

ОБ ОДНОМ СПОСОБЕ ПОСТРОЕНИЯ НАЧАЛЬНОГО ДОПУСТИМОГО БАЗИСА В ЗАДАЧАХ ОПТИМИЗАЦИИ

Е. А. Котельников

Институт вычислительной математики и математической геофизики СО РАН,
630090, Новосибирск, Россия

УДК 519.853.32

Предложен алгоритм построения начального допустимого базиса в задачах математического программирования с линейными ограничениями, заданными разреженными матрицами большой размерности. В данном случае тип целевой функции (линейная, квадратичная или нелинейная функция) несуществен.

Ключевые слова: оптимизация, линейные ограничения, начальный базис.

In paper the algorithm of construction of initial admissible basis in problems of mathematical programming with the linear restrictions set by rarefied matrixes of the big dimension is offered. The criterion function type can be any. It can be linear function, quadratic or nonlinear function of a general view.

Key words: optimization, linear restrictions, initial basis.

Введение. Пусть система ограничений исходной задачи оптимизации представлена в форме

$$Ax = b; \quad (1)$$

$$\alpha \leq x \leq \beta. \quad (2)$$

Здесь A — матрица размером $m \times n$ ($m < n$); $\alpha, x, \beta \in \mathbb{R}^n$; $b \in \mathbb{R}^m$. Если в исходной задаче среди условий (1) имеются ограничения типа неравенств, то, введя дополнительные неотрицательные переменные, такие ограничения можно преобразовать к равенствам. В расширенной таким образом матрице A дополнительным переменным соответствуют единичные орты (с тем или иным знаком) пространства $\mathbb{R}^m$. Исходная задача может содержать неограниченные (свободные) переменные x_j , что соответствует равенствам $\alpha_j = -\infty$, $\beta_j = +\infty$. Наличие таких переменных упрощает построение начального допустимого базиса.

Для построения начального допустимого базиса существует ряд методов. Во многих коммерческих пакетах (см. [1]) в фазе I (фазе поиска начального базиса) решается следующая вспомогательная задача линейного программирования: минимизировать $\sum_{j=1}^m x_{n+j}$ при условиях

$$a_i^T x \pm x_{n+i} = b_i, \quad i = 1, 2, \dots, m, \quad \alpha \leq x \leq \beta, \quad x_{n+i} \geq 0, \quad i = 1, 2, \dots, m. \quad (3)$$

Здесь a_i^T — i -я строка матрицы A .

Работа рекомендована к публикации Программным комитетом VIII Международной азиатской школы-семинара “Проблемы оптимизации сложных систем”.

Знак в ограничениях (3) определяется следующим образом. В точке $\bar{\mathbf{x}}$ с координатами

$$\bar{x}_i = \begin{cases} \alpha_i, & \text{если } \alpha_i > -\infty, \\ \beta_i, & \text{если } \alpha_i = -\infty, \beta_i < +\infty, \\ 0 & \text{в остальных случаях} \end{cases} \quad (4)$$

вычисляется вектор $\bar{\mathbf{b}} = \mathbf{b} - A\bar{\mathbf{x}}$. Если для некоторых i $\bar{b}_i < 0$, то в i -м ограничении (3) должен быть знак “минус”, в противном случае — знак “плюс”. Это позволяет выбрать в качестве начальной допустимой точки $\mathbf{x} \in \mathbb{R}^{n+m}$ вспомогательной задачи точку с координатами $x_i = \bar{x}_i$ ($i = 1, 2, \dots, n$); $x_{n+i} = |\bar{b}_i|$ ($i = 1, 2, \dots, m$), в качестве начальной базисной матрицы — диагональную матрицу размером $m \times m$ с элементами ± 1 на диагонали.

Очевидно, что если допустимое множество исходной задачи не пусто, то в оптимальном решении вспомогательной задачи $\sum_{j=1}^m x_{n+j} = 0$ и ее оптимальный базис может служить начальным для фазы II (фазы решения исходной задачи).

Все модификации данного метода сводятся к попыткам уменьшить число искусственных переменных x_{n+i} путем включения в начальный базис фазы I дополнительных или структурных переменных, на которые накладывается требование допустимости их значений. Будем считать, что дополнительная переменная x_j имеет границы допустимости $\alpha_j = 0$, $\beta_j = -\infty$, а искусственная — $\alpha_j = \beta_j = 0$.

Алгоритм. В настоящей работе предлагается способ построения начального допустимого базиса, позволяющий использовать на начальном этапе фазы I большее количество структурных и дополнительных переменных, уделяя основное внимание численной сбалансированности начальной базисной матрицы. Из всего набора столбцов матрицы A выбираем (если это возможно) m линейно независимых столбцов и формируем из них базисную матрицу B . Если это не удастся сделать, то недостающую часть дополняем единичной подматрицей с соответствующим набором искусственных переменных. Проблема построения начальной базисной матрицы B рассматривается ниже.

Столбцы матрицы A , не вошедшие в матрицу B , объединим в матрицу небазисных столбцов N . Любому небазисному столбцу $\mathbf{a}_j$ соответствует переменная x_{Nj} , значение которой всегда равно одному из ее граничных значений ($x_{Nj} = \alpha_{Nj}$ или $x_{Nj} = \beta_{Nj}$). Далее каждой небазисной переменной x_{Nj} каким-либо образом (например, по правилу, заданному в (4)) присваивается значение $\bar{x}_{Nj}$. Тогда, если все значения базисных переменных $\mathbf{x}_B = B^{-1}(\mathbf{b} - N\bar{\mathbf{x}}_N)$ имеют допустимые значения, построение начального допустимого базиса заканчивается. В противном случае переопределим значения базисных переменных следующим образом: если $x_{Bi} < \alpha_{Bi}$, то $\bar{x}_{Bi} = \alpha_{Bi}$; если $\alpha_{Bi} \leq x_{Bi} \leq \beta_{Bi}$, то $\bar{x}_{Bi} = x_{Bi}$; если $x_{Bi} > \beta_{Bi}$, то $\bar{x}_{Bi} = \beta_{Bi}$. Здесь $\alpha_B, \beta_B \in \mathbb{R}^m$ — векторы, составленные из компонент векторов α, β , которые (компоненты) соответствуют базисным переменным.

Пусть $\mathbf{r} = \mathbf{x}_B - \bar{\mathbf{x}}_B$ и $\mathbf{b}' = B\mathbf{r}$, тогда верно равенство $B\bar{\mathbf{x}}_B + N\bar{\mathbf{x}}_N + \mathbf{b}' = \mathbf{b}$. Поставим в соответствие столбцу $\mathbf{b}'$ искусственную переменную x_{n+1} и запишем систему ограничений (1) в виде системы

$$B\mathbf{x}_B + N\mathbf{x}_N + \mathbf{b}'x_{n+1} = \mathbf{b},$$

решением которой является точка $\mathbf{x}_0^T = (\bar{\mathbf{x}}_B^T, \bar{\mathbf{x}}_N^T, x_{n+1})$ при $x_{n+1} = 1$. Допустимое значение переменной x_{n+1} есть только нулевое значение, поэтому на следующем этапе устраним недопустимость переменной x_{n+1} , используя точку x_0 в качестве начальной.

Если столбец $\mathbf{b}'$ плохо масштабирован по сравнению со столбцами матрицы A , например если $\|\mathbf{b}'\|_\infty < \min \|\mathbf{a}_i\|_\infty$ или $\|\mathbf{b}'\|_\infty > \max \|\mathbf{a}_i\|_\infty$ ($\mathbf{a}_i$ — столбец матрицы A), то можно

нормировать вектор $\mathbf{b}'$: $\mathbf{b}' = \mathbf{b}'/q$, сбалансировав значения $\|\mathbf{b}'\|_\infty$ и $\|\mathbf{a}_i\|_\infty$, $i = 1, 2, \dots, n$, где $q > 0$ — некоторое число. Тогда новое начальное значение переменной x_{n+1} будет равно q . Переменную x_{n+1} нельзя оставлять небазисной, так как она не равна своему граничному значению. Поэтому, прежде чем устранить недопустимость переменной x_{n+1} , введем ее в базис, заменив ею одну из базисных переменных. Для этого необходимо заменить вектор $\mathbf{b}'$ на один из базисных столбцов, так чтобы не была нарушена невырожденность базисной матрицы и заменяемая базисная переменная x_{Bi} имела значение α_{Bi} или β_{Bi} .

Известно, что матрица B останется невырожденной при замене базисного столбца с номером i в базисе на столбец $\mathbf{b}'$ если i -я компонента вектора $B^{-1}\mathbf{b}'$ не равна нулю. Рассмотрим вектор $\mathbf{f} = B^{-1}\mathbf{b}' = B^{-1}(B\mathbf{r}/q) = \mathbf{r}/q$ и определим множество J номеров базисных переменных, таких что для любого $j \in J$ $x_{Bj} = \alpha_{Bj}$ или $x_{Bj} = \beta_{Bj}$. Множество J не пусто, так как в противном случае вектор $\mathbf{b}'$ был бы равен нулю. Пусть I — множество ведущих элементов в столбцах множества J , $i_0 = \arg \max_{i \in I} |f_i|$ и $j_0 \in J$ — номер столбца с номером ведущего элемента i_0 . Тогда столбец $\mathbf{a}_{j_0}$ можно вывести из матрицы B и перевести в разряд небазисных, а столбец $\mathbf{b}'$ ввести в базис с номером ведущего элемента i_0 . Новая базисная матрица невырождена, так как $|f_{i_0}| = \max_i |x_{Bi} - \bar{x}_{Bi}|/q$.

Последний этап фазы I состоит в исключении из базиса искусственной переменной x_{n+1} при условии допустимости всех остальных базисных переменных. С этой целью используем алгоритм, предложенный в [1]. В данном алгоритме на каждой итерации величина сдвига вдоль выбранного направления вычисляется таким образом, чтобы суммарная величина отклонений базисных переменных от их отрезков допустимости была минимальной. На каждой итерации фазы I коэффициенты целевой функции, подлежащей минимизации, задаются в виде

$$c_i = \begin{cases} -1, & \text{если } x_i < \alpha_i, \\ +1, & \text{если } x_i > \beta_i, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Напомним, что дополнительная переменная x_i имеет границы допустимости $\alpha_i = 0$, $\beta_i = +\infty$, а искусственная — $\alpha_i = \beta_i = 0$. Коэффициенты c_i используются для вычисления относительных оценок d_j небазисных столбцов $\mathbf{a}_j$ в модифицированном симплекс-методе [1]. На основе значений d_j принимается решение о целесообразности ввода в базис столбца $\mathbf{a}_j$. Если на каком-либо шаге ни один небазисный столбец не может быть введен в базис (выполнен признак оптимальности), то полученное решение является оптимальным. Будем считать, что при $x_j = 0$ j -я свободная переменная ($\alpha_j = -\infty$, $\beta_j = +\infty$) может быть небазисной, а знак относительной оценки d_j при выборе j -го столбца для ввода в базис несуществен. В оптимальном решении d_j свободной небазисной переменной x_j должно быть равно нулю.

Пусть на текущей итерации для ввода в базис выбран небазисный столбец $\mathbf{a}_j$. Тогда, как известно, вектор сдвига $\mathbf{p}_B$ базисных переменных можно вычислить по формуле $\mathbf{p}_B = -B^{-1}\mathbf{a}_j$. Величину сдвига λ вдоль направления $\mathbf{p}_B$ будем выбирать из условия минимума суммы отклонений $\sum_{i \in I_B} \delta_i$ (I_B — список базисных переменных) вектора $\bar{\mathbf{x}}_B = \mathbf{x}_B + \lambda \mathbf{p}_B$ от параллелепипеда $\alpha_B \leq \mathbf{x}_B \leq \beta_B$, где

$$\delta_i = \begin{cases} \alpha_{Bi} - \bar{x}_{Bi}, & \text{если } \bar{x}_{Bi} < \alpha_{Bi}, \\ \bar{x}_{Bi} - \beta_{Bi}, & \text{если } \bar{x}_{Bi} > \beta_{Bi}, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (5)$$

Будем считать, что небазисная переменная x_{Nj} , которая вводится в базис, не нарушает границы допустимости, т. е. выполняется условие $\lambda \leq \beta_{Nj} - \alpha_{Nj}$.

Задача минимизации $\sum_{i \in I_B} \delta_i$ сводится к задаче минимизации кусочно-линейной выпуклой функции $\varphi(x)$ одной переменной при условии $x \geq 0$. Функция φ является суммой кусочно-линейных функций φ_i , определенных для базисных переменных x_i ($i \in I_B$):

1. Если $x_{Bi} < \alpha_{Bi}$ и $p_{Bi} < 0$, то $\varphi_i(x) = 0$.
2. Если $x_{Bi} < \alpha_{Bi}$ и $p_{Bi} > 0$, то

$$\varphi_i(x) = \begin{cases} p_{Bi}(\gamma_i - x), & \text{если } x < \gamma_i, \\ p_{Bi}(x - \eta_i), & \text{если } x > \eta_i, \\ 0, & \text{если } \gamma_i \leq x \leq \eta_i, \end{cases}$$

где $\gamma_i = (\alpha_{Bi} - x_{Bi})/p_{Bi}$; $\eta_i = (\beta_{Bi} - x_{Bi})/p_{Bi}$.

3. Если $x_{Bi} > \beta_{Bi}$ и $p_{Bi} > 0$, то $\varphi_i(x) = 0$.
4. Если $x_{Bi} > \beta_{Bi}$ и $p_{Bi} < 0$, то

$$\varphi_i(x) = \begin{cases} -p_{Bi}(\gamma_i - x), & \text{если } x < \gamma_i, \\ -p_{Bi}(x - \eta_i), & \text{если } x > \eta_i, \\ 0, & \text{если } \gamma_i \leq x \leq \eta_i, \end{cases}$$

где $\gamma_i = -(x_i - \beta_{Bi})/p_{Bi}$; $\eta_i = -(x_{Bi} - \alpha_{Bi})/p_{Bi}$.

5. Если $\alpha_{Bi} \leq x_{Bi} \leq \beta_{Bi}$, то

$$\varphi_i(x) = \begin{cases} |p_{Bi}|(x - \eta_i), & \text{если } x > \eta_i, \\ 0, & \text{если } x \leq \eta_i, \end{cases}$$

где

$$\eta_i = \begin{cases} (x_{Bi} - \alpha_{Bi})/p_{Bi}, & \text{если } p_{Bi} < 0, \\ (\beta_{Bi} - x_{Bi})/p_{Bi}, & \text{если } p_{Bi} > 0. \end{cases}$$

6. Если $p_{Bi} = 0$, то $\varphi_i(x) = 0$.

Упорядочим величины γ_i , η_i ($i = 1, 2, \dots, m$) по возрастанию их значений, запомним в соответствующем порядке номера переменных и обозначим через σ_i элементы отсортированного массива. Тогда на каждом отрезке $[\sigma_i, \sigma_{i+1}]$ функция φ линейна, а в точках σ_i имеет изломы: $\varphi(0) = \sum_{i \in I_2} p_{Bi} - \sum_{i \in I_1} p_{Bi}$. Здесь $I_1 = \{i : x_{Bi} < \alpha_{Bi}\}$; $I_2 = \{i : x_{Bi} > \beta_{Bi}\}$. Нетрудно проверить, что $\varphi(0)$ равна относительной оценке небазисного столбца, вводимого в базис: $d_j = c_{Nj} - \mathbf{c}_B^T B^{-1} \mathbf{a}_j = \mathbf{c}_B^T \mathbf{p}_B = \varphi(0)$.

Целью минимизации функции φ является нахождение номера переменной, которая покидает список базисных столбцов, и величины сдвига вдоль $\mathbf{p}_B$. Для этого последовательно по возрастанию индекса перебираются элементы σ_i и вычисляются величины $\tau_i = \varphi(0) - \sum_{k=1}^i p_{Bi_k}$, где i_k — номер базисной переменной, стоящей на k -м месте после сортировки. Перебор осуществляется до тех пор, пока либо величина τ_i перестанет быть положительной, либо величина σ_i станет больше $\beta_{Nj} - \alpha_{Nj}$. В первом случае номер переменной i_k , при котором $\tau_{i_k} \leq 0$, выбирается в качестве номера столбца, выводимого из базисной матрицы, а в качестве искомой величины сдвига λ^* вдоль вектора $\mathbf{p}_B$ выбирается σ_{i_k} . Если после перебора всех σ_i величина τ_i остается положительной, то в качестве λ^* выбирается $\beta_{Nj} - \alpha_{Nj}$ и состав базиса не меняется, а небазисная переменная x_{Nj} меняет свое граничное значение.

Фаза I считается законченной, если на некоторой итерации значения всех базисных переменных находятся в пределах своих границ допустимости, что означает принадлежность найденной точки множеству, определенному условиями (1) и (2), или если существуют недопустимые базисные переменные, при этом все относительные оценки небазисных переменных удовлетворяют условию оптимальности, что соответствует отсутствию допустимого решения в исходной задаче.

Вернемся к задаче построения начальной базисной матрицы B . Суть задачи состоит в следующем: из всего набора столбцов матрицы A следует выбрать m линейно независимых, так чтобы степень заполненности матрицы B^{-1} незначительно превышала степень заполненности матрицы B , а численная устойчивость разложения матрицы $B^{-1} = U^{-1}L^{-1}$ была достаточно высокой. Для решения данной задачи используем процедуру перепостроения обратных базисных матриц P^3 (preassigned pivot procedure) [2], модифицировав ее с целью увеличения численной устойчивости разложения $B^{-1} = U^{-1}L^{-1}$. Процедура P^3 представляет собой эвристический алгоритм приведения с помощью перестановок строк и столбцов матрицы B к виду, по возможности близкому к нижней треугольной матрице. При этом предпринимается попытка минимизировать как число столбцов, имеющих ненулевые элементы над главной диагональю, так и количество нулей в этих столбцах над главной диагональю. Столбцы, не имеющие ненулевых элементов над главной диагональю, называются нормальными, а имеющие — столбцами-выступами. Если матрица B состоит из нормальных столбцов и столбцов-выступов, причем последние расположены в порядке возрастания высоты столбцов, то новые ненулевые элементы в матрице B^{-1} могут возникнуть только в столбцах, которые соответствуют столбцам-выступам матрицы B , причем на месте нулевых элементов над главной диагональю. В данном случае k — высота столбца $\mathbf{a}_j$, если $a_{kj} \neq 0$ и $a_{ij} = 0$ при $i < k$.

После упорядочивания строк и столбцов ведущие элементы выбираются на главной диагонали сверху вниз и слева направо. Однако при таком правиле выбора ведущих элементов обеспечить численную устойчивость разложения матрицы $B^{-1} = U^{-1}L^{-1}$ достаточно трудно. Поэтому при получении очередного ведущего элемента нужно проконтролировать его величину и в случае нарушения допустимого уровня перевести столбец в резерв, чтобы иметь возможность использовать его в более подходящих условиях. Для проверки величины ведущего элемента можно использовать стандартный контроль ведущих элементов на устойчивость и отложить использование столбца, если ведущий элемент g_i не удовлетворяет условию $|g_i| \geq u \max_{k \in K} |g_k|$, где $0 < u \leq 1$; K — список индексов столбцов, для которых еще не назначены ведущие элементы. В конце процесса разложения ведущие элементы в отложенных столбцах определяются путем частичного выбора в столбцах.

Описанная модификация процедуры P^3 использовалась в качестве процедуры перепостроения обратной базисной матрицы в первых версиях пакетов программ для решения задач линейного и квадратичного программирования и их целочисленных вариантов [3, 4].

Применение модифицированной процедуры P^3 для формирования начальной матрицы B из столбцов матрицы A упрощается за счет того, что на каждом шаге список столбцов-кандидатов на ввод в базис шире текущего списка кандидатов при перепостроении обратной базисной матрицы. При этом в рассматриваемом случае существует больше возможностей при наборе столбцов в так называемый нижний треугольник. Кандидатом на включение в “нижний треугольник” в текущий момент является столбец, у которого для всех номеров ненулевых элементов (кроме одного) уже определены ведущие элементы.

Процессы формирования начальной базисной матрицы B и вычисления вектора $\mathbf{b}'$ можно модифицировать, выполняя их параллельно. На каждом шаге по правилам процедуры P^3 набираем множество J — список номеров небазисных столбцов-кандидатов на ввод в базис, прошедших стандартный контроль устойчивости. Затем для каждого $j \in J$ вычисляем значение j -й переменной x_j по текущим значениям остальных переменных. В качестве первоначального значения всем небазисным переменным x_j присваиваем значения $\bar{x}_j$, определенные в (4). Для ввода в базис выбираем столбец с номером $j_0 = \arg \min_{j \in J} \{\delta_j \|\mathbf{a}_j\|_\infty\}$, где δ_j — величина отклонения, определенная в (5); вектор $\mathbf{b}'$ пересчитывается с учетом δ_{j_0} , а в качестве нового значения x_{j_0} выбирается величина

$$\bar{x}_{j_0} = \begin{cases} \alpha_{j_0}, & \text{если } x_{j_0} < \alpha_{j_0}, \\ \beta_{j_0}, & \text{если } x_{j_0} > \beta_{j_0}, \\ x_{j_0}, & \text{если } \alpha_{j_0} \leq x_{j_0} \leq \beta_{j_0}. \end{cases}$$

Данная процедура позволяет ожидать, что в конечном счете может быть получено наименьшее возможное значение $\|\mathbf{b}'\|_\infty$.

Последняя модификация процессов построения начальной базисной матрицы B и вычисления вектора $\mathbf{b}'$, а также процедура устранения недопустимости, описанная выше, были использованы для одного из двух вариантов построения начального допустимого базиса во всех пакетах программ при решении задач оптимизации с линейными ограничениями [3–5]. Результаты расчетов показывают высокую численную устойчивость вычислительного процесса построения начального базиса.

Список литературы

1. МУРТАФ Б. Современное линейное программирование. Теория и практика. М.: Мир, 1984.
2. HELLERMAN E., RARICK D. C. Reinversion with the pressigned pivot procedure // Math. Program. 1971. V. 1, N 2. P. 195–216.
3. ЗАБИНЯКО Г. И. Пакет программ целочисленного линейного программирования // Дискретный анализ и исследование операций. 1999. Сер. 2. Т. 6, № 2. С. 32–41.
4. ЗАБИНЯКО Г. И., КОТЕЛЬНИКОВ Е. А. Параллельный алгоритм целочисленного квадратичного программирования // Вычисл. технологии. 2004. Т. 9, № 1. С. 34–41.
5. ЗАБИНЯКО Г. И., КОТЕЛЬНИКОВ Е. А., РОЖИН В. Е. Программы минимизации нелинейных функций при линейных ограничениях: Отчет / ВЦ СО РАН. № ГР 01.9.30 001317; Инв. № 02.9.70 004793. Новосибирск, 1997.

Котельников Евгений Алексеевич — ст. науч. сотр. Института вычислительной математики и математической геофизики СО РАН; тел. (383) 330-60-66

Дата поступления — 2.04.12