ТРУДЫ ИНСТИТУТА МАТЕМАТИКИ И МЕХАНИКИ УрО РАН

Том 27 № 4 2021

УДК 519.658.4

ОБ ОДНОМ ПРИЕМЕ ПОВЫШЕНИЯ ГЛАДКОСТИ ВНЕШНИХ ШТРАФНЫХ ФУНКЦИЙ В ЛИНЕЙНОМ И ВЫПУКЛОМ ПРОГРАММИРОВАНИИ[1]

Л. Д. Попов

Предложены оригинальные конструкции внешних штрафных функций в линейном и выпуклом программировании, асимптотически сводящие задачи условной оптимизации к задачам безусловной оптимизации повышенной гладкости. Последние допускают эффективное решение методами второго порядка и в то же время не нуждаются в знании хотя бы одной внутренней допустимой точки исходной задачи. Более того, новые штрафные функции могут быть применены и к несобственным задачам линейного и выпуклого программирования (задачам с противоречивыми системами ограничений), для которых они способны вырабатывать некоторые обобщенные (компромиссные) решения. Приведены теоремы сходимости и данные численных экспериментов.

Ключевые слова: линейное программирование, несобственные задачи, обобщенные решения, метод штрафных функций, метод Ньютона.

L. D. Popov. On one method of increasing the smoothness of external penalty functions in linear and convex programming.

We propose original constructions of external penalty functions in linear and convex programming, which asymptotically reduce constrained optimization problems to unconstrained ones with increased smoothness. The latter admit an effective solution by second-order methods and, at the same time, do not require the knowledge of an interior feasible point of the original problem to start the process. Moreover, the proposed approach is applicable to improper linear and convex programs (problems with contradictory constraint systems), for which they can generate some generalized (compromise) solutions. Convergence theorems and the data of numerical experiments are presented.

Keywords: linear programming, improper (ill-posed) problems, generalized solutions, penalty functions, Newton method.

MSC: 47N05, 37N25, 37N40

DOI: 10.21538/0134-4889-2021-27-4-88-101

Введение

Штрафные функции представляют собой эффективный и широко применяемый инструмент построения приближенных итеративных алгоритмов решения задач условной оптимизации. Исследования на эту тему появились в научной литературе достаточно давно и до сих пор актуальны (см. многочисленные монографии и учебники [1–8]). В их основе лежит идея замещения исходной условно-экстремальной задачи задачей поиска безусловного экстремума некоторой вспомогательной функции, составленной из целевой функции исходной задачи и некоторого блока штрафных слагаемых, в той или иной мере ограничивающих выход переменных задачи за рамки исходных ограничений. Три наиболее известных класса такого рода формируют методы точных штрафных функций [1], методы внешних штрафных функций (см. [2–5]) и методы внутренних (барьерных) штрафных функций (см. [6–8]), тесно связанных с современными методами внутренней точки [9].

[1]Данное исследование выполнено в Уральском математическом центре при финансовой поддержке Министерства науки и высшего образования Российской Федерации (номер проекта 075-02-2021-1383).

Методы первой группы интересны тем, что дают решение исходной задачи уже при конечных (хотя и больших) значениях штрафного параметра. Платой за это выступает существенная негладкость вспомогательных подзадач безусловной оптимизации, что резко сужает возможности применения для их решения эффективных вычислительных алгоритмов. Внутренние (или барьерные) штрафные функции требуют неограниченного роста штрафного параметра и позволяют получать лишь приближенное решение исходной задачи (в асимптотике). Кроме того, для запуска этих алгоритмов необходимо изначально знать хотя бы одну внутреннюю точку допустимой области решаемой задачи. Нахождение таких точек обычно требует своего отдельного вычислительного процесса. Вместе с тем барьерные функции имеют очень высокие показатели гладкости, что позволяет использовать для их оптимизации высокоэффективные методы второго порядка. Широкое распространение компьютерных систем с большими объемами оперативной памяти способствовало росту интереса к таким алгоритмам.

Промежуточное положение по качеству гладкости вспомогательных оптимизационных подзадач занимают так называемые внешние штрафные функции, среди которых наиболее популярными являются квадратичные штрафы за отклонение текущего приближения от допустимой области задачи. Из-за пониженных свойств гладкости методы внешних штрафных функций ориентированы на применение методов оптимизации лишь первого порядка. Однако они не вызывают проблем, связанных с поиском стартовой точки вычислительного процесса. Более того, они обладают важным свойством, благодаря которому могут с успехом применяться не только к разрешимым задачам, но и к задачам с такой особенностью, как несовместность системы исходных ограничений.

Задачи условной оптимизации с противоречивыми ограничениями (называемые в научной литературе *несобственными* [10–12]) возникают достаточно часто как вследствие ошибок в собранных исходных данных математической модели, так и из-за наличия реальных внутренних противоречий у моделируемого объекта, скрытых на ранних ступенях его изучения. Здесь было бы удобнее, чтобы применяемый к таким задачам алгоритм не просто констатировал сам факт неразрешимости задачи, но и выяснял бы природу этой противоречивости, определял "узкие" места модели и давал рекомендации по их развязке и объему необходимых для этого ресурсов. Именно этими качествами обладают методы внешних штрафов. В данной работе автор предпринял очередную попытку так сконструировать новые типы штрафных функций, чтобы совместить перечисленные преимущества алгоритмов различных типов. Предлагаемое исследование является развитием более ранних работ [13–15].

1. Общее описание предлагаемого подхода

Общее описание предлагаемого подхода представим на примере задачи выпуклого программирования, которую запишем в виде

$$\min\{f_0(x)\colon f_j(x) \le 0,\ j = 1, \ldots, m\}; \tag{1.1}$$

здесь числовые функции $f_j\colon \mathbb{R}^n \to \mathbb{R},\ j = 0, 1, \ldots, m$ выпуклы и дважды непрерывно дифференцируемы, x — вектор переменных (неизвестных), значения которых надо определить.

Пусть задача (1.1) разрешима, а ее оптимальное множество ограничено.

Соотнесем с задачей (1.1) выпуклую штрафную функцию смешанного типа

$$F^{\varepsilon}(x, u) = f_0(x) - \varepsilon_2 \sum_{i=1}^{m} \ln(u_i - f_i(x)) + \frac{1}{2\varepsilon_1} \|u\|^2, \quad \varepsilon = (\varepsilon_1, \varepsilon_2) > 0. \tag{1.2}$$

Здесь к исходным переменным x добавлены дополнительные переменные u, отвечающие за некоторое ослабление первоначальных ограничений с тем, чтобы логарифмическая барьерная функция могла вобрать в себя допустимую область исходной задачи целиком, вместе с ее границей (и даже шире). Дополнительное штрафное слагаемое с нормой этого вектора отвечает за минимизацию вносимых в задачу изменений.

С помощью функции (1.2) внутри выпуклой открытой области

$$\Omega = \big\{(x,u)\colon x \in \mathbb{R}^n,\ u \in \mathbb{R}^m,\ u_i - f_i(x) > 0\ \ (i = 1,\dots,m)\big\}$$

построим вспомогательную задачу минимизации (фактически безусловной): найти

$$F^\varepsilon(\hat{x}_\varepsilon, \hat{u}_\varepsilon) = \min_{x,u\in\Omega} F^\varepsilon(x,u). \tag{1.3}$$

При сделанных предположениях функция (1.2) по крайней мере дважды непрерывно дифференцируема и имеет однозначно определяемые точки глобального минимума при всех $\varepsilon = (\varepsilon_1, \varepsilon_2) > 0$. Эти точки обязательно будут внутренними для области Ω. Последнее обстоятельство оправдывает запись задачи (1.3) как задачи безусловной оптимизации.

Ввиду выпуклости всех введенных конструкций и условий их гладкости необходимые и достаточные условия того, что точка $(\bar{x}, \bar{u})$ будет точкой глобального минимума функции (1.2) по совокупности переменных, можно записать в виде системы двух нелинейных уравнений

$$\nabla_u F^\varepsilon(\bar{x}, \bar{u}) = 0, \quad \nabla_x F^\varepsilon(\bar{x}, \bar{u}) = 0.$$

Последние, в частности, можно решать при помощи метода Ньютона.

Связь задачи (1.3) с поиском решения задачи (1.1) описывается следующим утверждением (его обоснование может быть проведено по схемам, примененным автором в предшествующих работах, посвященных штрафным функциям смешанного типа, см. [13–15]).

Утверждение 1. *Пусть допустимое (а значит, и оптимальное) множество задачи* (1.1) *ограничено. Тогда для любой пары параметров $\varepsilon = (\varepsilon_1, \varepsilon_2) > 0$ существует единственная пара векторов $\hat{x}_\varepsilon \in \mathbb{R}^n$, $\hat{u}_\varepsilon \in \mathbb{R}^n$, минимизирующая функцию* (1.2) *и решающая таким образом задачу* (1.3). *При этом стремление штрафных параметров к нулю гарантирует сходимость возникающей последовательности $\hat{x}_\varepsilon$ к оптимальному множеству задачи* (1.1).

Таким образом, решая задачу (гладкой) безусловной минимизации функции (1.3) при достаточно малых значениях параметров $\varepsilon_1 > 0$ и $\varepsilon_2 > 0$, можно получить решение исходной задачи со сколь угодно высокой точностью. Более детальные алгоритмы отыскания вспомогательного минимума и их дополнительные возможности обсудим ниже уже на примере задач линейного программирования.

2. Особенности линейной постановки

Рассмотрим теперь задачу линейного программирования вида

$$\max\{(c,x)\colon Ax \le b\}; \tag{2.1}$$

здесь числовая матрица $A = (a_{ij})_{m\times n}$ и векторы c и b заданы; x — вектор прямых переменных (неизвестных), значения которых надо определить; круглые скобки используются для обозначения скалярного произведения; $m > n = \operatorname{rank} A$.

Договоримся, что задача (2.1) может быть как разрешимой, так и несобственной (когда ее ограничения противоречивы, см. [3]). В последнем случае будем предполагать, что она может быть приведена к собственному (разрешимому) виду путем коррекции правых частей своих ограничений. Иными словами, будем предполагать совместность ограничений двойственной задачи

$$\min\{(b,y)\colon A^\top y = c,\ y \ge 0\} \tag{2.2}$$

и, даже более того, телесность ее допустимого множества.

Сразу оговорим, что именно в дальнейшем будет пониматься под *обобщенным решением* задачи (2.1) в случае ее несобственности. Для этого погрузим исходную задачу в параметрическое семейство задач вида

$$\max\{(c,x)\colon Ax \le b+u\} \quad \big(=:\operatorname{opt}(u)\big), \tag{2.3}$$

где $u \in \mathbb{R}^m$ — векторный параметр коррекции правых частей ограничений исходной задачи.

Введем обозначение $M(u) = \{x\colon Ax \le b+u\}$ и определим, например, вектор оптимальной коррекции по правилу

$$u_0 = \arg\min\{\|u\|\colon M(u) \ne \varnothing\},$$

где $\|\cdot\|$ — евклидова норма вектора. В силу свойств метрической проекции на выпуклое замкнутое множество такой вектор всегда существует и является единственным.

Определим обобщенное (аппроксимационное) решение несобственной задачи (2.1) как обычное решение уже разрешимой задачи

$$\max\{(c,x)\colon Ax \le b+u_0\}. \tag{2.4}$$

Телесность допустимой области задачи (2.2) обеспечивает не только разрешимость задачи (2.4), но и ограниченность ее оптимального множества.

Подчеркнем, что в случае разрешимости исходной задачи имеем $u_0 = 0$, и введенное выше обобщенное решение совпадает с ее обычным решением.

В завершение постановочной части укажем на альтернативное представление области

$$M(u_0) = \operatorname{Arg}\min_x \frac{1}{2}\|(Ax-b)_+\|^2,$$

где $(a)_+ = \max\{a,0\}$ — оператор положительной срезки (числа, вектора). Это представление является связующим звеном между обсуждаемой проблематикой и теорией штрафных функций (см. [11; 12]).

3. Штрафные конструкции для линейной задачи

Следуя вводным установкам из разд. 1, поставим в соответствие задаче (2.1) комбинированную штрафную функцию смешанного типа (теперь из-за специфики задачи вогнутую)

$$F^\varepsilon(x,u) = (c,x) + \varepsilon_2 \sum_{i=1}^{n} \ln\big(u_i + b_i - (a_i,x)\big) - \frac{1}{2\varepsilon_1}\|u\|^2, \quad \varepsilon = (\varepsilon_1,\varepsilon_2) > 0, \tag{3.1}$$

и свяжем с ней вспомогательную задачу поиска глобального максимума (фактически безусловного)

$$F^\varepsilon(\hat{x}_\varepsilon, \hat{u}_\varepsilon) = \max_{x,u} F^\varepsilon(x,u). \tag{3.2}$$

Очевидно, что при сделанных выше предположениях функция (3.1) имеет однозначно определяемые точки глобального максимума при всех $\varepsilon = (\varepsilon_1,\varepsilon_2) > 0$. Условие $u \ge 0$ будет выполняться в этих точках автоматически.

Связь задачи (3.2) с поиском обобщенного решения задачи (2.1) описывается серией следующих утверждений, доказательство которых легко провести по схемам, предложенным автором в более ранних публикациях по аналогичным конструкциям для несколько более специфических постановок (см. [13–15]).

Утверждение 2. *Пусть оптимальное множество задачи* (2.4) *ограничено. Тогда для любой пары параметров* $\varepsilon = (\varepsilon_1,\varepsilon_2) > 0$ *существует единственная пара векторов* $\hat{x}_\varepsilon \in \mathbb{R}^n$, $\hat{u}_\varepsilon \in \mathbb{R}^n$, *максимизирующая функцию* (3.1) *и решающая, таким образом, задачу* (3.2).

Утверждение 3. *Пусть допустимая область задачи* (2.2) *телесна и параметры в* (3.1) *стеснены условием* $0 < \varepsilon_1 < \bar{\varepsilon}_1$, $0 < \varepsilon_2 < \bar{\varepsilon}_2$. *Тогда векторы* $\hat{u}_\varepsilon$ *ограничены в совокупности.*

Утверждение 4. *Вектор* $\hat{x}_\varepsilon$ *является допустимым для задачи* (2.3) *при* $u = \hat{u}_\varepsilon$, *и для него выполнено неравенство*

$$\operatorname{opt}(\hat{u}_\varepsilon) - m\varepsilon_1 \le (c, \hat{x}_\varepsilon) \le \operatorname{opt}(\hat{u}_\varepsilon).$$

Следствие. *Пусть выполнены предположения утверждения* 3. *Тогда совокупность векторов* $\hat{x}_\varepsilon$ *также ограничена.*

Утверждение 5. *Пусть допустимая область задачи* (2.2) *телесна. Тогда*

$$\hat{u}_\varepsilon \to u_0, \quad (c, \hat{x}_\varepsilon) \to \operatorname{opt}(u_0)$$

при $0 < \varepsilon = (\varepsilon_1, \varepsilon_2) \to +0$.

Таким образом, решая задачу (гладкой) безусловной максимизации вспомогательной функции (3.1) при достаточно малых значениях параметров $\varepsilon_1 > 0$ и $\varepsilon_2 > 0$, можно получить аппроксимационное решение исходной задачи со сколь угодно высокой точностью.

4. Метод Ньютона для линейного случая

Положим ниже для краткости $e = (1, 1, \ldots, 1)$. Выпишем необходимые и достаточные условия того, что пара (x, u) является точкой максимума функции (3.1):

$$\nabla_x F^\varepsilon(x, u) = c - \varepsilon_2 A^\top \operatorname{diag}(u + b - Ax)^{-1} e = 0, \tag{4.1}$$

$$\nabla_u F^\varepsilon(x, u) = \varepsilon_2 \operatorname{diag}(u + b - Ax)^{-1} e - \frac{1}{\varepsilon_1} u = 0. \tag{4.2}$$

Представим полученную систему уравнений в нескольких, более удобных для дальнейшего анализа, видах.

Для начала введем вспомогательную переменную для $w = u + b - Ax > 0$. Получим эквивалентную перезапись соотношений (4.1), (4.2):

$$w - u + Ax - b = 0, \quad w > 0, \tag{4.3}$$

$$\varepsilon_2 A^\top \operatorname{diag}(w)^{-1} e - c = 0, \tag{4.4}$$

$$\varepsilon_2 \operatorname{diag}(w)^{-1} e - \frac{1}{\varepsilon_1} u = 0.$$

Из последнего соотношения выразим вектор $u = \varepsilon_1 \varepsilon_2 \operatorname{diag}(w)^{-1} e$ и исключим его из соотношений (4.3)–(4.4):

$$w - \varepsilon_1 \varepsilon_2 \operatorname{diag}(w)^{-1} e + Ax - b = 0, \tag{4.5}$$

$$\varepsilon_2 A^\top \operatorname{diag}(w)^{-1} e - c = 0. \tag{4.6}$$

Введя еще одну дополнительную переменную $y = \varepsilon_2 \operatorname{diag}(w)^{-1} e > 0$, преобразуем последние соотношения к виду, наиболее близко напоминающему соотношения метода центрального пути:

$$w - \varepsilon_1 y + Ax - b = 0, \quad w, y > 0, \tag{4.7}$$

$$\operatorname{diag}(y) \operatorname{diag}(w) e - \varepsilon_2 e = 0, \tag{4.8}$$

$$A^\top y - c = 0. \tag{4.9}$$

Наконец, введем вспомогательное отображение $\Phi(w,y,x)$ пространства $\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n$ в себя по правилу

$$\Phi(w,y,x) = \begin{pmatrix} w - \varepsilon_1\, y + Ax - b \\ \operatorname{diag}(y)\operatorname{diag}(w)\, e - \varepsilon_2\, e \\ A^\top y - c \end{pmatrix}.$$

С его помощью перепишем систему (4.7)–(4.9) как

$$\Phi(w,y,x) = 0. \tag{4.10}$$

Применим для решения системы (4.10) метод Ньютона. Последний заключается в построении последовательности приближений $z^k = (w^k, y^k, x^k)$ к решению этой системы по рекуррентным формулам

$$z^{k+1} = z^k - \alpha_k\, p^k, \quad \text{где} \quad p^k = \bigl(\nabla\,\Phi(z^k)\bigr)^{-1}\Phi(z^k), \quad k = 0,1,2,\ldots\,. \tag{4.11}$$

В начальном приближении $w^0,\ y^0$ положительные. Параметр шага $\alpha_k = 1$ обычно постоянен.

Заметим, что параметры $\varepsilon_1 > 0,\ \varepsilon_2 > 0$ также можно менять от итерации к итерации, постепенно сводя их значения к нулю и следя за тем, чтобы такое сведение не нарушало сходимости метода Ньютона и позволяло удерживать длину его шага на постоянном уровне.

Обсудим сложность реализации одной итерации метода (4.11). Основные усилия заключены в отыскании решения p^k системы уравнений

$$\nabla\,\Phi(z^k)\, p^k = \Phi(z^k),$$

или, в детальном виде, системы

$$\begin{pmatrix} E & -\varepsilon_1 E & A \\ Y & W & 0 \\ 0 & A^\top & 0 \end{pmatrix} \begin{pmatrix} p_w \\ p_y \\ p_x \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}, \tag{4.12}$$

где E — единичная матрица, $Y = \operatorname{diag}(y^k)$, $W = \operatorname{diag}(w^k)$, $p^k = (p_w, p_y, p_x)$. Перед нами блочная система линейных уравнений. Подвергнем ее ряду элементарных преобразований. Начнем с того, что умножим матрицы первой группы уравнений слева на Y и вычтем полученный результат из второй группы уравнений:

$$\begin{pmatrix} E & -\varepsilon_1\, E & A \\ 0 & W + \varepsilon_1\, Y & -Y\, A \\ 0 & A^\top & 0 \end{pmatrix} \begin{pmatrix} p_w \\ p_y \\ p_x \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 - Y\, h_1 \\ h_3 \end{pmatrix}.$$

Далее умножим коэффициенты второй группы уравнений слева на матрицу $H = (W + \varepsilon_1 Y)^{-1}$:

$$\begin{pmatrix} E & -\varepsilon_1\, E & A \\ 0 & E & -H\, Y\, A \\ 0 & A^\top & 0 \end{pmatrix} \begin{pmatrix} p_w \\ p_y \\ p_x \end{pmatrix} = \begin{pmatrix} h_1 \\ H\,(h_2 - Y\, h_1) \\ h_3 \end{pmatrix}.$$

Наконец, домножим коэффициенты второй группы уравнений слева на матрицу $A^\top$ и почленно вычтем их из уравнений третьей группы:

$$\begin{pmatrix} E & -\varepsilon_1\, E & A \\ 0 & E & -H\, Y\, A \\ 0 & 0 & A^\top H\, Y\, A \end{pmatrix} \begin{pmatrix} p_w \\ p_y \\ p_x \end{pmatrix} = \begin{pmatrix} h_1 \\ H\,(h_2 - Y\, h_1) \\ h_3 - A^\top H\,(h_2 - Y\, h_1) \end{pmatrix}.$$

Отсюда вытекает

Утверждение 6. *Решение системы* (4.12) *может быть последовательно получено по формулам*

$$p_x = (A^\top H\, Y\, A)^{-1}\big[h_3 - A^\top H\,(h_2 - Y\,h_1)\big],$$
$$p_y = H\,(h_2 - Y\,h_1) + H\,Y\,A\,p_x,$$
$$p_w = h_1 + \varepsilon_1\,p_y - A\,p_x,$$

где $H = (W + \varepsilon_1\,Y)^{-1}$.

Мы видим, что основной объем вычислений связан с обращением положительно определенной симметрической матрицы $Q = A^\top H\,Y\,A = A^\top (W + \varepsilon_1\,Y)^{-1} Y\,A$ размерности $n \times n$. Более того, вместо вычисления и обращения этой матрицы для решения последней из соответствующих линейных подсистем

$$(A^\top H\,Y\,A)\,p_x = h_3 - A^\top H\,(h_2 - Y\,h_1)$$

можно воспользоваться методом сопряженных направлений. Это позволит эффективно учесть также возможную разреженность исходных матриц.

В заключение раздела отметим, что y-компонента решения операторного уравнения (4.10) по мере уменьшения параметра $\varepsilon_{1,2} \to +0$ сходится к решению задачи, двойственной к исходной. Это с очевидностью вытекает из приведенных формул и утверждений.

5. Альтернативное решение вспомогательной системы

Вернемся к системе (4.5), (4.6)

$$w - \varepsilon_1\,\varepsilon_2\,\mathrm{diag}(w)^{-1} e + Ax - b = 0,$$
$$\varepsilon_2\,A^\top \mathrm{diag}(w)^{-1} e - c = 0.$$

Заметим, что первое уравнение определяет w как неявную функцию $w = w(x)$ переменной x. Фактически перед нами серия обыкновенных квадратных уравнений относительно компонент w_i. Они получаются, если обе части первого уравнения умножить слева на матрицу $\mathrm{diag}(w)$:

$$w_i^2 + w_i\big((a_i, x) - b_i\big) - \varepsilon_1\,\varepsilon_2 = 0, \quad i = 1, 2, \ldots, m;$$

здесь a_i — i-я строка матрицы A. Выбирая положительные корни этих уравнений, получаем аналитические зависимости

$$w_i(x) = \frac{b_i - (a_i, x) + \sqrt{[b_i - (a_i, x))]^2 + 4\,\varepsilon_1\,\varepsilon_2}}{2}, \quad i = 1, 2, \ldots, m, \tag{5.1}$$

и

$$\frac{\partial w_i(x)}{\partial x_j} = -\frac{a_{ij} w_i(x)}{2\,w_i(x) + (a_i, x) - b_i}, \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n. \tag{5.2}$$

Теперь можно посмотреть на второе соотношение исходной системы

$$\Psi(w(x)) = \varepsilon_2\,A^\top \mathrm{diag}(w(x))^{-1}\,e - c = 0 \tag{5.3}$$

как на неявное уравнение для отыскания x. В нем зависимости $w_i = w_i(x)$ представлены соотношениями (5.1), (5.2).

Чтобы применить для решения (5.3) метод Ньютона в пространстве переменных x, нам надо найти якобиан этой системы по x. Имеем

$$\Psi_j(w(x)) = \varepsilon_2 \sum_{s=1}^{m} \frac{a_{sj}}{w_s(x)} - c_j, \quad j = 1, 2, \ldots, n,$$

и, значит,

$$\frac{\partial \Psi_j(w(x))}{\partial x_i} = -\varepsilon_2 \sum_{s=1}^{m} \frac{a_{sj}}{w_s(x)^2} \frac{\partial w_s(x)}{\partial x_i} = \varepsilon_2 \sum_{s=1}^{m} \frac{a_{sj} a_{si}}{w_s(x)[2\, w_s + (a_s, x) - b_s]},$$

$$i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n.$$

Сам алгоритм работает в пространстве переменных x и описывается все теми же формулами

$$x^{k+1} = x^k - \alpha_k \big(\nabla \Psi(w(x^k))\big)^{-1} \Psi(w(x^k)), \quad k = 0, 1, 2, \ldots .$$

Снижение размерности пространства переменных сопровождается здесь повышенной сложностью расчета показателей неявной функции $w = w(x)$ и ее производных, так что вопрос об эффективности данного подхода нуждается в дополнительной проработке.

6. Каноническая задача линейного программирования

В виду прикладной важности рассмотрим дополнительно каноническую запись задачи линейного программирования

$$\min\{(c, x) \colon Ax = b,\ x \geq 0\};$$

здесь числовая матрица $A = (a_{ij})_{m \times n}$ и векторы c и b заданы; x — вектор прямых переменных (неизвестных), значения которых надо определить; круглые скобки используются для обозначения скалярного произведения; $n > m = \operatorname{rank} A$.

Для канонической задачи предлагаемая нами комбинированная штрафная функция примет вид

$$F^{\varepsilon}(x, u) = (c, x) + \frac{1}{2\,\varepsilon_1} \|u\|^2 - \varepsilon_2 \sum_{i=1}^{n} \ln\big(u_i + x_i\big), \quad \varepsilon = (\varepsilon_1,\, \varepsilon_2) > 0,$$

а порождаемая ею вспомогательная задача поиска ее глобального минимума будет уже включать в себя ограничения-уравнения: найти

$$F^{\varepsilon}(\hat{x}_{\varepsilon}, \hat{u}_{\varepsilon}) = \min_{u} \min_{x \colon Ax = b} F^{\varepsilon}(x, u). \tag{6.1}$$

Выпишем классические необходимые и достаточные условия (условия Лагранжа) того, что пара (x, u) является решением задачи (6.1):

$$\nabla_x F^{\varepsilon}(x, u) = c - \varepsilon_2 \operatorname{diag}(u + x)^{-1}\, e = A^{\top} y, \tag{6.2}$$

$$\nabla_u F^{\varepsilon}(x, u) = \frac{1}{\varepsilon_1}\, u - \varepsilon_2 \operatorname{diag}(u + x)^{-1} e = 0, \tag{6.3}$$

$$Ax = b. \tag{6.4}$$

Здесь $y = (y_1, \ldots, y_m)$ — вектор множителей Лагранжа, $e = (1, 1, \ldots, 1)$.

Представим полученную систему уравнений в нескольких более удобных для дальнейшего анализа видах.

Для начала введем вспомогательную переменную для $w = u + x > 0$. Получим эквивалентную перезапись соотношений (6.2)–(6.4)

$$w - u - x = 0, \quad w > 0, \tag{6.5}$$

$$A^{\top} y + \varepsilon_2 \operatorname{diag}(w)^{-1}\, e - c = 0,$$

$$Ax - b = 0,$$

$$\frac{1}{\varepsilon_1}\,u - \varepsilon_2\,\mathrm{diag}(w)^{-1}\,e = 0. \tag{6.6}$$

Из последнего соотношения выразим вектор $u = \varepsilon_1\,\varepsilon_2\,\mathrm{diag}(w)^{-1}\,e$ и исключим его из соотношений (6.5), (6.6):

$$w - \varepsilon_1\,\varepsilon_2\,\mathrm{diag}(w)^{-1}\,e - x = 0,$$

$$A^\top y + \varepsilon_2\,\mathrm{diag}(w)^{-1}\,e - c = 0,$$

$$Ax - b = 0.$$

Введем еще одну дополнительную переменную $v = \varepsilon_2\,\mathrm{diag}(w)^{-1}\,e > 0$ и преобразуем последние соотношения к виду, наиболее близко напоминающему соотношения метода центрального пути:

$$\varepsilon_1\,v + x - w = 0, \tag{6.7}$$

$$A^\top y + v - c = 0, \tag{6.8}$$

$$\mathrm{diag}(v)\,\mathrm{diag}(w)e - \varepsilon_2\,e = 0, \tag{6.9}$$

$$Ax - b = 0. \tag{6.10}$$

Вновь объединим соотношения последней системы уравнений в единое целое при помощи вспомогательного отображения

$$\Phi(w,v,x,y) = \begin{pmatrix} \varepsilon_1\,v + x - w \\ A^\top y + v - c \\ V\,W\,e - \varepsilon_2\,e \\ Ax - b \end{pmatrix}, \tag{6.11}$$

которое переводит пространство $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ в себя. Здесь для краткости введены обозначения $V = \mathrm{diag}(v),\ W = \mathrm{diag}(w)$.

С помощью отображения (6.11) препишем систему (6.7)–(6.10) как

$$\Phi(w,v,x,y) = 0.$$

Применим для решения системы (4.10) метод Ньютона. Как уже многократно отмечалось, последний заключается в построении последовательности приближений $z^k = (w^k, v^k, x^k, y^k)$ к решению этой системы по рекуррентным соотношениям

$$z^{k+1} = z^k - \alpha_k\,p^k, \quad \text{где} \quad p^k = \bigl(\nabla\,\Phi(z^k)\bigr)^{-1}\,\Phi(z^k), \quad k = 0,1,2,\ldots\,. \tag{6.12}$$

В начальном приближении надо брать положительные $w^0,\ v^0$. Обычно $\alpha_k = 1$.

Обсудим сложность реализации одной итерации метода (6.12).

Как всегда, основные усилия заключены в отыскании решения p^k системы уравнений

$$\nabla\,\Phi(z^k)\,p^k = \Phi(z^k),$$

или, в детальном виде, системы

$$\begin{pmatrix} E & -\varepsilon_1\,E & -E & 0 \\ 0 & E & 0 & A^\top \\ V & W & 0 & 0 \\ 0 & 0 & A & 0 \end{pmatrix} \begin{pmatrix} p_w \\ p_v \\ p_x \\ p_y \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{pmatrix}, \tag{6.13}$$

где снова для краткости обозначено $V = \mathrm{diag}(v^k)$, $W = \mathrm{diag}(w^k)$, $p^k = (p_w, p_v, p_x, p_y)$.

Перед нами блочная система линейных уравнений. Подвергнем ее ряду элементарных преобразований. Начнем с того, что умножим матрицы первой группы уравнений слева на V и вычтем полученный результат из третьей группы уравнений:

$$\begin{pmatrix} E & -\varepsilon_1 E & -E & 0 \\ 0 & E & 0 & A^\top \\ 0 & W+\varepsilon_1 V & V & 0 \\ 0 & 0 & A & 0 \end{pmatrix} \begin{pmatrix} p_w \\ p_v \\ p_x \\ p_y \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ h_3 - V h_1 \\ h_4 \end{pmatrix}.$$

Далее умножим коэффициенты второй группы уравнений слева на матрицу $M = W + \varepsilon_1 V$ и также почленно вычтем из уравнений третьей группы:

$$\begin{pmatrix} E & -\varepsilon_1 E & -E & 0 \\ 0 & E & 0 & A^\top \\ 0 & 0 & V & -M A^\top \\ 0 & 0 & A & 0 \end{pmatrix} \begin{pmatrix} p_w \\ p_v \\ p_x \\ p_y \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ h_3 - V h_1 - M h_2 \\ h_4 \end{pmatrix}.$$

Наконец, домножим коэффициенты третьей группы уравнений слева на матрицу V^{-1}, а затем почленно вычтем их из уравнений третьей группы:

$$\begin{pmatrix} E & -\varepsilon_1 E & -E & 0 \\ 0 & E & 0 & A^\top \\ 0 & 0 & E & -V^{-1}M A^\top \\ 0 & 0 & 0 & A V^{-1} M A^\top \end{pmatrix} \begin{pmatrix} p_w \\ p_v \\ p_x \\ p_y \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ V^{-1}(h_3 - V h_1 - M h_2) \\ h_4 - A V^{-1}(h_3 - V h_1 - M h_2) \end{pmatrix}.$$

Отсюда вытекает

Утверждение 7. *Решение системы* (6.13) *может быть последовательно получено по формулам*

$$\begin{aligned} p_y &= (A V^{-1} M A^\top)^{-1}\big[h_4 - A V^{-1}(h_3 - V h_1 - M h_2)\big], \\ p_x &= V^{-1}(h_3 - V h_1 - M h_2) + V^{-1} M A^\top p_y, \\ p_v &= h_2 - A^\top p_y, \\ p_w &= h_1 + \varepsilon_1 p_v + p_x, \end{aligned}$$

где $M = W + \varepsilon_1 V$.

И снова мы видим, что основной объем вычислений связан с обращением положительно определенной симметрической матрицы $Q = A^\top V^{-1} M A^\top$ размерности $m \times m$. Более того, вычисления и обращения этой матрицы можно избежать, если для решения подсистемы

$$(A V^{-1} M A^\top)\, p_y = h_4 - A V^{-1}(h_3 - V h_1 - M h_2)$$

воспользоваться методом сопряженных направлений. Это позволит эффективно учесть также возможную разреженность исходных матриц.

7. Вычислительный эксперимент

Вычислительный эксперимент проводился на задачах линейного программирования средней размерности (3000×10000) в среде MATLAB на вычислительном комплексе УРАН ИММ УрО РАН. Разреженные матрицы коэффициентов тестовых задач (50 штук) генерировались при помощи датчика случайных чисел с равномерным распределением от -1 до 1. Заполненность матриц ненулевыми элементами составляла 5–6%. Правые части ограничений и коэффициенты целевой функции подбирались таким образом, чтобы решение задачи было единственным и совпадало с некоторым заданным заранее. На каждой задаче сравнивались показатели

Таблица 1

Характеристики траектории метода для задачи ЛП на неравенствах размерности 3000×10000 **при** $\varepsilon_{1,2} = 0.001$

iter	$\|\Psi(z^k)\|$	$\|x^k - x_0\|$	$\|(Ax-b)_+\|$	Δopt	$\|y^k - y_0\|$
1	630.49	31.623	191.9	1.0	44.72
2	63.101	3.1653	9.56	0.1001	4.504
3	0.01049	0.01043	0.00524	$9.30e-07$	0.0907
4	0.00020	0.12031	0.00069	$1.70e-07$	0.5019
5	$1.84e-07$	0.11677	0.000711	$3.44e-07$	0.4995

Таблица 2

Характеристики траектории метода для задачи ЛП на неравенствах размерности 3000×10000 **при** $\varepsilon_{1,2} = 0.00001$

iter	$\|\Psi(z^k)\|$	$\|x^k - x_0\|$	$\|(Ax-b)_+\|$	Δopt	$\|y^k - y_0\|$
1	614.61	31.623	188.11	1	44.72
2	61.461	3.1623	9.689	0.1	4.472
3	$9.953e-05$	0.00010	$4.937e-05$	$4.883e-09$	0.00086
4	$1.140e-08$	$3.906e-06$	$7.069e-08$	$1.434e-10$	0.00541
5	$3.509e-13$	$3.858e-06$	$7.312e-08$	$1.154e-10$	0.00541
6	$1.065e-13$	$3.853e-06$	$7.310e-08$	$1.152e-10$	0.00541

Таблица 3

Характеристики траектории метода для задачи ЛП на неравенствах размерности 3000×10000 **при** $\varepsilon_{1,2} = 0.0000001$

iter	$\|\Psi(z^k)\|$	$\|x^k - x_0\|$	$\|(Ax-b)_+\|$	Δopt	$\|y^k - y_0\|$
1	628.22	31.623	184.31	1	44.721
2	62.82	3.1623	9.3347	0.1	4.4721
3	$1.063e-06$	$1.006e-06$	$5.25e-07$	$4.318e-10$	$9.398e-06$
4	$3.483e-12$	$2.719e-09$	$1.77e-11$	$6.847e-15$	0.000126
5	$3.400e-13$	$2.725e-09$	$1.79e-11$	$9.415e-15$	0.000126

эффективности алгоритма из разд. 4 (для задачи с ограничениями-неравенствами) и алгоритма из разд. 6 (для задачи в каноническом формате; такие задачи находятся в отношении двойственности друг к другу). Корректировка шагового параметра в методе Ньютона применялась только для предотвращения выхода отдельных последовательностей в отрицательную область. Стартовые приближения во всех случаях брались одинаковыми.

Результаты расчетов приведены в табл. 1—6. Обозначения колонок: ε — единое значение штрафного параметра, *iter* — порядковый номер итерации метода Ньютона, $\|x^k - x_0\|$ — отклонение полученного прямого решения от точного, $\|y^k - y_0\|$ — отклонение полученного двойственного решения от точного, $||(Ax-b)_+||$ — достигнутая точность выполнения ограничений общего вида, $||(-x)_+||$ — достигнутая точность выполнения требований неотрицательности переменных в канонической задаче, Δopt — достигнутая точность (относительная) по целевому функционалу.

Как видно из приведенных данных, оба метода показывают более быструю сходимость по сравнению с методами из работ [13–15]. При этом в ходе уменьшения штрафного параметра

Таблица 4

Характеристики траектории метода для канонической задачи ЛП размерности 10000×3000 **при** $\varepsilon_{1,2} = 0.001$

iter	$\|\Psi(z^k)\|$	$\|x^k - x_0\|$	$\|(-x)_+\|$	Δopt	$\|y^k - y_0\|$
1	6290.7	77.46	54.772	4.0119	0
2	629.06	7.7408	4.4811	0.4012	0.00168
3	0.00364	0.01710	0.00335	$2.796e-06$	0.00217
4	$5.918e-06$	0.04020	0.00016	$8.301e-08$	0.00969
5	$1.323e-10$	0.04149	0.00016	$8.324e-08$	0.00968

Таблица 5

Характеристики траектории метода для канонической задачи ЛП размерности 10000×3000 **при** $\varepsilon_{1,2} = 0.00001$

iter	$\|\Psi(z^k)\|$	$\|x^k - x_0\|$	$\|(-x)_+\|$	Δopt	$\|y^k - y_0\|$
1	6304	77.46	54.77	3.987	0
2	630.41	7.7459	4.472	0.398	$1.77e-05$
3	$3.559e-05$	0.000159	$3.21e-05$	$2.650e-08$	$2.14e-05$
4	$8.328e-10$	$1.802e-05$	$2.08e-08$	$7.263e-12$	0.00015
5	$2.468e-12$	$1.862e-05$	$2.13e-08$	$7.291e-12$	0.00015

Таблица 6

Характеристики траектории метода для канонической задачи ЛП размерности 10000×3000 **при** $\varepsilon_{1,2} = 0.0000001$

iter	$\|\Psi(z^k)\|$	$\|x^k - x_0\|$	$\|(-x)_+\|$	Δopt	$\|y^k - y_0\|$
1	6279.3	77.46	54.77	3.998	0
2	627.93	7.746	4.472	0.3998	$1.76e-07$
3	$3.60e-07$	$1.647e-06$	$3.25e-07$	$2.684e-10$	$2.22e-07$
4	$3.16e-12$	$4.381e-09$	$3.17e-12$	$6.002e-16$	$2.46e-06$
5	$2.16e-12$	$4.665e-09$	$3.26e-12$	$7.503e-16$	$2.46e-06$
6	$1.78e-12$	$4.665e-09$	$3.26e-12$	$7.503e-16$	$2.46e-06$

увеличение точности получаемых решений происходит несколько медленнее роста точности решения самих операторных уравнений. В обоих случаях сходимость прямого решения выражена более ярко по сравнению со сходимостью двойственного решения.

Заключение

В настоящей работе предложены оригинальные конструкции внешних штрафных функций в линейном и выпуклом программировании, асимптотически сводящие задачи условной оптимизации к задачам безусловной оптимизации повышенной гладкости. Последние допускают эффективное решение методами второго порядка и в то же время не требуют для своего обоснования существования внутренних допустимых точек исходной задачи. Более того, новые штрафные функции могут быть применены и к несобственным задачам линейного и выпуклого программирования (к задачам с противоречивыми системами ограничений), для которых они способны вырабатывать обобщенные (компромиссные) квазирешения.

СПИСОК ЛИТЕРАТУРЫ

1. **Еремин И. И.** Метод штрафов в выпуклом программировании // Докл. АН СССР. 1967. Т. 173, № 4. С. 748–751.
2. **Фиакко А., Мак-Кормик Г.** Нелинейное программирование. Методы последовательной безусловной минимизации. М.: Мир, 1972. 240 с.
3. **Зангвилл У. И.** Нелинейное программирование. Единый подход. М.: Сов. радио, 1973. 312 с.
4. **Полак Э.** Численые методы оптимизации. М.: Мир, 1974. 376 с.
5. **Моисеев Н. Н., Иванилов Ю. П., Столярова Е. М.** Методы оптимизации. М.: Наука, 1978. 351 с.
6. **Евтушенко Ю. Г.** Методы решения экстремальных задач и их применение в системах оптимизации. М.: Наука, 1982. 432 с.
7. **Гилл Ф., Мюррей У., Райт М.** Практическая оптимизация. М.: Мир, 1985. 509 с.
8. **Васильев Ф.П.** Численные методы решения экстремальных задач. М.: Наука, 1988. 552 с.
9. **Roos C., Terlaky T., Vial J.-Ph.** Theory and algorithms for linear optimization. Chichester: John Wiley & Sons Ltd, 1997. 484 p.
10. **Еремин И.И.** Двойственность для несобственных задач линейного и выпуклого программирования // Докл. АН СССР. 1981. Т. 256, № 2. С. 272–276.
11. **Еремин И. И., Мазуров Вл. Д., Астафьев Н. Н.** Несобственные задачи линейного и выпуклого программирования. М.: Наука, 1983. 336 с.
12. **Кочиков И.В., Матвиенко А.Н., Ягола А.Г.** Обобщенный принцип невязки для решения несовместных уравнений // Журн. вычисл. математики и мат. физики. 1984. Т. 24, № 7. С. 1087–1090.
13. **Попов Л. Д.** Комбинированные штрафы и обобщенные решения несобственных задач линейного и выпуклого программирования 1-го рода // Тр. Ин-та математики и механики УрО РАН. 2010. Т. 16, № 3. С. 217–226.
14. **Попов Л. Д.** Поиск обобщенных решений несобственных задач линейного и выпуклого программирования с помощью барьерных функций // Изв. Иркутского гос. ун-та. Сер. Математика. 2011. Т. 4, № 2. С. 134–146.
15. **Попов Л. Д.** Применение барьерных функций для оптимальной коррекции несобственных задач линейного программирования 1-го рода // Автоматика и телемеханика. 2012. Вып. 3. С. 3–11.

Поступила 19.05.2021
После доработки 20.07.2021
Принята к публикации 26.07.2021

Попов Леонид Денисович
д-р физ.-мат. наук, ведущий науч. сотрудник
Институт математики и механики им. Н. Н. Красовского УрО РАН;
профессор
Уральский федеральный университет
г. Екатеринбург
e-mail: popld@imm.uran.ru

REFERENCES

1. Eremin I. The “penalty” method in convex programming. *Soviet Math. Dokl.*, 1967, vol. 8, pp. 459–462.
2. Fiacco A., McCormick G. *Nonlinear programming: Sequential unconstrained minimization techniques.* New York: Wiley, 1968, 210 p. doi: 10.1137/1.9781611971316 . Translated to Russian under the title *Nelineinoe programmirovanie: Metody posledovatel'noi bezuslovnoi minimizatsii*, Moscow: Mir Publ., 1972, 240 p.
3. Zangwill W.I. *Nonlinear programming: A unified approach.* Englewood Cliffs: Prentice-Hall, 1969, 356 p. ISBN: 0136235794 . Translated to Russian under the title *Nelineinoe programmirovanie: Edinyi podkhod*, Moscow: Sov. Radio Publ., 1973, 312 p.
4. Polak E. *Computational methods in optimization: A unified approach.* NY: Acad. Press, 1971, 329 p. ISBN: 008096091X . Translated to Russian under the title *Chislennye metody optimizatsii*, Moscow: Mir Publ., 1974, 376 p.

5. Moiseev N.N., Ivanilov Yu.P., Stolyarova E.M. *Metody optimizatsii* [Optimization methods]. Moscow: Nauka Publ., 1978, 351 p. ISBN: 978-5-9500751-2-4 .
6. Evtushenko Yu.G. *Numerical optimization techniques*. NY: Springer-Verlag, 1985, 562 p. ISBN: 978-1-4612-9530-3 . Original Russian text published in Evtushenko Yu.G. *Metody resheniya ekstremal'nykh zadach i ikh primenenie v sistemakh optimizatsii*, Moscow: Nauka Publ., 1982, 432 p.
7. Gill P., Murray W., Wright M. *Practical optimization*. London: Acad. Press, 1982, 401 p. ISBN: 0122839528 . Translated to Russian under the title *Prakticheskaya optimizatsiya*, Moscow: Mir Publ., 1985, 509 p.
8. Vasil'ev F.P. *Chislennye metody resheniya ekstremal'nykh zadach* [Numerical methods for solving extremal problems]. Moscow: Nauka Publ., 1988, 552 p. ISBN: 5-02-013796-0 .
9. Roos C., Terlaky T., Vial J.-Ph. *Theory and algorithms for linear optimization*. Chichester: John Wiley & Sons Ltd, 1997, 484 p. ISBN: 0471956767 .
10. Eremin I.I. Duality for improper problems of linear and convex programming. *Sov. Math. Dokl.*, 1981, vol. 23, pp. 62–66.
11. Eremin I.I., Mazurov V.D., and Astaf'ev N.N. *Nesobstvennye zadachi lineinogo i vypuklogo programmirovaniya* [Improper problems of linear and convex programming]. Moscow: Nauka Publ., 1983, 336 p.
12. Kochikov I.V., Matvienko A.N., Yagola A.G. The generalized residual principle for the solution of inconsistent equations. *U.S.S.R. Comput. Math. Math. Phys.*, 1984, vol. 24, no. 4, pp. 78–80. doi: 10.1016/0041-5553(84)90233-7 .
13. Popov L.D. Combined penalties and generalized solutions for improper problems of linear and convex programming of the first kind. *Trudy Inst. Mat. i Mekh. UrO RAN*, 2010, vol. 16, no. 3, pp. 217–226 (in Russian).
14. Popov L.D. Search of generalized solutions to improper linear and convex programming problems using barrier functions. *The Bulletin of Irkutsk State University. Series Mathematics*, 2011, vol. 4, no. 2, pp. 134–146 (in Russian).
15. Popov L.D. Use of barrier functions for optimal correction of improper problems of linear programming of the 1st kind. *Autom. Remote Control*, 2012, vol. 73, no. 3, pp. 417–424. doi: 10.1134/S0005117912030010 .

Received May 19, 2021
Revised July 20, 2021
Accepted July 26, 2021

Funding Agency: This study is a part of the research carried out at the Ural Mathematical Center and supported by the Ministry of Science and Higher Education of the Russian Federation (agreement no. 075-02-2021-1383).

Leonid Denisovich Popov, Dr. Phys.-Math. Sci., Krasovskii Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences, Yekaterinburg, 620108 Russia; Ural Federal University, Yekaterinburg, 620000 Russia, e-mail: popld@imm.uran.ru .

Cite this article as: L. D. Popov. On one method of increasing the smoothness of external penalty functions in linear and convex programming, *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2021, vol. 27, no. 4, pp. 88–101 .