

УДК 330.105

МЕТОД ВЕТВЕЙ И ГРАНИЦ ДЛЯ КВАДРАТИЧНОЙ ЗАДАЧИ ДВУХУРОВНЕВОГО ПРОГРАММИРОВАНИЯ*)

М. С. Ершова

Рассматривается задача двухуровневого программирования с квадратичными целевыми функциями и линейными ограничениями. Для её решения предлагается метод ветвей и границ, основанный на разбиении слабого допустимого множества и неявном разбиении индуктивной области. Нижняя оценка оптимального значения целевой функции достигается оптимизацией на эллипсоиде, аппроксимирующем допустимую область извне. В качестве верхней оценки используется локальное оптимальное значение целевой функции верхнего уровня при условии равенства нулю невязки двойственности в задаче нижнего уровня.

Введение

Задача двухуровневого программирования является задачей оптимизации целевой функции с экстремальными ограничениями. В задаче с одним экстремальным ограничением требуется найти вектор x^* из допустимого множества, на котором достигается минимум функция $F(x, y(x))$, где $y(x)$ — решение задачи минимизации функции $f(x, y)$ при векторном ограничении $g(x, y) \leq 0$ и фиксированном значении x .

Полагая, что допустимые значения переменной x удовлетворяют векторному неравенству $G(x, y) \leq 0$, задачу можно записать в виде: найти

$$\min_x F(x, y) \quad (1)$$

при ограничениях

$$G(x, y) \leq 0, \quad (2)$$

$$\min_y f(x, y), \quad (3)$$

$$g(x, y) \leq 0. \quad (4)$$

*) Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект 03-01-00518).

Следует заметить, что в этой постановке не подразумевается однозначное понятие оптимальной пары $(x^*, y^* = y(x^*))$, поскольку при неединственности решения задачи (3), (4) множество значений переменной x , на которых достигается минимум целевой функции F , зависит от конкретной реализации минимума в задаче (3), (4). Если в задаче (3), (4) строго выпуклы функции $f(x, y)$ и $g(x, y)$ при любом фиксированном x , то точка $y(x) \in \operatorname{Argmin}\{f(x, y) \mid g(x, y) \leq 0\}$ единственна и оптимальным решением считается любая точка $x^* \in \operatorname{Argmin}_x \{F(x, y(x)) \mid G(x, y(x)) \leq 0\}$. Будем рассматривать двухуровневую задачу при этом условии.

Задача двухуровневого программирования моделирует принятие оптимальных решений в иерархической системе и соответствует следующей игре двух лиц. Первый игрок, называемый игроком верхнего уровня, располагает переменной x и решает задачу (1), (2) при условии, что его целевая функция и ограничения зависят от значения переменной y , выбранной игроком нижнего уровня. Игроку верхнего уровня известно, что значение y^* переменной y будет определено в результате решения задачи (3), (4) после того, как зафиксирована величина x^* .

С помощью задач двухуровневого программирования моделируются многие экономические ситуации, например, проблема определения эффективной политики государства в отношении развития отдельных отраслей производства. Рассмотрим следующую постановку. Государство заинтересовано в том, чтобы стимулировать производство биотоплива, получаемого из сельскохозяйственных культур и известного своей экологической безопасностью по сравнению с обычным топливом. Инструментами государства являются кредиты и субсидии, предоставляемые фермерам и промышленности на производство соответствующей продукции. Рассматривая государство как верхний уровень, который стремится минимизировать затраты на разрабатываемую программу, а фермерские хозяйство — как нижний уровень, чей интерес состоит в максимизации прибыли, получаем задачу двухуровневого программирования [3]. Эта задача относится к классу квадратичных двухуровневых задач с невыпуклой квадратичной функцией на верхнем уровне и линейными ограничениями.

В общем случае двухуровневая задача является многоэкстремальной и NP-трудной, даже если все функции, входящие в формулировку, линейны [3, 6].

Среди задач непрерывного двухуровневого программирования наиболее исследованными являются линейные задачи. Для получения гло-

бальных оптимумов в линейных задачах разработан ряд методов, многие из которых могут применяться для решения линейно квадратичной задачи, когда целевая функция верхнего уровня и функции ограничений линейны, а целевая функция нижнего уровня является строго выпуклой квадратичной.

Задача с линейными ограничениями, строго выпуклой квадратичной целевой функцией нижнего уровня и с нелинейной целевой функцией на верхнем уровне является более сложной для решения. Разработанные методы позволяют получать либо только локальный оптимум в такой задаче [6], либо глобальный оптимум, но лишь для частных видов целевой функции верхнего уровня: выпуклой, вогнутой, а также представимой в виде разности двух выпуклых функций [3].

Квадратичные задачи двухуровневого программирования со строго выпуклой квадратичной целевой функцией нижнего уровня, произвольной квадратичной целевой функцией на верхнем уровне и линейными ограничениями исследованы мало, и очевидна необходимость в разработке эффективных методов достижения глобального оптимума в таких задачах.

1. Задача и её свойства

Рассматривается задача двухуровневого программирования: найти

$$\min_x F(x, y) \quad (5)$$

при ограничениях

$$G_i(x, y) \leq 0, \quad i = 1, \dots, P; \quad (6)$$

$$\min_y f(x, y); \quad (7)$$

$$g_i(x, y) \leq 0, \quad i = 1, \dots, p. \quad (8)$$

Здесь $x \in \mathbb{E}^N$, $y \in \mathbb{E}^n$ — векторы переменных соответственно верхнего и нижнего уровня; F и f — квадратичные по совокупности переменных функции, причём f — строго выпуклая по y при фиксированном x ; G_i и g_i — линейные функции. Пусть в задаче нижнего уровня (7), (8) допустимая область ограничена.

Ограничения (6) и (8) будем записывать в форме векторных неравенств

$$G(x, y) \leq 0, \quad g(x, y) \leq 0.$$

Функции, фигурирующие в постановке, имеют вид:

$$F(x, y) = \frac{1}{2}x^T W_1 x + x^T W_2 y + \frac{1}{2}y^T W_3 y + w_1^T x + w_2^T y,$$

$$f(x, y) = \frac{1}{2}y^T Q_1 y + x^T Q_2 y + q^T y,$$

$$G(x, y) = A_1 x + B_1 y - b_1,$$

$$g(x, y) = A_2 x + B_2 y - b_2,$$

а также $G_i(x, y) = (A_1)_i x + (B_1)_i y - (b_1)_i$ и $g_i(x, y) = (A_2)_i x + (B_2)_i y - (b_2)_i$. Здесь $W_1, W_2, W_3, Q_1, Q_2, A_1, B_1, A_2, B_2$ — матрицы, а w_1, w_2, q, b_1, b_2 — векторы соответствующих размерностей; $(A)_i$ и $(B)_i$ — i -е строки матриц A и B , b_i — i -я компонента вектора b . Матрицы W_1, W_3 размера $N \times N$ и матрица Q_1 размера $n \times n$ — симметричные, причём Q_1 положительно определена.

Двухуровневая задача (5)–(8) может быть представлена как одноуровневая: найти

$$\min_{x, y} F(x, y) \text{ при } (x, y) \in \Psi,$$

где Ψ — *индуктивная область*. Она является частью слабого допустимого множества $\Omega = \{(x, y) \mid G(x, y) \leq 0, g(x, y) \leq 0\}$ и записывается в виде

$$\Psi = \left\{ (x, y) \mid (x, y) \in \Omega, y \in \text{Arg min}_{\eta} \{f(x, \eta) \mid g(x, \eta) \leq 0\} \right\}.$$

Индуктивная область может быть невыпуклой, а также несвязной: она состоит из точек множества Ω , удовлетворяющих некоторому кусочно-линейному равенству. Точнее, она включает в себя часть границы множества Ω и те точки гиперплоскости $\{(x, y) \in \mathbb{E}^{N+n} \mid \nabla_y f(x, y) = 0\}$, которые принадлежат Ω [3].

2. Метод ветвей и границ

Для решения рассматриваемой задачи предлагается метод ветвей и границ, осуществляющий направленный перебор фрагментов индуктивной области.

В результате последовательного разбиения индуктивной области при помощи гиперплоскостей на каждой итерации предлагаемого метода имеется набор \mathcal{P} подмножеств индуктивной области, каждое из которых может содержать точку глобального оптимума. Для элемента разбиения $\hat{\Psi} \in \mathcal{P}$ определяются нижняя $\underline{F}(\hat{\Psi})$ и верхняя $\overline{F}(\hat{\Psi})$ оценки оптимального значения целевой функции на этом множестве:

$$\underline{F}(\hat{\Psi}) \leq \min\{F(x, y) \mid (x, y) \in \hat{\Psi}\} \leq \overline{F}(\hat{\Psi}).$$

Вычисляются $\underline{F} = \min\{\underline{F}(\hat{\Psi}), \hat{\Psi} \in \mathcal{P}\}$, $\overline{F} = \min\{\overline{F}(\hat{\Psi}), \hat{\Psi} \in \mathcal{P}\}$, так что

$$\underline{F} \leq \min\{F(x, y) \mid (x, y) \in \Psi\} \leq \overline{F}.$$

Если $\overline{F} - \underline{F}$ — достаточно малая величина, работа алгоритма завершается и в качестве решения выбирается допустимая точка, в которой значение целевой функции равно \overline{F} . Иначе выбирается такое множество $\hat{\Psi} \in \mathcal{P}$, что оценка $\underline{F}(\hat{\Psi})$ наименьшая, производится его разбиение и т. д. Если для элемента разбиения $\hat{\Psi}$ нижняя оценка $\underline{F}(\hat{\Psi})$ больше, чем наименьшая среди полученных верхних оценок \overline{F} , то из перспективного набора \mathcal{P} множество $\hat{\Psi}$ удаляется как не содержащее оптимального решения.

Ввиду того, что индуктивная область Ψ задана неявно, её разбиение осуществляется также неявно на основе разбиения слабого допустимого множества Ω . Элемент разбиения $\hat{\Omega}$ слабого допустимого множества задаётся линейными неравенствами и содержит фрагмент $\hat{\Psi}$ индуктивной области $\hat{\Omega}$.

3. Внешняя аппроксимация слабого допустимого множества

Для построения разбиений допустимой области и для оценивания оптимального значения целевой функции будем использовать внешнюю аппроксимацию многогранника при помощи эллипсоида наименьшего объёма.

Сначала заметим, что линейное неравенство, задающее совместно с другими неравенствами ограниченное множество, может быть представлено как квадратичное. Геометрически это означает замену полупространства, заданного линейным неравенством, полосой — областью, ограниченной двумя параллельными гиперплоскостями. Действительно, пусть неравенство $c^T z \leq h^+$ есть одно из ограничений, задающих выпуклую ограниченную область C в пространстве \mathbb{E}^m . Вычислим $h^- = \min\{c^T z \mid z \in C\}$ и определим полосу

$$H = \{z \in \mathbb{E}^m \mid h^- \leq c^T z \leq h^+\}.$$

Понятно, что $C \subset H$. Пусть $d^+ = \frac{h^+ + h^-}{2}$, $d^- = \frac{h^+ - h^-}{2}$. Тогда справедливо эквивалентное представление полосы при помощи квадратичной формы с матрицей ранга 1:

$$\begin{aligned} H &= \{z \mid (c^T z - d^+)^2 \leq (d^-)^2\} = \{z \mid z^T c c^T z - 2d_+ c^T z \leq (d^-)^2 - (d^+)^2\} \\ &= \{z \mid (z - z^0)^T c c^T (z - z^0) \leq (d^-)^2\}, \end{aligned}$$

где z^0 — любой такой вектор, что $c^T z^0 = d^+$.

Опишем эллипсоид вокруг множества C , заданного одним выпуклым квадратичным неравенством и линейными ограничениями. Множество

C после соответствующего преобразования координат можно записать в виде

$$C = \{z \in \mathbb{E}^m \mid z^T U z \leq 1, (z - z_i^0)^T c_i c_i^T (z - z_i^0) \leq (d_i^-)^2, i = 1, \dots, s\}, \quad (9)$$

где U — положительно определённая симметричная матрица. Область

$$S = \{z \in \mathbb{E}^m \mid \alpha_0 z^T U z + \sum_{i=1}^s \alpha_i (z - z_i^0)^T c_i c_i^T (z - z_i^0) \leq \alpha_0 + \sum_{i=1}^s \alpha_i (d_i^-)^2\}, \quad (10)$$

где α_i — любые неотрицательные коэффициенты, задаёт эллипсоид, в котором содержится множество C . Будем полагать, что $\sum_{i=0}^s \alpha_i = 1$, а наилучшим выбором для α_i будем считать такой, при котором объём линейной комбинации (10) наименьший. Благодаря строгой квазивыпуклости объёма такой линейной комбинации множеств относительно своих коэффициентов [1] последняя задача является задачей выпуклого программирования.

В случае $s = 1$ оптимальные коэффициенты, которые обеспечивают эллипсоиду S наименьший объём, определяются явно следующим образом. Пусть выполнено преобразование координат такое, что справедливо нормирующее условие $c_1^T U^{-1} c_1 = 1$. Тогда α_0 вычисляется как больший корень квадратного уравнения [1]

$$(m+1)(d^+)^2 \alpha^2 + (1 - (d^+)^2 - (d^-)^2) \alpha - (m-1)(d^-)^2 = 0,$$

а $\alpha_1 = 1 - \alpha_0$. Для решения задачи в случае произвольного s можно воспользоваться той же техникой, применяя её последовательно: на первом шаге к первым двум неравенствам из (9), на каждом следующем шаге — к результату предыдущего шага и очередному ограничению из (9). С каждым шагом уменьшается объём аппроксимирующего множества [2].

Поскольку множество Ω образовано линейными ограничениями, указанную технику можно применить для аппроксимации множества Ω эллипсоидом возможно меньшего объёма. Сначала при помощи линейной комбинации неравенств определяется внешний эллипсоид и его объём последовательно сокращается за счёт включения в линейную комбинацию вида (10) одного за другим всех линейных ограничений. Обозначим через S внешнюю эллипсоидальную аппроксимацию.

4. Разбиение допустимого множества

Разбиение фрагмента слабого допустимого множества будем определять разбиением его внешней аппроксимации. Введём переменную $z = (x^T, y^T)^T \in \mathbb{E}^{N+n}$. Пусть после аффинного преобразования координат, обеспечивающего приведение к главным осям,

$$S = \{z \in \mathbb{E}^{N+n} \mid (z - z_c)^T \Lambda (z - z_c) \leq 1\},$$

где $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_{N+n}\}$. Пусть $\lambda_{i_0} = \min\{\lambda_1, \dots, \lambda_{N+n}\}$. Возьмём вектор $c = (0, \dots, 0, 1, 0, \dots, 0)^T \in \mathbb{E}^{N+n}$ с i_0 -й единичной компонентой и положим $S^1 = \{z \in S \mid c^T(z - z_c) \leq 0\}$, $S^2 = \{z \in S \mid c^T(z - z_c) \geq 0\}$. Другими словами, локализующий эллипсоид делится гиперплоскостью, проходящей через его центр ортогонально наибольшей оси. Следовательно, нормаль секущей плоскости параллельна наибольшей оси. Соответственно $\hat{\Omega}$ и $\hat{\Psi}$ разбиваются на множества

$$\hat{\Omega}^1 = \hat{\Omega} \cap S^1, \hat{\Omega}^2 = \hat{\Omega} \cap S^2 \text{ и } \hat{\Psi}^1 = \hat{\Psi} \cap S^1, \hat{\Psi}^2 = \hat{\Psi} \cap S^2.$$

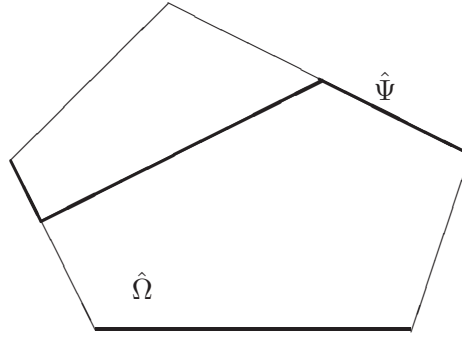


Рис. 1. Элемент разбиения

Если нормаль секущей плоскости определяется неоднозначно, следует выбрать то её направление, которому в ходе работы алгоритма до сих пор соответствовало наименьшее число делений.

На рис. 1 показан некоторый элемент разбиения слабого допустимого множества $\hat{\Omega}$ и жирной линией выделен соответствующий элемент разбиения индуктивной области $\hat{\Psi}$. Рис. 2 иллюстрирует технику разбиений.

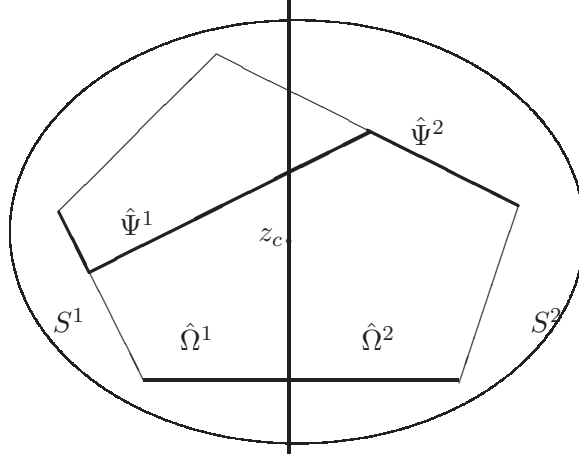


Рис. 2. Разбиение допустимой области

5. Верхняя оценка

Рассмотрим задачу нижнего уровня

$$\min_y \frac{1}{2} y^T Q_1 y + x^T Q_2 y + q^T y; \quad A_2 x + B_2 y \leq b_2, \quad (11)$$

которая является выпуклой квадратичной задачей математического программирования. Обозначив через $u \in \mathbb{E}^p$ вектор двойственных переменных, запишем задачу, двойственную к (11) [4]: найти

$$\max_u \varphi(x, u) \text{ при } u \geq 0.$$

Здесь целевая функция определяется следующим образом:

$$\varphi(x, u) = \frac{1}{2} u^T D_1 u + \frac{1}{2} x^T D_2 x + x^T D_3 u + c_1^T u + c_2^T x + d,$$

где $D_1 = -B_2^T Q_1^{-1} B_2$, $D_2 = -Q_2 Q_1^{-1} Q_2^T$, $D_3 = -Q_2 Q_1^{-1} B_2^T + A_2^T$, $c_1 = -B_2 Q_1^{-1} q - b_2$, $c_2 = -c^T Q_1^{-1} Q_2^T$, $d = -\frac{1}{2} q^T Q_1^{-1} q$. Заметим, что D_1 и D_2 — отрицательно определённые симметричные матрицы.

Невязка двойственности в (11) имеет вид

$$\begin{aligned} \Delta(x, y, u) \equiv f(x, y) - \varphi(x, u) &= \frac{1}{2} y^T Q_1 y + x^T Q_2 y + q^T y \\ &\quad - \left(\frac{1}{2} u^T D_1 u + \frac{1}{2} x^T D_2 x + x^T D_3 u + c_1^T u + c_2^T x + d \right) \end{aligned}$$

и является невыпуклой квадратичной функцией. Согласно теории двойственности её значения неотрицательны при всех $(x, y) \in \Omega$, $u \geq 0$, и равны нулю при $(x, y) \in \Psi$, $u \geq 0$.

Поставим задачу: найти

$$\min F(x, y) \text{ при ограничениях } \Delta(x, y, u) = 0, (x, y) \in \hat{\Omega}, u \geq 0.$$

Ограничения определяют множество $\hat{\Psi}$ — элемент разбиения индуктивной области. Ввиду сложности этой задачи ограничимся получением её локального решения. Пусть (\bar{x}, \bar{y}) — решение задачи: найти

$$\text{loc min } F(x, y) \text{ при ограничениях } \Delta(x, y, u) = 0, (x, y) \in \hat{\Omega}, u \geq 0. \quad (12)$$

Тогда $\bar{F}(\hat{\Psi}) \equiv F(\bar{x}, \bar{y})$ является верхней оценкой оптимального значения целевой функции на $\hat{\Psi}$: $F(x, y) \leq \bar{F}(\hat{\Psi})$ при любых $x, y \in \hat{\Psi}$. В качестве стартовой точки для локального спуска в (12) возьмём некоторую точку в множестве $\hat{\Omega}$.

Если задача (12) несовместна, то $\hat{\Psi} = \emptyset$ и элемент разбиения $\hat{\Omega}$ исключается из дальнейшего рассмотрения. Процедура локального поиска может показать несовместность задачи даже в случае, когда она имеет допустимые решения. Поэтому такой вывод следует подтвердить получением глобального оптимума задачи: найти

$$\min \Delta(x, y, u) \text{ при ограничениях } (x, y) \in \hat{\Omega}, u \geq 0. \quad (13)$$

Это задача минимизации невыпуклой квадратичной функции на многограннике и может быть эффективно решена после сведения к задаче целочисленного программирования. Если оптимальное значение в (13) положительно, то действительно $\hat{\Psi} = \emptyset$ и элемент разбиения $\hat{\Omega}$ не включается в \mathcal{P} и далее не рассматривается. Если оптимальное значение равно нулю, то решение (\bar{x}, \bar{y}) задачи (13) даёт верхнюю оценку $\bar{F} \equiv F(\bar{x}, \bar{y})$.

6. Нижняя оценка

Для получения нижней оценки оптимального значения целевой функции верхнего уровня на множестве $\hat{\Psi}$ используется эллипсоид S , аппроксимирующий извне область $\hat{\Omega}$. Нижняя оценка есть оптимальное значение в задаче нахождения

$$\min F(x, y) \text{ при } (x, y) \in S, \quad (14)$$

которая разрешима полиномиально даже в случае невыпуклой целевой функции [8]. Пусть $(\underline{x}, \underline{y})$ — оптимальное решение задачи (14), а оптимальное значение равно $\underline{F}(\hat{\Psi}) \equiv F(\underline{x}, \underline{y})$. Тогда $F(x, y) \geq \underline{F}(\hat{\Psi})$ при всех $(x, y) \in \hat{\Psi}$.

Предложенная оценка обеспечивает сходимость метода. Однако она может оказаться очень плохой, поскольку не учитывает сам элемент разбиения индуктивной области $\hat{\Psi}$, а учитывает только содержащее его множество $\hat{\Omega}$. Между тем $\hat{\Psi}$ может быть любым фрагментом границы $\hat{\Omega}$. Чтобы улучшить оценку, в дополнение к основному разбиению введём ещё одно. Его цель — удалить из $\hat{\Omega}$ фрагмент, на границе которого нет точек из $\hat{\Psi}$, и тем самым увеличить «вес» $\hat{\Psi}$ в $\hat{\Omega}$. Тогда есть больше оснований судить о значениях целевой функции на $\hat{\Psi}$ по её значениям на $\hat{\Omega}$, и на таком элементе разбиения предложенная нижняя оценка оказывается более качественной.

Грани исходного слабого допустимого множества Ω задаются уравнениями $G_i(x, y) = 0$, $1 \leq i \leq P$, и $g_i(x, y) = 0$, $1 \leq i \leq p$. Элемент разбиения $\hat{\Omega}$ является многогранником, и его грани определяются этими уравнениями и уравнениями, которые задают разделяющие гиперплоскости. Предположим, что $\hat{\Omega}$ задаётся s линейными неравенствами и соответствующие уравнения будем записывать в виде $g_i(x, y) = 0$, $i = 1, \dots, s$.

Каждую грань множества $\hat{\Omega}$ попытаемся перенести так, чтобы она прошла через некоторую точку множества $\hat{\Psi}$. Для этого решим последовательно задачи: найти

$$\text{loc max } g_i(x, y) \text{ при ограничениях } \Delta(x, y, u) = 0, (x, y) \in \hat{\Omega}, u \geq 0 \quad (15)$$

при $i = 1, \dots, s$. В качестве стартовой точки локального подъёма целесообразно выбрать решение задачи $\text{max } g_i(x, y)$ при $(x, y) \in \hat{\Omega}$.

Если оптимальное значение g_i^* в задаче (15) для некоторого i отрицательно, то используем равенство $g_i(x, y) = g_i^*$ для дополнительного разбиения. Такая разбивающая гиперплоскость параллельна грани, заданной уравнением $g_i(x, y) = 0$, и проходит через точку множества $\hat{\Psi}$, в которой получено оптимальное решение в (15) (см. рис. 3 и 4). Если оптимальное значение равно нулю, то грань содержит точки множества $\hat{\Psi}$ и не используется при дополнительном разбиении.

Исследуем произведённое разбиение. Если в вспомогательной задаче (15) удалось достичь глобального оптимума, то множество

$$\hat{\Omega}^+ = \{x \in \hat{\Omega} \mid g_i(x, y) \geq g_i^*\}$$

содержит только те точки индуктивной области, которые уже включены в множество $\hat{\Omega}^- = \{x \in \hat{\Omega} \mid g_i(x, y) \leq g_i^*\}$. Значит, $\hat{\Omega}^+$ можно исключить из дальнейшего рассмотрения (рис. 3). Если же получена стационарная точка, не доставляющая глобального максимума, то множество $\hat{\Omega}^+$ может содержать часть $\hat{\Psi}$, не включённую в $\hat{\Omega}^-$ (рис. 4).

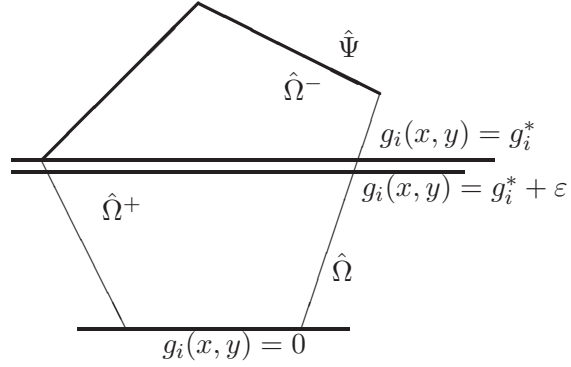


Рис. 3. Изменение элемента разбиения

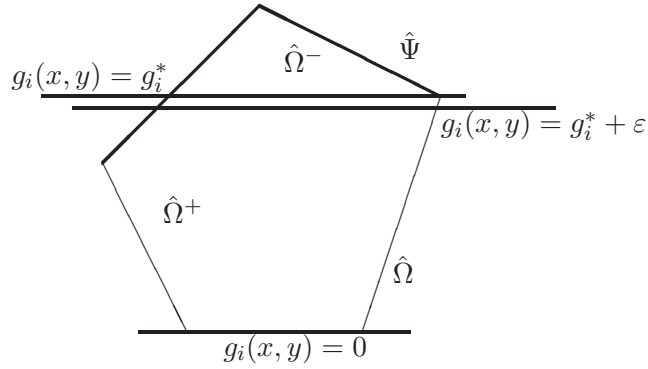


Рис. 4. Дополнительное разбиение

Чтобы удалить из дальнейшего рассмотрения неперспективный элемент разбиения и сохранить тот, который может содержать глобальное решение исходной задачи, необходимо выяснить, какая из этих двух ситуаций имеет место. Для этого решим задачу: найти

$$\begin{aligned} \text{loc min } F(x, y) \text{ при ограничениях } \Delta(x, y, u) = 0, \\ (x, y) \in \hat{\Omega}, g_i(x, y) \geq g_i^* + \varepsilon, u \geq 0 \end{aligned} \quad (16)$$

при некотором достаточно малом ε . Если процедура локального поиска не обнаруживает допустимых точек, то сделать точный вывод о том, что допустимая область пуста, поможет решение задачи: найти

$$\min \Delta(x, y, u) \text{ при условиях } (x, y) \in \hat{\Omega}, g_i(x, y) \geq g_i^* + \varepsilon, u \geq 0. \quad (17)$$

Если оптимальное значение в (17) положительно, то область $\hat{\Omega}^+$ является собственным подмножеством множества $\hat{\Omega}^-$ и далее рассматривается

(рис. 3). В этом случае элемент разбиения $\hat{\Omega}$ заменяем множеством $\hat{\Omega}^-$ и решаем задачу (15) для следующего i . Если оптимальное значение в (17) равно нулю или процедура локального поиска обнаружила допустимые точки в задаче (16), то оба элемента разбиения для $\hat{\Omega}^+$ и $\hat{\Omega}^-$ должны рассматриваться далее. Для этого следует заменить $\hat{\Omega}$ на множество $\hat{\Omega}^-$, в набор \mathcal{P} перспективных элементов разбиения добавить множество $\hat{\Omega}^+$ и не решать задачу (15) при других i .

Когда задача (15) решена для каждого $1 \leq i \leq s$, либо этот процесс остановлен на некотором i , $i < s$, то для преобразованного элемента разбиения $\hat{\Omega}$ можно построить внешний эллипсоид и найти $\underline{F}(\hat{\Omega})$. Если образовался дополнительный элемент разбиения, то для него повторяется описанная процедура и получаются нижняя и верхняя оценки.

Не исключено, что указанным способом не удаётся изменить элемент разбиения. Тогда нижняя оценка этим методом не может быть улучшена и дополнительное разбиение не используется.

Если полученная нижняя оценка не превосходит наименьшей верхней оценки среди множеств из \mathcal{P} : $\underline{F}(\hat{\Omega}) \leq \overline{F}$, то элемент разбиения $\hat{\Omega}$ включается в набор \mathcal{P} и подвергается новому разбиению в случае, когда $\underline{F}(\hat{\Omega})$ — наименьшая оценка по всем множествам из \mathcal{P} .

7. Обоснование сходимости

Заметим, что при осуществлении очередного разбиения формируются полуэллипсоиды S^1 и S^2 , содержащие новые элементы разбиения $\hat{\Psi}^1$ и $\hat{\Psi}^2$ соответственно (см. рис. 2). Для формирования внешней аппроксимации на следующем шаге такой полуэллипсоид достаточно погрузить в эллипсоид M наименьшего объёма. Оценки в этом случае будут хуже, а сходимость медленнее чем при аппроксимировании с помощью оптимальной линейной комбинации неравенств. Тем не менее уже такой вариант метода даёт сходимость к оптимальному решению.

Сопоставим множеству M его диаметр $\delta(M)$ (для эллипсоида он совпадает с длиной наибольшей оси). Элемент разбиения l -й по глубине в дереве предложенного алгоритма ветвей и границ обозначим через $\hat{\Psi}_l$, а через M_l соответствующий внешний эллипсоид, описанный вокруг полуэллипсоида. Рассмотрим бесконечную цепочку делений и аппроксимаций: исходный эллипсоид делится пополам ортогонально своей наибольшей оси, полученный полуэллипсоид погружается в эллипсоид наименьшего объёма, он, в свою очередь, делится пополам и т. д. После первых шагов такого процесса диаметр очередного эллипсоида будет больше диаметра исходного, однако через число шагов, равное размерности пространства,

диаметр уменьшается, а при бесконечном продолжении процесса стремится к нулю. Это утверждение формулируется следующим образом.

Лемма 1. При любом $l \geq 0$ справедливы неравенства

$$\delta(M_{l+(N+n)}) < \delta(M_l) \text{ и } \lim_{l \rightarrow \infty} \delta(M_l) = 0.$$

Другими словами, аппроксимирующий эллипсоид, а вместе с ним и содержащийся в нём элемент разбиения, «стягиваются в точку». Данное утверждение доказано в [2].

По построению верхняя и нижняя оценки оптимального значения целевой функции на элементе разбиения есть значения функции в некоторых точках внешнего эллипсоида. Поскольку диаметр эллипсоида стремится к нулю, разность оценок также стремится к нулю, что утверждается в следующей лемме.

$$\text{Лемма 2. } \lim_{l \rightarrow \infty} (\overline{F}(\hat{\Psi}_l) - \underline{F}(\hat{\Psi}_l)) = 0.$$

Процедура оценивания с таким свойством называется *состоятельной* [7, определение IV.4, замечание IV.3 (ii)].

Напомним, что очередному разбиению подвергается то из перспективных множеств $\hat{\Psi}$, для которого оценка $\underline{F}(\hat{\Psi})$ наименьшая. О подобной процедуре выбора подмножества для разбиения говорят, что она улучшает оценку [7, определение IV.6]. Согласно теории методов ветвей и границ сходимость процесса обеспечивается этими свойствами процедур.

Теорема 1 [7, теорема IV.3]. Если в бесконечном алгоритме ветвей и границ процедура оценивания состоятельна, а процедура выбора подмножества для разбиения улучшает оценку, то алгоритм сходится, т. е.

$$\lim_{l \rightarrow \infty} \underline{F}_l = \lim_{l \rightarrow \infty} \overline{F}_l = F(\bar{x}, \bar{y}) = F_*.$$

Здесь \underline{F}_l , \overline{F}_l — достигнутые к данному шагу наименьшие нижняя и верхняя оценки, F_* — оптимальное значение целевой функции в рассматриваемой задаче.

Поскольку в предлагаемом методе условия теоремы обеспечиваются, то он приводит к глобальному оптимуму.

Теорема 2. Последовательность точек, формируемая описанным методом ветвей и границ, сходится к точке глобального оптимума в задаче (5)–(8).

8. Численный пример

Рассмотрим работу метода на следующем простом примере. Найти

$$\min_x \{-0,5(x-3)^2 + 2(y-2)^2 - 5(x-3)(y-2)\}$$

при условиях

$$\min_y \{(y-1)^2 - 1,5xy\}$$

$$\begin{aligned} -3x + y &\leq -3, & x - 0,5y &\leq 4, \\ x + y &\leq 7, & y &\geq 0. \end{aligned}$$

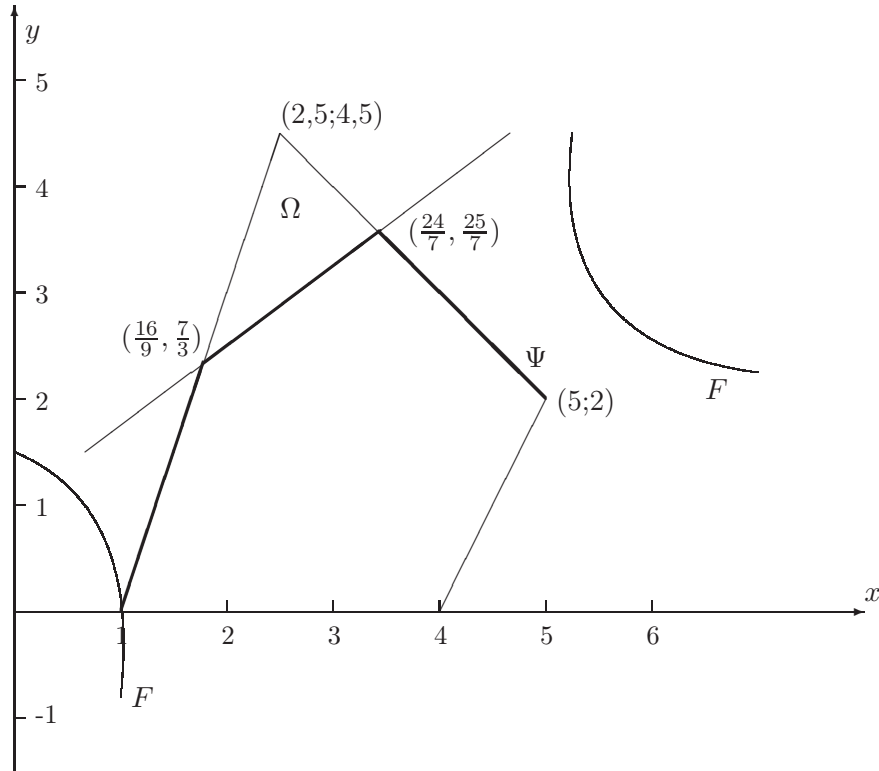


Рис. 5. Иллюстрация численного примера

Слабое допустимое множество, соответствующее этой задаче, показано на рис. 5, индуктивная область выделена жирной линией. Изображена также линия уровня целевой функции $F(x, y)$, минимум которой на индуктивной области достигается в точке $(1; 0)$ с оптимальным значением $F_* = -14$.

После первых двух итераций имеем следующий набор элементов разбиения слабой допустимой области (рис. 6): $\mathcal{W} = \{\hat{\Omega}_1, \hat{\Omega}_2, \hat{\Omega}_3, \hat{\Omega}_4\}$, где

$$\Omega_1 = \{(x, y) \mid -3x + y \leq -3; y \geq 0; y \leq 1,5; x \leq 3,08\},$$

$$\Omega_2 = \{(x, y) \mid -3x + y \leq -3; x + y \leq 7; y \geq 1,5; x \leq 3,08\},$$

$$\Omega_3 = \{(x, y) \mid x - 0,5y \leq 4; y \geq 0; y \leq 1,5; x \geq 3,08\},$$

$$\Omega_4 = \{(x, y) \mid x - 0,5y \leq 4; x + y \leq 7; y \geq 1,5; x \geq 3,08\}.$$

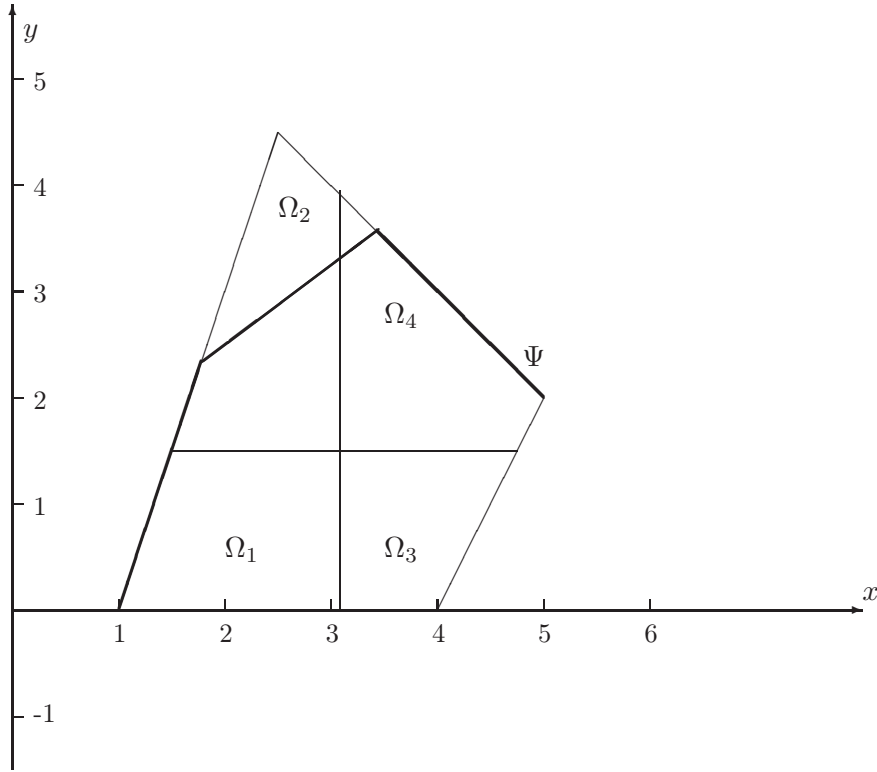


Рис. 6. Разбиение допустимой области в численном примере

Попытка решения задач (12) и (13) продемонстрировала, что элемент разбиения Ω_3 не содержит точек индуктивной области. Поэтому данный элемент был исключен из дальнейшего рассмотрения. Для улучшения нижних оценок были решены задачи (15), и области Ω_1 , Ω_2 и Ω_4 были преобразованы к виду (жирным шрифтом выделены изменившиеся неравенства)

$$\Omega_1 = \{(x, y) \mid -3x + y \leq -3; y \geq 0; y \leq 1,5; \mathbf{x} \leq \mathbf{1,51}\},$$

$$\Omega_2 = \{(x, y) \mid -3x + y \leq -3; \mathbf{x} + \mathbf{y} \leq \mathbf{6,4}; y \geq 1,5; x \leq 3,08\},$$

$$\Omega_4 = \{(x, y) \mid x - 0,5y \leq 4; x + y \leq 7; \mathbf{y} \geq \mathbf{2}; x \geq 3,08\}.$$

Полученное разбиение показано на рис. 7.

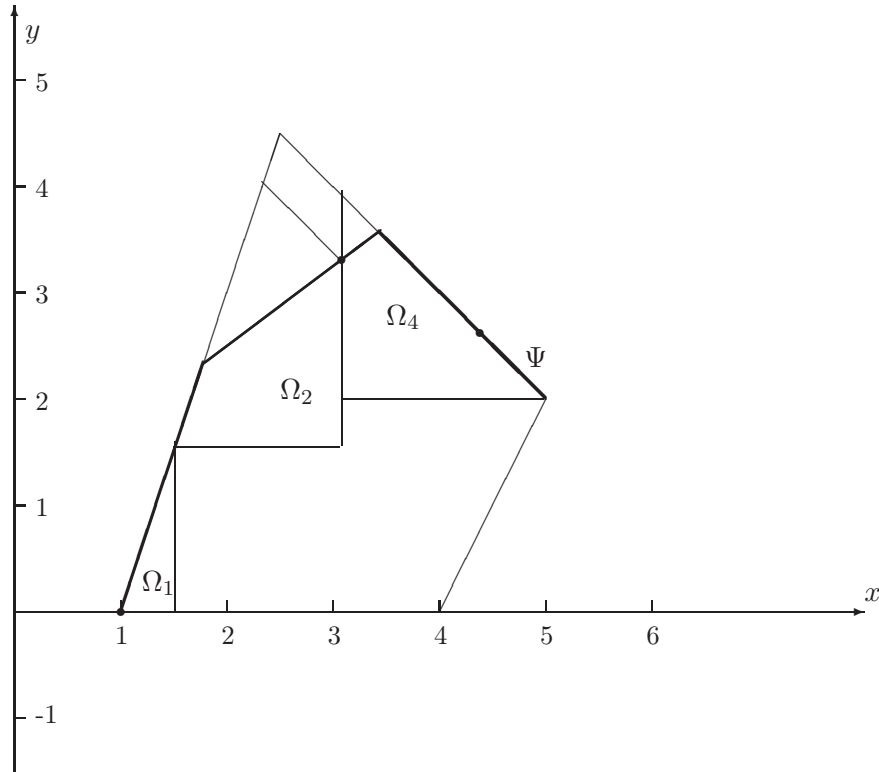


Рис. 7. Перспективные элементы разбиения допустимой области в численном примере

Перспективным элементам разбиения соответствуют следующие оценки оптимального значения целевой функции:

$$\overline{F}(\Psi_1) = -14; \overline{F}(\Psi_2) = 2,90; \overline{F}(\Psi_4) = -4,46;$$

$$\underline{F}(\Psi_1) = -42,5; \underline{F}(\Psi_2) = -4,77; \underline{F}(\Psi_4) = -7,7.$$

Верхние оценки достигаются в точках $(1; 0)$, $(3,08; 3,31)$ и $(4,38; 2,62)$, которые также показаны на рис. 7. Для следующего разбиения была выбрана область Ω_1 и процесс продолжался до достижения решения $(1; 0)$ с заданной точностью. Расчёты выполнялись в системе автоматического моделирования GAMS [5].

ЛИТЕРАТУРА

1. **Анциферов Е. Г.** К методу эллипсоидов в выпуклом программировании // Численные методы анализа и их приложения. Иркутск: СЭИ СО АН СССР, 1987. С. 3–12.
2. **Нечаева М. С., Хамисов О. В.** Метод ветвей и границ для задачи минимизации невыпуклой квадратичной функции при выпуклых квадратичных ограничениях // Дискрет. анализ и исслед. операций. Сер.2. 2000. Т. 7, № 2. С. 74–88.
3. **Bard J. F.** Practical bilevel optimization. Dordrecht: Kluwer Academic Publishers, 1998.
4. **Bazaraa M. S., Shetti C. M.** Nonlinear programming. Theory and algorithms. New York: John Wiley & Sons, Inc., 1993.
5. **Brooke A., Kendrick D., Meeraus A.** GAMS. Release 2.25. A User's Guide // GAMS Development Corporation, 1996.
6. **Dempe S.** Foundations of bilevel programming. Dordrecht: Kluwer Academic Publishers, 2002.
7. **Horst R., Tuy H.** Global optimization. Deterministic approaches. Berlin: Springer-Verlag, 1993.
8. **Ye Y.** On the affine scaling algorithm for nonconvex quadratic programming // Mathematical Programming. 1992. V. 56, N 2. P. 285–300.

Адрес автора:

Институт систем энергетики им.
Л.А.Мелентьева СО РАН,
ул.Лермонтова, 130,
664033 Иркутск, Россия.

Статья поступила

26 октября 2004 г.

Переработанный вариант —

13 ноября 2005 г.