к треугольному виду. Для этого нужно провести серию вращений Гивенса, аннулировать в каждом активном столбце после j-го по одному элементу ниже главной диагонали. Затем вышеизложенным образом добавляется очередная активная переменная. В разделе 3 при построении начального базиса для двойственного симплекс-метода условие неотрицательности переменных не проверяется. Когда в невырожденном случае число активных переменных в системе (3) равняется числу уравнений $m+1$, начальный базис определен.

В двух следующих разделах рассматривается определение начального базиса в невырожденном случае. Последний раздел посвящен задачам с вырожденными базисами.

Конечность алгоритма следует из [1].

2. Прямой симплекс-метод

Предположим, что все допустимые базисы невырождены. Рассмотрим определение начального базиса на основе предложенного в [1] подхода, использующего третий метод из [2] (гл. 24). В этом методе все столбцы расширенной матрицы преобразуются на каждом шаге. Кроме A и b требуются еще три вспомогательных n-вектора. Ортогональные матрицы преобразования системы не запоминаются.

Основная часть алгоритма совпадает с алгоритмом VRM1 из работы [1]. Не вводя дополнительных обозначений, считаем в системе (3) A $(m+1)\times(n+1)$-матрицей и b $(m+1)$-вектором. Опишем алгоритм определения начального базиса.

Алгоритм VRMA $(A, b, IJ, x, z, F, G, m, n, \varepsilon_1, \varepsilon_2, ET)$.

1. Присоединить к матрице коэффициентов A в качестве последней $m+1$-й строки вектор c и дополнительный столбец (см. систему (3)).
2. Положить число активных переменных $k:=0$.
3. Найти n-векторы F и G с координатами

$$F_j=(a_j, b)=\sum_{i=k+1}^{m+1} a_{ij}b_i, \quad G_j=\|a_j\|=\sqrt{\sum_{i=k+1}^{m+1} a_{ij}^2}, \quad j=1, \ldots, n. \qquad (5)$$

4. Определить очередную активную переменную $x_{j_{k+1}}$ по решению задачи

$$\max_{j=1,\cdots,n} \frac{F_j}{G_j}=\max_{j=1,\cdots,n} \frac{\sum_{i=k+1}^{m+1} a_{ij}b_i}{\sqrt{\sum_{i=k+1}^{m+1} a_{ij}^2}}=\mathrm{RE}. \qquad (6)$$

5. Если $\mathrm{RE}<\varepsilon_1$, перейти к шагу 6, в противном случае к 9.
6. Проверить неравенство $|b_i|\leqslant\varepsilon_2$ для $i=k+1, \ldots, m+1$. Если все они выполняются, то начальный базис вырожденный и в таком случае перейти к шагу 24, в противном случае к 7.
7. Если $z=0$, то активизируется z; перейти к шагу 9.
8. Задача (1) несовместна. Стоп.
9. Увеличить число активных переменных на единицу, $k:=k+1$.
10. Проверить равенство $k=m+1$. Если оно выполняется, перейти к шагу 12.
11. Выполнить преобразование Хаусхолдера с направляющим столбцом, определенным на 4-м или 7-м шаге.
12. Вычислить новые F_j и G_j по формулам

$$\tilde{F}_j=F_j-\tilde{a}_{kj}\tilde{b}_k, \quad \tilde{G}_j^2=G_j^2-(\tilde{a}_{kj})^2, \quad j=1, \ldots, n. \qquad (7)$$

13. Проверить неравенство $k<m\times ET+1$. Если оно выполняется, перейти к шагу 4.
14. Решить треугольную систему для определения активных переменных.
15. Если $ET=1$, перейти к шагу 24.
16. Проверить неравенства $x_j\geqslant0$, $j\in IJ$. Если все неравенства выполняются, перейти к шагу 23.
17. Удалить из актива отрицательную переменную x_j, которая активизировалась позже остальных отрицательных переменных.
18. Привести матрицу из активных столбцов вращениями Гивенса к треугольному виду (после удаления из актива j-го столбца).

19. Найти новые F_j и G_j по формулам

$$\tilde{F}_j=F_j+\tilde{a}_{kj}\tilde{b}_k, \quad \tilde{G}^2_j=G^2_j+(\tilde{a}_{kj})^2, \quad j=1, \ldots, n. \tag{8}$$

20. Положить число активных переменных $k:=k-1$.
21. Перейти к шагу 4.
22. Проверить неравенство $k<m+1$. Если оно не выполняется, перейти к шагу 24.
23. Проверить неравенства $|b_i|\leqslant\varepsilon_2$ для $i=k+1, \ldots m+1$. Если все они выполняются, то начальный базис вырожденный, перейти к шагу 24, в противном случае к 4.
24. Привести треугольную систему, соответствующую активным переменным, к диагональному виду.
25. Начальное базисное решение найдено.

Замечания: 1. До величины $k=\mathrm{m}\times ET+1$ условие $xj\geqslant 0$ не проверяется (шаг 13), так как при малых k это условие почти всегда выполняется. При этом $ET\leqslant 1$ задается при обращении к подпрограмме.

2. Критерий оптимальности в задаче (2) (при фиксированном z)

$$\max_{j=1,\cdots,n} F_j=0. \tag{9}$$

3. Для определения $m+1$-й активной переменной иногда нежелательно пользоваться формулой (6), так как величины F_j и G_j вычисляются приближенно и многие из них близки к нулю. Тогда при $k=m$ вместо формулы (6) можно воспользоваться формулой

$$F_j/G_j=\tilde{b}_{m+1}\,\mathrm{sign}\,(\tilde{a}_{m+1j}).$$

Для всех активных переменных $F_j=G_j=0$, для неактивных переменных $F_j/G_j=\pm\tilde{b}_{m+1}$ или $F_j/G_j=0$. Так как при активизации $m+1$-й переменной преобразование Хаусхолдера не выполняется, то в случае $x_j<0$ проще не исключать x_j из актива, а перебирать в качестве $m+1$-й все переменные, для которых $F_j/G_j=|\tilde{b}_{m+1}|$ пока не получено неотрицательное решение. Если ни для одной переменной такого решения нет, придется исключить $x_j<0$ (шаг 17).

4. Если k_0 столбцов системы (3) единичны или отличаются от них на постоянный множитель и соответствующие переменные неотрицательны, то можно их сразу включить в актив и начать VRMA не с $k=0$, а с $k=k_0$. Только предварительно надо строки переставить так, чтобы сначала были активные, затем пассивные строки.

Пример 1.

$$\begin{array}{rcrcrcrcrcl} x_1 & + & x_2 & + & x_3 & - & x_4 & - & 2x_5 & & =1{,}50 \\ & & x_2 & - & x_3 & & & & & & =0 \\ & & & & x_3 & & & - & x_5 & & =1 \\ & & x_2 & + & x_3 & + & x_4 & & & - z & =0. \end{array} \tag{10}$$

Находим $\dfrac{F_j}{G_j}=\dfrac{1{,}50}{1}, \dfrac{1{,}50}{\sqrt{3}}\ \dfrac{2{,}50}{2}\ \dfrac{-1{,}50}{\sqrt{2}}\ \dfrac{-4}{\sqrt{5}}$

Здесь F_j и G_j вычисляются по формуле (5). Последнее уравнение в этой системе соответствует целевой функции $x_2+x_3+x_4=z\to\min$, $M=0$.

Согласно критерию (6) активизируется x_1. Выполним преобразование Хаусхолдера для столбцов и правой части системы (10) при $v==(1, 0, 0, 0)^{\mathrm{T}}$, см [2].

Шаг	Строка	x_1	x_2	x_3	x_4	x_5	z	b
II	1	−1	−1	1	1	2	0	−1,50
	2	0	1	−1	0	0	0	0
	3	0	0	1	0	−1	0	1
	4	0	1	1	1	0	−1	0
	$\frac{F_j}{G_j}$	$\frac{0}{0}$	$\frac{0}{\sqrt{2}}$	$\frac{1}{\sqrt{3}}$	$\frac{0}{1}$	$\frac{-1}{1}$		
III	2	0	0	1,73	0,58	−0,58	−0,58	0,58
	3	0	0,37	0	−0,21	−0,79	0,21	0,79
	4	0	1,37	0	0,79	0,21	−0,79	−0,21
	$\frac{F_j}{G_j}$	$\frac{0}{0}$	$\frac{0}{\sqrt{2}}$	$\frac{0}{0}$	$\frac{-0{,}33}{\sqrt{0{,}66}}$	$\frac{-0{,}66}{\sqrt{0{,}66}}$		
IV	3	0	1,22	0	0,82	0,41	−0,82	−0,41
	3	0	0,71	0	0	−0,71	0	0,71
	4 $\frac{F_j}{G_j}$	$\frac{0}{0}$	$\frac{0{,}50}{\sqrt{0{,}50}}$	$\frac{0}{0}$	$\frac{0}{0}$	$\frac{-0{,}50}{\sqrt{0{,}50}}$		
V	1	0,50	0,50	2	0	−1,50	−0,50	1,25
	2	0,29	−0,87	0	1,15	−0,87	0,87	0,72
	3	0,71	1,41	0	−0,71	−1,41	0	1,06
	4	−0,41	0	0	0,41	0	0	0,20
	$\frac{F_j}{G_j}$	$\frac{-0{,}08}{0{,}41}$	$\frac{0}{0}$	$\frac{0}{0}$	$\frac{0{,}08}{0{,}41}$	$\frac{0}{0}$		
VI	1	0	0	1	0	−1	0	1
	2	−1	0	0	0	−2	1	2,5
	3	0	1	0	0	−1	0	1
	4	−1	0	0	1	0	0	0,5

После первого шага из треугольной системы $-x_1=-1{,}50$ (строка 1 шага II) определим $x_1^1=1{,}50>0$. На k-м шаге строки 1, ..., $k-1$ не преобразуются и поэтому в таблице они не приводятся. Шаг II: активизируется x_3, $x_1^2=1{,}16$, $x_3^2=0{,}33$. Шаг III: выполняется условие оптимальности (9) при $z=0$. Теперь согласно седьмому пункту алгоритма VRMA активизируется z. Шаг IV: активизируется x_2, решение треугольной системы

$$\begin{array}{rl}
-x_1-x_2-x_3 & =-1{,}50 \\
1{,}73x_3-0{,}58z & =\ 0{,}58 \\
1{,}22x_2\qquad -0{,}82z & =-0{,}41 \\
0{,}71x_2 & =\ 0{,}71,
\end{array}$$

$x_1^4=-0{,}50$, $x_2^4=1$, $x_3^4=1$, $z^4=2$ не удовлетворяет условию неотрицательности. Для исключения x_1 из актива нужно провести серию вращений Гивенса, чтобы оставшиеся активные столбцы приняли треугольную форму. Нужно аннулировать последовательно перед x_3, z и x_2 соответственно элементы второй, третьей и четвертой строки (шаг V).

Теперь активизируется x_4. Решение треугольной систем $x^5_2=x^5_3=1$, $x^5_4=0{,}5$, $z^5=2{,}5$ неотрицательно. Осталось исключить активные переменные из остальных уравнений, т. е. преобразовать систему из треугольного вида к диагональному (шаг VI). Целевой функции z соответствует вторая строка последней таблицы. Найденное начальное решение x^5 оказывается оптимальным.

5. Если переменная z в ходе выполнения алгоритма VRMA не активизировалась, то перед применением симплекс-метода надо ее ввести в базис вместо какой-либо переменной. Выводимой переменной соответствует $\max b_i/a_{in+1}$, который находится по всем $a_{in+2}<0$, $i=1, \ldots, m+1$. Если все $a_{in+2}\geqslant 0$, целевая функция задачи (1) не ограничена.

6. Если имеется подозрение, что целевая функция в задаче (1) не ограничена, нет необходимости применять симплекс-метод. M следует приравнять значению, меньшему возможного значения целевой функции, см. [3]. Если при таком M задача (2) имеет допустимое решение, целевая функция задачи (1) не ограничена.

3. Двойственный симплекс-метод

Рассмотрим нахождение начального базиса для двойственного симплекс-метода. Для этого используются метод НК и прямой симплекс-метод. Такой метод эффективен в случае, если в исходной задаче много переменных, неограниченных по знаку.

По-прежнему предполагаем, что все базисы невырождены.

В этом алгоритме на первом этапе применения метода НК находится некоторое базисное решение задачи (1), не удовлетворяющее, возможно, условию $x\geqslant 0$. В алгоритме VRMA предыдущего раздела условие неотрицательности не проверяется, полагается $ET=1$, после 13-го шага всегда переходят к 4-у шагу.

На втором этапе нужно добиться неположительности оценок всех переменных, $z_j-c_j\leqslant 0$, $j=1, \ldots, n$. Для этого используется прямой симплекс-метод. Если это условие выполняется, мы имеем двойственно-допустимое базисное решение, см. [4].

На третьем этапе находится оптимальное решение задачи с помощью двойственного симплекс-метода.

Рассмотрим решение примера 1 предыдущего раздела. Алгоритм VRMA при $ET=1$ определяет актив x_1, x_3, z, x_2 и система (10) принимает вид

$$\begin{aligned}
-x_1 - x_2 - x_3 \quad + x_4 \quad + 2x_5 \quad &= -1{,}50\\
1{,}73x_3 + 0{,}58x_4 - 0{,}58x_5 - 0{,}58z &= 0{,}58\\
1{,}22x_2 \quad + 0{,}82x_4 + 0{,}41x_5 - 0{,}82z &= -0{,}41\\
0{,}71x_2 \quad - 0{,}71x_5 \quad &= 0{,}71.
\end{aligned}$$

Для окончания первого этапа нужно активные столбцы преобразовать из треугольного вида к диагональному, исключить переменные в порядке x_2, z, x_3, x_1

$$\begin{array}{llr}
1 & x_1 - x_4 & = -0{,}5\\
2 & x_3 - x_5 & = 1\\
3 & -x_4 - 2x_5 + z & = 2\\
4 & x_2 - x_5 & = 1.
\end{array}$$

Второй этап применения прямого симплекс-метода пропускается, так как оценки всех небазисных переменных отрицательны.

На третьем этапе по правилам двойственного симплекс-метода из базиса выводится x_1 и вводится x_4.

$$\begin{array}{rrrrrrl}
-x_1 & & & +x_4 & & & =0{,}5 \\
 & & x_3 & & -x_5 & & =1 \\
-x_1 & & & & -2x_5 & +z & =2{,}5 \\
 & x_2 & & & -x_5 & & =1.
\end{array}$$

Решение $x^*=(0, 1, 1, 1/2, 0)^{\mathrm{T}}$ оптимальное, $z^*=2{,}5$.

4. Выродженный базис

Если определенный методом НК базис вырожденный, число шагов метода меньше числа уравнений. Тогда недостающие базисные переменные можно определить либо непосредственным исключением, либо методом искусственного базиса или M-методом.

Пример 2. Запишем задачу в виде (3).

	x_1	x_2	x_3	x_4	x_5	z	b
1	−1	−1	2	−1	0	0	0
2	0	1	−3	0	0	0	0
3	0	−1	−1	−1	−1	0	−1
4	0	2	−1	1	0	−1	0

Величины $\dfrac{F_j}{G_j}=\dfrac{0}{1},\ \dfrac{1}{\sqrt{7}},\ \dfrac{1}{\sqrt{15}},\ \dfrac{1}{\sqrt{3}},\ \dfrac{1}{1}.$

Последняя строка соответствует целевой функции $2x_2-x_3+x_4= =z\to\min$. На первом шаге активизируется x_5. Выполним для столбцов системы преобразование Хаусхолдера при $V=(0, 0, -1, 0)^{\mathrm{T}}$.

	x_1	x_2	x_3	x_4	x_5	z	b
1	0	−1	−1	−1	−1	0	−1
2	0	1	−3	0	0	0	0
3	−1	−1	2	−1	0	0	0
4	0	2	−1	1	0	−1	0

Находим $\dfrac{F_j}{G_j}=\dfrac{0}{1},\ \dfrac{0}{\sqrt{6}},\ \dfrac{0}{\sqrt{14}},\ \dfrac{0}{\sqrt{2}},\ \dfrac{0}{0}.$

Так как все $F_j=0$, начальное вырожденное базисное решение найдено, $x_5=1$, остальные $x_j=0$. Третьей строке можно сопоставить базисную переменную $x_1=0$, четвертой $z=0$. При дальнейшем исключении допустимы и отрицательные направляющие элементы, так как в этих строках $b_i=0$. Для большей надежности лучше выбрать направляющим элемент с максимальным модулем. Поэтому по второй строке исключим x_3. Решение $x=(0, 0, 0, 0, 1)^{\mathrm{T}}$ оптимальное, базис составляет переменные x_5, x_3, x_1, z.

	x_1	x_2	x_3	x_4	x_5	z	b
1	0	4/3	0	1	1	0	1
2	0	−1/3	1	0	0	0	0
3	1	1/3	0	1	0	0	0
4	0	−5/3	0	−1	0	1	0

ЛИТЕРАТУРА

1. *Юби Э.* Изв. АН Эстонии. Физ. Матем., 1989, 38, № 4, 423—432.
2. *Лоусон Ч., Хенсон Р.* Численное решение задач метода наименьших квадратов. М., Наука, 1986.
3. *Пападимитриу Х., Стайглиц К.* Комбинаторная оптимизация. Алгоритмы и сложность. М., Мир, 1985.
4. *Giesen, G.* Unternehmensforschung, 1961, 5, H. 3, 132—139.

Поступила в редакцию
13/VI 1990

Evald ÜBI

ALGBAAS VÄHIMRUUTUDE MEETODI ABIL. PRIMAARNE JA DUAALNE SIMPLEKSMEETOD

Algbaasi määramiseks on kaks korda kasutatud varem esitatud alamääratud lineaarse võrrandisüsteemi mittenegatiivse lahendi leidmise meetodit.

Evald ÜBI

AN INITIAL BASIS VIA THE LEAST SQUARES METHOD. THE DIRECT AND DUAL SIMPLEX METHOD

Determining of the initial basis for a linear programming problem is reduced to finding a solution to an extended system of underdetermined linear equations $Ax=b$, $(c,x)-z=M$, where all variables x_j are non-negative, $j=1, \ldots, n$; it is recommendable that $M\leqslant z^*=(c,x^*)$, where M is given before solution as an expected value of minimum z^*.

To solve the problem we use the previously suggested scheme of the least squares method application. At the first step the function $\varphi(x,z)=1/2\|Ax-b\|^2+$ $+1/2[(c,x)-z-M]^2$ is minimized on positive orthant with fixed $z=0$ on the second step with z — free.

The initial basis defined in such a way depends also on the coefficients of the objective function, and quite often it happens to be optimal.