ТРУДЫ ИНСТИТУТА МАТЕМАТИКИ И МЕХАНИКИ УрО РАН

Том 31 № 3 2025

УДК 519.658.4

О ПРИНЦИПАХ ПОСТРОЕНИЯ И СВОЙСТВАХ ШТРАФНЫХ ФУНКЦИЙ КОМПОЗИТНОГО ТИПА (НА ПРИМЕРЕ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ)[1]

Л. Д. Попов

В данной статье продолжено исследование штрафных функций композитного типа в применении к задачам линейного программирования. Термин "композитные" объясняется тем, что графики таких функций получаются операцией гладкой склейки разнотипных графиков ряда привычных функций внутреннего и внешнего штрафов, что позволяет сохранять положительные качества склеиваемых компонент и устранять их специфические недостатки. В частности, эти конструкции сохраняют высокие свойства гладкости, позволяющие использовать для их минимизации методы второго порядка, и в то же время оказываются применимыми не только к задачам, допустимые области которых имеют непустую внутренность, но и к некорректным (несобственным, противоречивым, плохо поставленным) задачам, вовсе не имеющим допустимых планов; для последних композитные функции способны находить их так называемые аппроксимационные решения. Автор предлагает строгую аксиоматику таких функций, расширяя тем самым их перечень, а также доказывает отвечающие новой аксиоматике теоремы сходимости метода.

Ключевые слова: линейное программирование, комбинированные методы, методы внешней точки, методы внутренней точки, несобственные задачи, обобщенные решения.

L. D. Popov. Principles of construction and properties of penalty functions of composite type (on the example of linear programming problems).

In this paper the author continues his research on the application of penalty functions of composite type for solving linear programming problems. The term "composite" is explained by the fact that the graphs of such functions are obtained by the operation of smooth gluing of different-type graphs of a number of usual functions of internal and external penalties. Such an operation allows one to preserve the positive qualities of the glued components and eliminate their specific shortcomings. In particular, these constructions preserve the smoothness properties, allowing the use of second-order methods for their minimization, and at the same time are applicable not only to problems whose admissible regions have a non-empty interior, but also to ill-posed (improper, contradictory, poorly posed) problems that do not have admissible plans at all; for the latter, composite functions are capable of finding their so-called approximation solutions. The author proposes a rigorous axiomatization of such functions, thus extending their list, and also proves convergence theorems corresponding to the new axiomatization of the method.

Keywords: linear programming, combinations of the methods, interior penalties, exterior penalties, improper programs, generalizes solutions.

MSC: 90C05, 90C51, 90C53

DOI: 10.21538/0134-4889-2025-31-3-215-232

Введение

Технология сведения задач математического, и в частности линейного, программирования к решению серии (конечной или бесконечной) более простых задач безусловной оптимизации, построенных на базе применения различных штрафных функций, давно стала классикой (см. [1–10] и др.). Разнообразие используемых для этого штрафных функций чрезвычайно велико, и семейство таких функций постоянно расширяется за счет применения к ним различных схем регуляризации или посредством перехода от исходной оптимизационной задачи к задаче поиска седловой точки ее функции Лагранжа (см., например, [11–16] и др.).

[1]Работа выполнена в рамках исследований, проводимых в Уральском математическом центре при финансовой поддержке Министерства науки и высшего образования Российской Федерации (номер соглашения 075-02-2025-1549).

Обычно выделяют точные штрафные функции, внешние функции штрафа и барьерные функции (последние еще называют внутренними штрафными функциями). Выбирая при решении конкретной задачи то или иное из перечисленных направлений, приходится считаться с их специфическими чертами, как сильными, так и слабыми.

Например, точные штрафы [1; 2] бывает сложно использовать из-за их негладкости, но зато пользователю достаточно решить лишь одну включающую их вспомогательную задачу. К положительным свойствам точных штрафных функций также относится возможность их применения не только к регулярным, разрешимым задачам математического программирования, но и к некорректным (несобственным) задачам условной оптимизации [11; 12], т.е. к задачам с противоречивыми ограничениями, для которых эти методы способны находить их обобщенные решения. Данное свойство может оказаться полезным на начальных стадиях моделирования прикладных задач, когда надо быстро определить причины и размеры «узких” мест построенной математической модели, если такие обнаружены.

Методы внешнего штрафа (наиболее популярна здесь квадратичная функция от «положительной” срезки невязок ограничений исходной задачи [3–10]) порождают более гладкие вспомогательные задачи безусловной оптимизации, для решения которых подходят самые разнообразные варианты методов градиентного спуска. Кроме того, как и методы точных штрафных функций, методы внешних штрафов можно применять к плохо поставленным задачам условной оптимизации. Однако при использовании гладких штрафов возникает необходимость решать уже бесконечную серию вспомогательных задач оптимизации, свойства которых с ростом штрафного параметра становятся все более и более неудобными в плане вычислений.

Что касается барьерных функций или функций внуреннего штрафа, то они традиционно предъявляли высокие требования к допустимой области исходной задачи. В частности, упомянутая область должна обязательно содержать внутренние точки [4–10]. Однако возможность применения к барьерным функциям методов оптимизации второго порядка сделала эти функции настолько привлекательными, что одна из них, а именно логарифмическая, легла в основу современных методов внутренних точек и центрального пути [16–18]. Последние являются признанными лидерами среди коммерческих систем условной оптимизации. Более того, в настоящее время известно несколько подходов, позволяющих применять метод барьерных функций и построенные на их основе методы центрального пути к задачам, допустимые области которых не имеют внутренних точек и даже могут быть пустыми. Среди этих подходов выделяются приемы симметричной регуляризации функции Лагранжа исходной задачи и первые идеи разработки штрафных функций композитного типа (см., например, [19; 20]), показавшие в предварительных компьютерных экспериментах весьма высокую эффективность.

В настоящей работе мы продолжаем свои ранние исследования по композитным штрафным функциям. Напомним, что последние представляют собой гладкую “склейку” обычной внешней штрафной функции с различными функциями барьерного и квазибарьерного типов, что позволяет объединить их положительные свойства и свести на нет их отрицательные качества. В новом исследовании мы предлагаем строгий аксиоматический подход к таким функциям, расширяем число приводимых конкретных конструкций композитного типа, а также приводим подробное обоснование применения данных функций не только к обычным, разрешимым задачам линейного программирования, но и к задачам, называемым несобственными или плохо поставленными (т. е. к задачам с противоречивыми системами ограничений), для которых новые методы находят их псевдорешения.

Работа организована следующим образом. В разд. 1 аксиоматически описаны предлагаемые новые смешанные конструкции штрафных функций различных типов. В разд. 2 обоснована сходимость оптимальных значений вспомогательных задач, порождаемых такими конструкциями, к оптимальному значению исходной задачи в случае ее разрешимости. В разд. 3 разработаны детали применения к новым конструкциям метода Ньютона и его обобщений. В разд. 5 исследовано поведение новых функций в применении к несобственным задачам линейного программирования 1-го рода (задачам, не имеющим решения из-за противоречивости их ограничений). Для таких задач метод находит их псевдорешения. В заключении подведены итоги исследования.

1. Постановка задачи и исходные построения

Рассмотрим общую задачу линейного программирования (ЛП) вида

$$\max\big\{(c,x)\colon\, Ax\le b\big\}, \tag{1.1}$$

ограничения которой записаны в виде неравенств, и двойственную к ней задачу, представленную в каноническом формате:

$$\min\big\{(b,y)\colon\, A^T y=c,\quad y\ge 0\big\}. \tag{1.2}$$

Здесь векторы $c\in\mathbb{R}^n$, $b\in\mathbb{R}^m$ и матрица $A=(a_{ji})_{m\times n}$ заданы, векторы $x\in\mathbb{R}^n$ и $y\in\mathbb{R}^m$ обозначают переменные выписанных задач (их оптимальные значения предстоит найти), круглые скобки означают скалярное произведение векторов, $m\ge n=rank(A)$. Неравенства прямой задачи будем также записывать как

$$l_j(x)=(a_j,x)-b_j\le 0 \quad (j=1,2,\dots,m),$$

где a_j — строки матрицы A. Именно для задач приведенного формата нагляднее и проще всего изложить основы штрафных функций композитного типа.

На первом этапе исследования будем считать задачу (1.1), а значит, и задачу (1.2) разрешимыми (в этом случае они разделяют общее оптимальное значение $\bar v$). Их допустимые множества будем обозначать через M и M^*, а их оптимальные множества — через $\bar M$ и $\bar M^*$ соответственно. Напомним, что произвольные оптимальные векторы этих задач $\bar x\in\bar M$ и $\bar y\in\bar M^*$ связаны друг с другом классическими условиями дополнительности

$$(\bar x, A^T\bar y-c)=0,\quad (\bar y, A\bar x-b)=0,$$

откуда следует, что $(\bar x, A^T\bar y)=(b,\bar y)=(c,\bar x)=\bar v$.

Как уже говорилось во введении, одним из традиционных подходов к численному анализу задач (1.1), (1.2) является асимптотическая замена их серией вспомогательных задач безусловной оптимизации, заключающихся в поиске

$$\sup_x \mathcal{H}(x,\lambda,\omega),\quad \text{где}\quad \mathcal{H}(x,\lambda,\omega)=(c,x)-H(Ax-b,\lambda,\omega). \tag{1.3}$$

Здесь так называемая функция штрафа (внешнего или внутреннего) имеет вид

$$H(\cdot,\lambda,\omega):\mathbb{R}^m\to\mathbb{R},\quad H(z,\lambda,\omega)=\omega\sum_{j=1}^m h\big(z_j,\lambda\big);$$

при этом функция $h(\cdot,\lambda):\mathbb{R}\to\mathbb{R}$ играет роль унифицированного “строительного блока” для всей конструкции, штрафной параметр λ как правило неограниченно растет, ω — положительный масштабный множитель (постоянный или переменный).

Наиболее популярной функцией внешнего штрафа является сумма квадратов “положительных срезок”

$$H(Ax-b,\lambda,1)=\lambda\big\|(Ax-b)^+\big\|^2=\lambda\sum_{j=1}^m\big[l_j(x)^+\big]^2,$$

где $a^+=\max(0,a)$ для числа a, и $z^+=(z_1^+,\dots,z_m^+)$ для вектора $z=(z_1,\dots,z_m)$. Данная функция выпукла и определена всюду, но имеет непрерывные производные только первого порядка, что делает невозможным применение к ней методов оптимизации типа Ньютона.

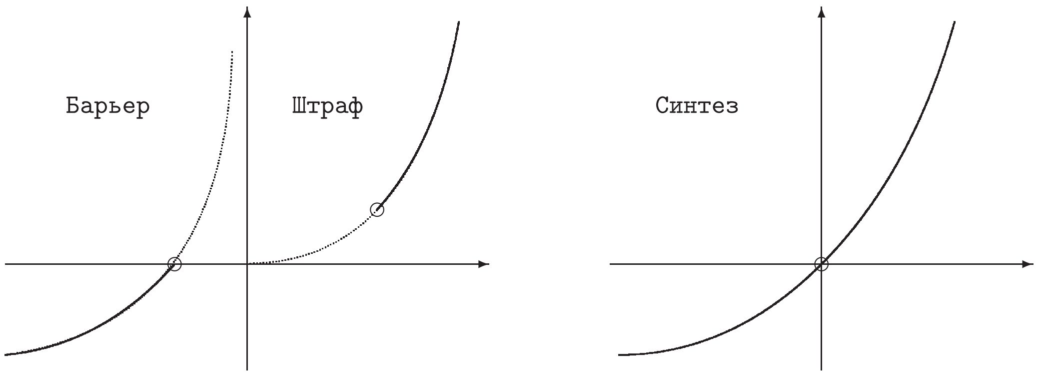

Рис. 1. Геометрический смысл операции “склейки”.

Среди функций внутреннего штрафа (барьерных функций) наиболее популярны обратная, логарифмическая и степенная функции:

$$H(Ax-b,\lambda,1)=-\sum_{j=1}^{m}\frac{1}{\lambda l_j(x)},\quad H(Ax-b,\lambda,1)=-\frac{1}{\lambda}\sum_{j=1}^{m}\ln(-l_j(x))$$

и

$$H(Ax-b,\lambda,1)=-\frac{1}{\lambda}\sum_{j=1}^{m}\sqrt{-l_j(x)}.$$

Эти функции выпуклы и уже дважды непрерывно дифференцируемы, но определены только внутри допустимой области исходной задачи (а последняя функция — также и на ее границе), что порождает определенные сложности при поиске точек их минимума.

Основная идея построения нового класса штрафных функций, предлагаемая в настоящей статье, состоит в том, что мы постараемся соединить в единый гладкий график отрицательную ветвь той или иной выбранной барьерной или квазибарьерной функции с положительной ветвью квадратичной функции штрафа (рис. 1). Именно поэтому мы назвали этот класс функций композитными штрафными функциями.

Специфическая сложность операции “склейки” заключается в необходимости обеспечить высокую степень гладкости итоговой штрафной функции. В частности, нам удалось получить по крайней мере три примера строительных блоков для штрафных функций композитного типа. Это

$$h_1(t,\lambda)=\begin{cases}\dfrac{t}{1-\lambda t} & \text{при } t\le 0,\\ t+\lambda t^2 & \text{при } t>0,\end{cases}\qquad h_2(t,\lambda,)=\begin{cases}-\dfrac{2}{\lambda}\ln(1-\lambda t) & \text{при } t\le 0,\\ 2t+\lambda t^2 & \text{при } t>0,\end{cases}$$

и

$$h_3(t,\lambda)=\begin{cases}\dfrac{8}{\lambda}\left(1-\sqrt{1-\lambda t}\right) & \text{при } t\le 0,\\ 4t+\lambda t^2 & \text{при } t>0.\end{cases}$$

Подчеркнем, что все приведенные конструкции имеют непрерывные производные 1-го и 2-го порядков:

$$h_1'(t,\lambda)=\begin{cases}\dfrac{1}{(1-\lambda t)^2} & \text{при } t\le 0,\\ 1+2\lambda t & \text{при } t>0,\end{cases}\qquad h_1''(t,\lambda)=\begin{cases}\dfrac{2\lambda}{(1-\lambda t)^3} & \text{при } t\le 0\\ 2\lambda & \text{при } t>0.\end{cases}$$

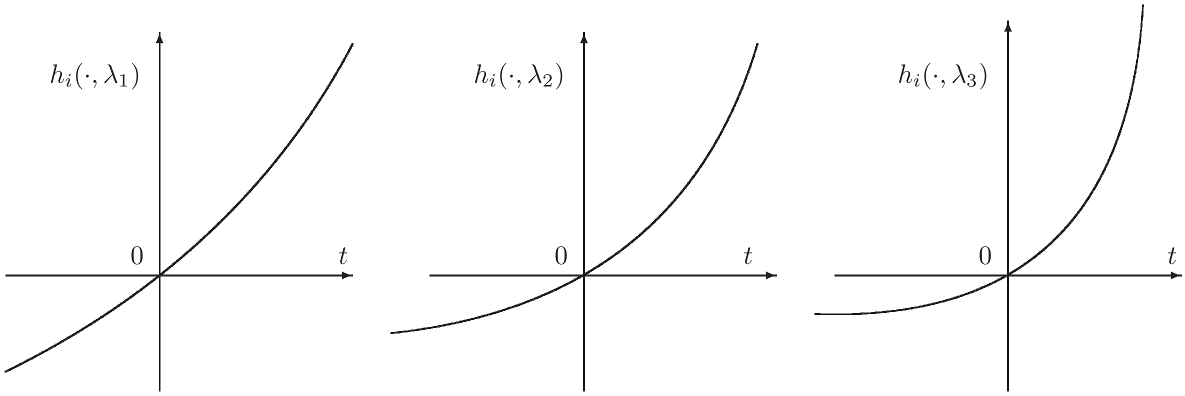

Рис. 2. Характер зависимостей $h_i(\cdot,\lambda)$ при $\lambda_1 < \lambda_2 < \lambda_3$.

$$h_2'(t,\lambda) = \begin{cases} \dfrac{2}{1-\lambda t} & \text{при } t \le 0, \\ 2 + 2\lambda t & \text{при } t > 0, \end{cases} \qquad h_2''(t,\lambda) = \begin{cases} \dfrac{2\lambda}{(1-\lambda t)^2} & \text{при } t \le 0, \\ 2\lambda & \text{при } t > 0, \end{cases}$$

$$h_3'(t,\lambda) = \begin{cases} \dfrac{4}{\sqrt{1-\lambda t}} & \text{при } t \le 0, \\ 4 + 2\lambda t & \text{при } t > 0, \end{cases} \qquad h_3''(t,\lambda) = \begin{cases} \dfrac{2\lambda}{(1-\lambda t)^{3/2}} & \text{при } t \le 0, \\ 2\lambda & \text{при } t > 0. \end{cases}$$

На рис. 2 условно изображены графики предложенных функций при различных (возрастающих) значениях штрафного параметра.

Чтобы унифицировать дальнейшие исследования, попробуем кратко сформулировать основные качества (свойства) приведенных выше примеров.

Свойство A1. Функция $h(\cdot,\lambda) : \mathbb{R} \to \mathbb{R}$ дважды непрерывно дифференцируема, причем ее первая и вторая производные всюду положительны, а $h(0,\lambda) = 0$.

Свойство A2. При всех $t \ge 0$ верны оценки снизу $h(t,\lambda) \ge t + \lambda t^2$.

Для формулировки третьего свойства рассмотрим две вспомогательные функции:

$$\mathcal{C}_1(\sigma,\lambda,\omega) := \max_{t\le 0} \left[\sigma t - \omega h(t,\lambda)\right] \ge 0 \quad \text{и} \quad \mathcal{C}_2(\sigma,\lambda,\omega) := \max_{t\le 0} \left[\frac{\sigma}{\lambda+1} t - \omega h(t,\lambda)\right] \ge 0.$$

Свойство A3. При всех $\sigma > 0$ имеет место сходимость

$$\limsup_{\omega\to +0,\ \omega\lambda\to+\infty} \mathcal{C}_2(\sigma,\lambda,\omega) = 0.$$

Поскольку при всех $\sigma > 0$ обе вспомогательные функции неотрицательны и $\mathcal{C}_1(\sigma,\lambda,\omega) \le \mathcal{C}_2(\sigma,\lambda,\omega)$, то при выполнении свойства A3 также

$$\limsup_{\omega\to +0,\ \omega\lambda\to+\infty} \mathcal{C}_1(\sigma,\lambda,\omega) = 0.$$

Примеры. Рассмотрим каждую из функций $h_i(t,\lambda)$, $i = 1,2,3$, по отдельности и покажем, что все они удовлетворяют свойству A3 (выполнение A1 и A2 не вызывает вопросов). С этой целью в случае $i = 1$ воспользуемся вспомогательной функцией $\mathcal{H}_1(t,\sigma) = \sigma t - t/(1-t)$, в которой $\sigma > 0$ — числовой параметр и t — числовой аргумент. Рассматриваемая на отрицательной полуоси $t \le 0$, функция $\mathcal{H}_1(t,\sigma)$ при любом значении параметра будет вогнутой и ограниченной сверху. В самом деле, при $\sigma \ge 1$ ее первая производная $\mathcal{H}_1'(t,\sigma) = \sigma - 1/(1-t)^2$ неотрицательна на этой полуоси, а потому сама функция достигает на ней максимума в нуле $\mathcal{H}_1(0,\sigma) = 0$. При $0 < \sigma < 1$ та же самая производная обращается в ноль в точке

$t' = (\sqrt{\sigma} - 1)/\sqrt{\sigma} < 0$, которая тем самым является точкой максимума рассматриваемой функции на отрицательной полуоси, при этом сам максимум равен $\mathcal{H}_1(t', \sigma) = \big(1 - \sqrt{\sigma}\big)^2$. В любом случае

$$W_1(\sigma) := \sup_{t \le 0} \mathcal{H}_1(t, \sigma) \le 1.$$

Поэтому при $\omega\lambda \to +\infty$ и $\omega =\to +0$ имеем равномерную по t сходимость

$$\frac{\sigma t}{\lambda+1} - \omega h_1(t, \lambda) = \frac{\sigma t}{\lambda+1} - \frac{\omega t}{1-\lambda t} = \frac{\omega}{\lambda}\Big(\frac{\sigma\lambda t}{\omega(\lambda+1)} - \frac{\lambda t}{1-\lambda t}\Big)$$

$$= \frac{\omega}{\lambda}\mathcal{H}_1\Big(\lambda t, \frac{\sigma}{\omega(\lambda+1)}\Big) \le \frac{\omega}{\lambda} W_1\Big(\frac{\sigma}{\omega(\lambda+1)}\Big) \le \frac{\omega}{\lambda} \to +0.$$

При $i = 2$ воспользуемся схожей вспомогательной функцией $\mathcal{H}_2(t, \sigma) = \sigma t + \ln(1-t)$, в которой $\sigma > 0$ — числовой параметр и t — числовой аргумент. Рассматриваемая на отрицательной полуоси $t \le 0$, эта функция также вогнута и всегда ограничена сверху, так как при $\sigma \ge 1$ она также не положительна, а при $0 < \sigma < 1$ ее производная $\mathcal{H}_2'(t, \sigma) = \sigma - 1/(1-t)$ обращается в ноль в точке $t' = (\sigma - 1)/\sigma < 0$, в которой сама функция принимает максимальное значение $\mathcal{H}_2(t', \sigma) = \sigma - 1 - \ln\sigma > 0$. В любом из двух случаев

$$W_2(\sigma) := \sup_{t \le 0} \mathcal{H}_2(t, \sigma) \le \sigma - 1 - \ln\sigma.$$

Отсюда при $\omega\lambda \to +\infty$ и $\omega =\to +0$ также имеем равномерную по t сходимость

$$\frac{\sigma t}{\lambda+1} - \omega h_2(t, \lambda) = \frac{\sigma t}{\lambda+1} - \frac{2\omega}{\lambda}\ln(1-\lambda t)$$

$$= \frac{2\omega}{\lambda}\Big(\frac{\sigma\lambda t}{2\omega(\lambda+1)} - \ln(1-\lambda t)\Big) = \frac{2\omega}{\lambda}\mathcal{H}_2\Big(\lambda t, \frac{\sigma}{2\omega(\lambda+1)}\Big) \le \frac{2\omega}{\lambda} W_2\Big(\frac{\sigma}{2\omega(\lambda+1)}\Big)$$

$$= \frac{\sigma}{\lambda(\lambda+1)} - \frac{2\omega(1+\ln\sigma)}{\lambda} + \frac{2\omega\ln\big(2\omega\big)}{\lambda} + \frac{2\omega\ln(\lambda+1)}{\lambda} \to +0.$$

Наконец, при $i = 3$ воспользуемся свойствами функции $\mathcal{H}_3(t, \sigma) = \sigma t - \big(1 - \sqrt{1-t}\big)$, в которой $\sigma > 0$ — числовой параметр и t — числовой аргумент. Рассматриваемая на отрицательной полуоси $t \le 0$, эта функция также оказывается вогнутой и ограниченной сверху. Действительно, она не положительна при $\sigma \ge 1/2$, а при $0 < \sigma < 1/2$ ее производная $\mathcal{H}_3'(t, \sigma) = \sigma - 1/(2\sqrt{1-t})$ обращается в ноль в точке $t' = (4\sigma^2 - 1)/(4\sigma^2) < 0$, в которой сама функция принимает максимальное на полуоси значение $\mathcal{H}_3(t', \sigma) = (1-2\sigma)^2/(4\sigma) > 0$. В любом из этих случаев

$$W_3(\sigma) = \sup_{t \le 0} \mathcal{H}_3(t, \sigma) \le \frac{1}{4\sigma}.$$

Соответственно при $\omega\lambda \to +\infty$ и $\omega =\to +0$ вновь имеем равномерную по t сходимость

$$\frac{\sigma t}{\lambda+1} - \omega h_3(t, \lambda) = \frac{\sigma t}{\lambda+1} - \frac{8\omega}{\lambda}\Big(1 - \sqrt{1-\lambda t}\Big)$$

$$= \frac{8\omega}{\lambda}\Big(\frac{\sigma\lambda t}{8\omega(\lambda+1)} - \Big(1 - \sqrt{1-\lambda t}\Big)\Big) = \frac{8\omega}{\lambda}\mathcal{H}_3\Big(\lambda t, \frac{\sigma}{8\omega(\lambda+1)}\Big)$$

$$\le \frac{8\omega}{\lambda} W_3\Big(\frac{\sigma}{8\omega(\lambda+1)}\Big) = \frac{16\omega^2(\lambda+1)}{\lambda\sigma} \to +0.$$

З а м е ч а н и е 1. Очевидно, что сходимость масштабного множителя ω к нулю имеет принципиальное значение только для третьей из рассматриваемых композитных функций (см. последнюю формулу). Для первых двух функций достаточно просто ограниченности этого множителя сверху (например, единицей). Однако для сходимости метода в случае несобственности исходных задач сходимость к нулю коэффициента масштаба будет очень важна.

2. Связь композитных функций штрафа с исходной задачей

Теперь мы готовы изучить связь задачи (1.3) с решением исходной задачи (1.1).

Утверждение 1. *Пусть задача (1.1) разрешима, а среди допустимых векторов двойственной к ней задачи имеются положительные. Пусть также функция $h(t,\lambda)$, лежащая в основе функции штрафа $\mathcal{H}(x,\lambda,\omega)$, обладает свойствами A1–A3. Тогда*

$$\lim_{\omega\to+0,\ \lambda\omega\to+\infty}\Big(\sup_x \mathcal{H}(x,\lambda,\omega)\Big)=\bar v. \tag{2.1}$$

Более того, точная верхняя грань здесь достигается всякий раз в единственной точке $\bar x(\lambda,\omega)$, и последовательность таких точек сходится к оптимальному множеству задачи (1.1) по расстоянию.

Д о к а з а т е л ь с т в о проведем в три этапа.

Этап 1. Обоснуем равенство (2.1). Начнем с того, что в силу разрешимости исходной задачи и свойства A1 при всех $\omega>0$, $\lambda>0$ имеем неравенство

$$\sup_x \mathcal{H}(x,\lambda,\omega)\ge \mathcal{H}(\bar x,\lambda,\omega)=(c,\bar x)-H(A\bar x-b,\lambda,\omega)=\bar v-\underbrace{H(A\bar x-b,\lambda,\omega)}_{\le 0}\ge\bar v. \tag{2.2}$$

Чтобы получить обратное неравенство, выберем в допустимом множестве двойственной задачи (1.2) произвольный положительный вектор y и зафиксируем его (по сделанным предположениям это можно сделать). Поскольку для этого вектора $A^Ty=c$ и $(b,y)\ge\bar v$, то при всех $x\in\mathbb{R}^n$ выполняются равенства

$$\mathcal{H}(x,\lambda,\omega)=(c,x)-H(Ax-b,\lambda,\omega)=(y,Ax)-H(Ax-b,\lambda,\omega)$$
$$=(y,b)+(y,Ax-b)-H(Ax-b,\lambda,\omega).$$

Далее разделим множество индексов $J=\{1,2,\dots,m\}$ на два непересекающихся индексных подмножества $J_+(x)=\{j\in J\colon l_j(x)>b_j\}$ и $J_-(x)=\{j\in J\colon l_j(x)\le b_j\}$. В результате

$$\mathcal{H}(x,\lambda,\omega)=(b,y)+\sum_{j\in J_+(x)}\big[y_jl_j(x)-\omega h(l_j(x),\lambda)\big]+\sum_{j\in J_-(x)}\big[y_jl_j(x)-\omega h(l_j(x),\lambda)\big],$$

где слагаемые каждой группы вычисляются по своим правилам.

Слагаемые первой, “положительной” группы, пользуясь свойством A2, можно оценить как

$$y_jl_j(x)-\omega h(l_j(x),\lambda)\ \le\ y_jl_j(x)-\omega\lambda l_j^2(x)=\frac{y_j^2}{4\lambda\omega}-\left[\frac{y_j}{2\sqrt{\lambda\omega}}-\sqrt{\lambda\omega}\,l_j(x)\right]^2\le\frac{1}{4\lambda\omega}y_j^2.$$

Слагаемые второй, “отрицательной” группы оцениваются на основе свойства A3 как

$$y_jl_j(x)-\omega h(l_j(x),\lambda)\le\mathcal{C}_1(y_j,\lambda,\omega).$$

Суммируя эти две группы слагаемых, выводим огрубленную оценку сверху для

$$\mathcal{H}(x,\lambda,\omega)\le(b,y)+\frac{\|y\|^2}{4\lambda\omega}+\sum_{j=1}^m\mathcal{C}_1(y_j,\lambda,\omega).$$

Здесь правая часть неравенства не зависит от x, в силу чего

$$\sup_x\mathcal{H}(x,\lambda,\omega)\le(b,y)+\frac{\|y\|^2}{4\lambda\omega}+\sum_{j=1}^m\mathcal{C}_1(y_j,\lambda,\omega),$$

и после перехода здесь к пределу слева и справа получаем неравенство

$$\lim_{\omega\to+0,\ \omega\lambda\to+\infty}\Big(\sup_x \mathcal{H}(x,\lambda,\omega)\Big)\le (b,y).$$

Здесь мы еще раз воспользовались свойством A3.

Остается подставить на место фиксированной положительной точки y последовательность аналогичных точек y_k, сходящуюся к $\bar y$ (такую легко построить), и еще раз перейти к пределу, теперь уже по последовательности $y_k\to\bar y$. В силу предположения о разрешимости исходной задачи верно $(b,y_k)\to\bar v$, и мы получим противоположное (2.2) неравенство

$$\lim_{\omega\to+0,\ \omega\lambda\to+\infty}\Big(\sup_x \mathcal{H}(x,\lambda,\omega)\Big)\le \bar v.$$

Равенство (2.1) доказано.

Этап 2. Убедимся в ограниченности хотя бы одного из непустых множеств уровня функции $\mathcal{H}(x,\lambda,\omega)$. Обозначим такое множество через

$$\Omega_K^{\lambda,\omega}:=\big\{x\in\mathbb{R}^n\colon \mathcal{H}(x,\lambda,\omega)=(c,x)-H(Ax-b,\lambda,\omega)\ge K\big\}.$$

Поскольку понижение уровня K сохраняет непустоту множеств уровня, будем считать $K<\bar v$.

Отметим предварительно, что по условиям доказываемого утверждения существуют положительные точки y, для которых $c=A^Ty$. Поэтому оптимальное множество прямой задачи $\bar M=\{x\in\mathbb{R}^n\colon Ax\le b,\ (c,x)\ge\bar v\}$ ограничено [5], а вместе с ним ограниченным будет и любое непустое множество вида

$$\bar M_{\Delta b}^{\delta}=\big\{x\in\mathbb{R}^n\colon Ax\le b+\Delta b,\ (c,x)\ge\bar v-\delta\big\}. \tag{2.3}$$

Покажем, как по параметрам K, λ, y можно подобрать такие $\Delta b=\Delta b(K,\lambda,y)$ и $\delta=\delta(K,\lambda,y)$, которые обеспечили бы включение $\Omega_K^{\lambda,\omega}\subseteq\bar M_{\Delta b}^{\delta}$.

Итак, пусть $x\in\Omega_K^{\lambda,\omega}$, т. е. $\mathcal{H}(x,\lambda,\omega)\ge K$. Начнем с оценки невязки прямых ограничений, для чего проследим следующую цепочку неравенств:

$$\begin{aligned}\lambda\omega\|(Ax-b)^+\|^2&\le\sum_{j\in J_+(x)}\omega h(l_j(x),\lambda)=-\mathcal{H}(x,\lambda,\omega)+(c,x)-\sum_{j\in J_-(x)}\omega h(l_j(x),\lambda)\\&=-\mathcal{H}(x,\lambda,\omega)+(y,b)+(y,Ax-b)-\sum_{j\in J_-(x)}\omega h\big(l_j(x),\lambda\big)\\&\le -K+(y,b)+\|y\|\,\big\|(Ax-b)^+\big\|+\sum_{j=1}^m\mathcal{C}_1(y_j,\lambda,\omega).\end{aligned}$$

Здесь мы снова воспользовались свойством A3.

Полученная оценка представляет собой квадратное неравенство относительно числовой переменой $t=\|(Ax-b)^+\|$, а именно,

$$\lambda\omega\,t^2-\|y\|\,t-K_1(y,\lambda,\omega)\le 0,$$

где в силу сделанных предположений и теоремы слабой двойственности верно неравенство $K_1(y,\lambda,\omega):=(b,y)-K+\sum_{j=1}^m\mathcal{C}_1(y_j,\lambda,\omega)>0$. Отсюда вытекает искомая оценка сверху для

$$\|(Ax-b)^+\|=t\le K_2(y,\lambda,\omega):=\frac{\|y\|+\sqrt{\|y\|^2+4\lambda\omega K_1(y,\lambda,\omega)}}{2\lambda\omega}. \tag{2.4}$$

Это позволяет нам определить Δb, подходящее для выполнения включения $\Omega_K^{\lambda,\omega}\subseteq\bar M_{\Delta b}^{\delta}$, а именно $\Delta b(K,\lambda,\omega)=K_2(y,\lambda,\omega)\,e$.

Перейдем теперь к подбору подходящего δ.

Для этого построим оценку снизу для величины (c,x), где $x \in \Omega_K^{\lambda,\omega}$. Во-первых, возьмем грубую оценку вида

$$(c,x) = \mathcal{H}(x,\lambda,\omega) + H(Ax-b,\lambda,\omega) \geq K + \sum_{j\in J_-(x)} \omega h\big(l_j(x),\lambda\big).$$

Во-вторых, если y — положительный допустимый вектор двойственной задачи, существующий по исходным предположениям, то

$$(c,x) = (y, Ax-b) + (y,b) = (b,y) + \sum_{j\in J_+(x)} y_j l_j(x) + \sum_{j\in J_-(x)} y_j l_j(x).$$

Умножим первое соотношение на $\mu = (\lambda+1)/\lambda > 1$ и вычтем из него почленно второе соотношение, предварительно умноженное на $1/\lambda$. С учетом свойства A3 получим

$$\begin{aligned}(c,x) &\geq \frac{(\lambda+1)K-(y,b)}{\lambda} - \sum_{j\in J_+(x)} \frac{y_j l_j(x)}{\lambda} - \frac{\lambda+1}{\lambda}\sum_{j\in J_-(x)} \Big[\frac{1}{\lambda+1} y_j l_j(x) - \omega h\big(l_j(x),\lambda\big)\Big] \\ &\geq \frac{(\lambda+1)K-(y,b)}{\lambda} - \frac{1}{\lambda}\|y\|\,\big\|(Ax-b)^+\big\| - \frac{\lambda+1}{\lambda}\sum_{j=1}^m \mathcal{C}_2(y_j,\lambda,\omega). \end{aligned} \tag{2.5}$$

Отсюда согласно найденным ранее оценкам для Δb и для $\|(Ax-b)^+\|$ выводим формулу для определения параметра

$$\delta = \delta(K,\lambda,y) := \bar{v} - \frac{(\lambda+1)K-(b,y)}{\lambda} + \frac{1}{\lambda}\|y\| K_2(y,\lambda,\omega) + \frac{\lambda+1}{\lambda}\sum_{j=1}^m \mathcal{C}_2(y_j,\lambda,\omega),$$

при котором $(c,x) \geq \bar{v} - \delta$ и следовательно $\Omega_K^{\lambda,\omega} \subseteq \bar{M}_{\Delta b}^{\delta}$.

Таким образом, множество уровня $\Omega_K^{\lambda,\omega}$ оказалось частью ограниченного множества вида (4.6) и в силу непрерывности и строгой вогнутости функции $\mathcal{H}(x,\lambda,\omega)$ верхняя граница в (1.3) конечна и достигается всякий раз в единственной точке $\bar{x}(\lambda,\omega)$, зависящей от λ, ω.

Этап 3. Нам осталось показать сходимость таких точек к оптимальному множеству задачи (1.1) при определенном правиле согласованного поведения параметров метода.

Итак, пусть выбраны последовательности $\omega_k \to +0$, $\omega_k\lambda_k \to +\infty$ и пусть $\bar{x}_k$ — отвечающая им последовательность точек максимума функций $\mathcal{H}(x,\lambda_k,\omega_k)$. По этой последовательности выстроим последовательность уровней

$$N_k = \sup_x \mathcal{H}(x,\lambda_k,\omega_k) - \varepsilon_k,$$

где ε_k — произвольная последовательность положительных чисел, $\varepsilon_k \to +0$. Следует заметить, что все $\bar{x}_k \in \Omega_{N_k}^{\lambda_k,\omega_k}$ и потому в силу соотношений (2.1), (4.7), (4.8) и свойств A1–A3 имеем

$$\|(A\bar{x}_k - b)^+\| \leq \frac{\|y\|}{2\lambda_k\omega_k} + \sqrt{\Big(\frac{\|y\|}{2\lambda_k\omega_k}\Big)^2 + \frac{1}{\lambda_k\omega_k}\Big[(b,y) - N_k + \sum_{j=1}^m \mathcal{C}_1(y_j,\lambda_k,\omega_k)\Big]} \to +0,$$

$$(c,\bar{x}_k) \geq \frac{(\lambda_k+1)N_k-(b,y)}{\lambda_k} - \frac{1}{\lambda_k}\|y\|\,\big\|(A\bar{x}_k-b)^+\big\| - \frac{\lambda_k+1}{\lambda_k}\sum_{j=1}^m \mathcal{C}_2(y_j,\lambda_k,\omega_k) \to \bar{v}.$$

Утверждение полностью доказано.

Замечание 2. Для первых наших двух примеров композитных функций последнее утверждение о сходимости метода остается справедливым даже в случае простой ограниченности масштабного множителя $\omega_k \equiv 1$. Для третьего примера условие сходимости $\omega_k \to +0$ существенно (это подтверждают и численные эксперименты). Тем самым для третьей композитной функции повышаются требования к темпу роста штрафного параметра λ (ведь надо обеспечить сходимость $\omega_k\lambda_k \to +\infty$).

3. Новые штрафы и метод Ньютона

Для решения вспомогательных задач из серии (1.3) можно применять метод Ньютона или его обобщения [8;21–23]. Как известно, метод Ньютона, отталкиваясь от произвольного начального приближения x^0, формирует последовательность x^k все более точных приближений к решению интересующей нас задачи по правилу

$$x^{k+1} = x^k - \tau \left(\nabla^2 \mathcal{H}(x^k, \lambda, \omega)\right)^{-1} \nabla \mathcal{H}(x^k, \lambda, \omega), \quad k = 0, 1, 2, \dots,$$

где числовой параметр шага τ обычно равен 1 (может регулироваться по правилу Армихо). Метод Ньютона характеризуется квадратичной скоростью сходимости (локальной).

В нашем случае

$$\frac{\partial \mathcal{H}(x, \lambda, \omega)}{\partial x_k} = c_k - \sum_{j=1}^{m} \omega h'\big(l_j(x), \lambda\big)\, a_{jk}, \quad \frac{\partial^2 \mathcal{H}(x, \lambda, \omega)}{\partial x_k \partial x_s} = -\sum_{j=1}^{m} \omega h''\big(l_j(x), \lambda\big)\, a_{jk}\, a_{js}$$

или, на векторно-матричном языке,

$$\nabla \mathcal{H}(x, \lambda, \omega) = c - A^T U(x, \lambda, \omega) e, \quad \nabla^2 \mathcal{H}(x, \lambda, \omega) = -A^T V(x, \lambda, \omega) A,$$

где $U(x, \lambda, \omega)$ и $V(x, \lambda, \omega)$ — диагональные матрицы с элементами $\omega h'\big(l_j(x), \lambda\big)$ и $\omega h''\big(l_j(x), \lambda\big)$ на своих диагоналях соответственно (здесь e — вектор, состоящий из единиц).

Заметим, что в силу сделанных предположений относительно исходной задачи и структуры ее композитной функции штрафа гессиан этой функции представляет собой симметричную отрицательно определенную матрицу.

Для приложений важно, что последовательность векторов x^k, порождаемых процессом Ньютона или его обобщениями, позволяет также определить последовательность приближений $y^k = U(x^k, \lambda, \omega)\, e$ к решению двойственной задачи (1.2). Приведем соответствующее утверждение.

Утверждение 2. *Пусть задача* (1.1) *разрешима, и ее оптимальное множество ограничено. Пусть также* $\bar{x}(\lambda, \omega)$ *— точка максимума функции* $\mathcal{H}(x, \lambda, \omega)$. *Тогда точка* $\bar{y}(\lambda, \omega) = U\big(\bar{x}(\lambda, \omega)\big)e$ *обладает свойствами*

$$\bar{y}_\lambda > 0, \quad A^T \bar{y}_\lambda = c, \quad \limsup_{\omega \to +0,\ \omega\lambda \to +\infty} (b, \bar{y}_\lambda) \le \bar{v}.$$

Доказательство. Выберем последовательности λ_k и ω_k такие, что $\omega_k \to +0$, $\omega_k \lambda_k \to +\infty$, и обозначим для краткости $\bar{x}^k = \bar{x}(\lambda_k, \omega_k)$ и $\bar{y}^k = U(\bar{x}^k, \lambda_k, \omega_k)\, e$. Поскольку соотношения $A^T \bar{y}^k = c$, $\bar{y}^k > 0$ очевидны, то достаточно показать, что $\limsup_k\, (b, \bar{y}^k) \le \bar{v}$.

Для этого вначале вспомним, что согласно предыдущим утверждениям последовательность $\bar{x}^k$ сходится по расстоянию к оптимальному множеству исходной задачи, которое по предположению ограничено. Поэтому ограничена и сама последовательность $\bar{x}^k$, следствием чего является конечность величины $\mathcal{C}_3 := \inf_k \sum_{j \in J_-(\bar{x}^k)} h'(l_j(\bar{x}^k, \lambda_k) l_j(\bar{x}^k)$. Далее имеем простую цепочку равенств и неравенств

$$(b, \bar{y}^k) = \pm\big(A\bar{x}^k, \bar{y}^k\big) + (b, \bar{y}^k) = \big(\bar{x}^k,\ A^T \bar{y}^k\big) - \big(A\bar{x}^k - b,\ \bar{y}^k\big) = (c, \bar{x}^k) - \big(A\bar{x}^k - b,\ \bar{y}^k\big)$$
$$= (c, \bar{x}^k) - \sum_{j \in J_+(\bar{x}^k)} \underbrace{\omega_k h'\big(l_j(\bar{x}^k, \lambda_k) l_j(\bar{x}^k)}_{\ge 0} - \sum_{j \in J_-(\bar{x}^k)} \omega_k h'\big(l_j(\bar{x}^k, \lambda_k) l_j(\bar{x}^k) \le (c, \bar{x}^k) - \omega_k \mathcal{C}_3.$$

Переходя здесь к пределу, выводим искомое неравенство

$$\limsup_k\, (b, \bar{y}^k) \le \bar{v}.$$

Утверждение доказано.

Следствие 1. *При дополнительном условии ограниченности оптимального множества двойственной задачи* (1.2) *последовательность точек* $\bar{y}^k$ *сходится к нему по расстоянию.*

4. Несобственные задачи ЛП 1-го рода

Рассмотрим отдельно случай, когда задача (1.1) — несобственная 1-го рода. Такие задачи не имеют обычного решения вследствие пустоты своей допустимой области, но могут быть приведены к разрешимому виду путем корректировки правых частей своих ограничений. В соответствии с общепринятой методологией [11] погрузим исходную задачу в параметрическое семейство задач вида

$$\max\{(c,x)\colon Ax \le b + \Delta b\}, \tag{4.1}$$

где $\Delta b \in \mathbb{R}^m$ — вектор параметров коррекции. Обозначим через $M(\Delta b)$ допустимое множество этой задачи и введем в рассмотрение вектор оптимальной (минимальной в евклидовой норме) коррекции правых частей ее ограничений

$$\bar{\Delta b} = \operatorname{Arg\,min}\{\|\Delta b\|\colon M(\Delta b) \ne \varnothing\}.$$

Как известно, вектор $\bar{\Delta b}$ определяется единственным образом (как проекция нуля на выпуклое замкнутое множество, состоящее из всех векторов коррекции, приводящих ограничения исходной задачи к совместному виду), а допустимое множество оптимально скорректированной задачи

$$\max\{(c,x)\colon Ax \le b + \bar{\Delta b}\} \tag{4.2}$$

на самом деле совпадает с множеством точек минимума популярной квадратичной штрафной функции $M(\bar{\Delta b}) = \operatorname{Arg\,min}_x (1/2)\big\|(Ax-b)^+\big\|^2$.

Предположение о разрешимости скорректированной задачи (4.2) влечет за собой разрешимость двойственной к ней задачи

$$\min\{(b+\bar{\Delta b}, y)\colon A^T y = c, \quad y \ge 0\}, \tag{4.3}$$

оптимальное значение которой будет тем же, что и у прямой задачи. По-прежнему будем обозначать это общее оптимальное значение через $\bar{v}$, а произвольно выбранные оптимальные векторы этих задач — через $\bar{x}$ и $\bar{y}$ соответственно. Последние связаны соотношениями

$$\big(\bar{x}, c - A^T\bar{y}\big) = 0, \quad \big(\bar{y}, A\bar{x} - b - \bar{\Delta b}\big) = 0, \quad (c,\bar{x}) = (b + \bar{\Delta b}, \bar{y}) = \bar{v}.$$

Разумеется, в собственном случае $\bar{\Delta b} = 0$. В несобственном случае, к сожалению, вектор $\bar{\Delta b}$ заранее не известен, но, как мы покажем ниже, его можно автоматически получить, применяя к исходной задаче (1.1) метод композитных штрафных функций и не внося при этом в алгоритм никаких явных изменений (тем самым алгоритм композитных функций становится универсальным).

Утверждение 3. *Пусть задача (1.1) — несобственная 1-го рода, т. е. аппроксимирующая ее задача (4.2) разрешима, и пусть среди допустимых планов двойственной к ней задачи (4.3) имеются положительные. Пусть также выполнены свойства A1–A3. Тогда точная верхняя грань в (1.3) достигается всякий раз в единственной точке $\bar{x}(\lambda,\omega)$ и последовательность таких точек, построенная по серии значений $\omega_k \to +0$, $\omega_k\lambda_k \to +\infty$, сходится к оптимальному множеству аппроксимационной задачи (4.2).*

Доказательство. Дальнейшие рассуждения, чуть более сложные по сравнению с использованными выше, вновь разобьем на три этапа.

Этап 1. Разберем поведение оптимума в (1.3). Прежде всего, ввиду свойств A1, A2 имеем

$$\begin{aligned}\sup_x \mathcal{H}(x,\lambda,\omega) \ge \mathcal{H}(\bar{x},\lambda,\omega) = (c,\bar{x}) - H(A\bar{x} - b, \lambda,\omega) \ge (c,\bar{x}) - H\big(\bar{\Delta b},\lambda,\omega\big) \\ \ge \bar{v} - \omega\|\bar{\Delta b}\|\|e\| - \omega\lambda\|\bar{\Delta b}\|^2.\end{aligned} \tag{4.4}$$

Отсюда
$$\liminf_{\omega\to+0,\ \omega\lambda\to\infty}\Big(\frac{1}{\omega\lambda}\sup_x \mathcal{H}(x,\lambda,\omega)\Big)\geq -\|\bar\Delta b\|^2. \tag{4.5}$$

Чтобы получить обратное неравенство, выберем в допустимом множестве двойственной задачи (4.3) какой-нибудь положительный вектор y и зафиксируем его. Тем самым
$$A^Ty=c,\quad y>0,\quad (b+\bar\Delta b,y)\geq \bar v.$$

Поскольку нас интересует ситуация, когда $\omega\lambda\to+\infty$, то, не снижая общности, можно считать, что $\omega\lambda>\max\big\{1,\ \max_j(y_j),\ \|y\|/\|\bar\Delta b\|\big\}$. Построение обратной оценки начнем с равенства
$$\mathcal{H}(x,\lambda,\omega)=(c,x)-H(Ax-b,\lambda,\omega)=(y,Ax-b)+(b,y)-H(Ax-b,\lambda,\omega).$$

Разобьем множество индексов $J=\{1,2,\dots,m\}$ на два непересекающихся индексных подмножества $J_+(x)=\{j\in J\colon l_j(x)>b_j\}$ и $J_-(x)=\{j\in J\colon l_j(x)\leq b_j\}$. В результате имеем
$$\mathcal{H}(x,\lambda,\omega)=(b,y)+\sum_{j\in J_+(x)}\big[y_jl_j(x)-\omega h(l_j(x),\lambda)\big]+\sum_{j\in J_-(x)}\big[y_jl_j(x)-\omega h(l_j(x),\lambda)\big],$$
где слагаемые каждой группы вычисляются по своим формулам. Применяя свойство A3, сумму "отрицательных" слагаемых оценим сверху как
$$\sum_{j\in J_-(x)}\big[y_jl_j(x)-\omega h(l_j(x),\lambda)\big]\leq\sum_{j=1}^m\mathcal{C}_1(y_j,\lambda,\omega).$$

Сумму "положительных" слагаемых оценим сверху путем выделения полных квадратов как
$$\sum_{j\in J_+(x)}\big[y_jl_j(x)-\omega h(l_j(x),\lambda)\big]\leq(y,(Ax-b)^+)-\lambda\omega\|(Ax-b)^+\|^2$$
$$\leq\ \|y\|\big\|(Ax-b)^+\big\|-\lambda\omega\big\|(Ax-b)^+\big\|^2=\|y\|\big\|(Ax-b)^+\big\|-\lambda\omega\big\|(Ax-b)^+\big\|^2\pm\frac{\|y\|^2}{4\lambda\omega}$$
$$=\frac{\|y\|^2}{4\lambda\omega}-\lambda\omega\Big(\big\|(Ax-b)^+\big\|-\frac{\|y\|}{2\lambda\omega}\Big)^2\leq\frac{\|y\|^2}{4\lambda\omega}-\lambda\omega\Big(\|\bar\Delta b\|-\frac{\|y\|}{2\lambda\omega}\Big)^2.$$
Здесь мы учли свойство A2 и монотонный рост квадратичной функции на положительной оси, а также то, что при сделанных предположениях относительно параметра λ верно неравенство
$$\big\|(Ax-b)^+\big\|-\frac{\|y\|}{2\lambda\omega}\geq\|\bar\Delta b\|-\frac{\|y\|}{2\lambda\omega}>0.$$

В итоге при всех достаточно больших значениях произведения $\omega\lambda$
$$\mathcal{H}(x,\lambda,\omega)\leq(b,y)+\frac{\|y\|^2}{2\lambda\omega}-\lambda\omega\Big(\|\bar\Delta b\|-\frac{\|y\|}{\lambda\omega}\Big)^2+\sum_{j=1}^m\mathcal{C}_1(y_j,\lambda,\omega).$$

Поскольку правая часть этого неравенства не зависит от x, то и
$$\sup_x\mathcal{H}(x,\lambda,\omega)\leq(b,y)+\frac{\|y\|^2}{2\lambda\omega}-\lambda\omega\Big(\|\bar\Delta b\|-\frac{\|y\|}{\lambda\omega}\Big)^2+\sum_{j=1}^m\mathcal{C}_1(y_j,\lambda,\omega),$$
а значит, с учетом свойства A3, после перехода здесь к пределу, получаем
$$\limsup_{\omega\to+0,\ \omega\lambda\to\infty}\Big(\frac{1}{\lambda\omega}\sup_x\mathcal{H}(x,\lambda,\omega)\Big)\leq-\|\bar\Delta b\|^2.$$

Принимая во внимание (4.5), окончательно заключаем, что

$$\lim_{\omega\to+0,\ \omega\lambda\to+\infty}\Big(\frac{1}{\lambda\omega}\sup_x \mathcal{H}(x,\lambda,\omega)\Big) = -\|\bar{\Delta b}\|^2.$$

Этап 2. Докажем, что при исходных предположениях множества уровня функции $\mathcal{H}(x,\lambda,\omega)$ ограничены. Напомним, что у нас есть положительная точка y, для которой $c = A^T y$. Сформируем одно такое множество:

$$\Omega_K^{\lambda,\omega} := \big\{x\in\mathbb{R}^n\colon \mathcal{H}(x,\lambda,\omega) = (c,x) - H(Ax-b,\lambda,\omega) \geq K\big\},$$

где K — достаточно малое число, обеспечивающее непустоту такого множества. В частности, будем полагать $K < \bar{v} - \max\big(\lambda\omega\|\bar{\Delta b}\|^2, \|y\|/\|\bar{\Delta b}\|\big)$.

Обосновывая ограниченность выписанного множества, как и ранее, будем опираться на то, что при наличии относительно внутренних точек в допустимом множестве двойственной задачи (4.3) любое непустое множество вида

$$\bar{M}_{\Delta b}^{\delta} = \{x\in\mathbb{R}^n\colon Ax\leq b+\Delta b,\ (c,x)\geq \bar{v}-\delta\} \tag{4.6}$$

будет ограниченным. Покажем, как в этом случае по параметрам K, ω и λ можно подобрать Δb и δ, обеспечивающие включение $\Omega_K^{\lambda,\omega}\subseteq \bar{M}_{\Delta b}^{\delta}$.

Итак, пусть $x\in\Omega_K^{\lambda,\omega}$. Начнем с оценки для невязки прямых ограничений:

$$\lambda\omega\|(Ax-b)^+\|^2 \leq -\mathcal{H}(x,\lambda,\omega) + (c,x) - \sum_{j\in J_-(x)}\omega h(l_j(x),\lambda)$$

$$= -\mathcal{H}(x,\lambda,\omega) + (b+\bar{\Delta b}, y) - (\bar{\Delta b}, y) + (y, Ax-b) - \sum_{j\in J_-(x)}\omega h(l_j(x),\lambda)$$

$$\leq -K + \bar{v} + (y, Ax-b) - \sum_{j\in J_-(x)}\omega h(l_j(x),\lambda)$$

$$= -K + \bar{v} + (y,(Ax-b)^+) + \sum_{j\in J_-(x)}\big[y_j l_j(x) - \omega h(l_j(x),\lambda)\big]$$

$$\leq -K + \bar{v} + \|y\|\,\big\|(Ax-b)^+\big\| + \sum_{j=1}^m \mathcal{C}_1(y_j,\lambda,\omega).$$

Здесь мы воспользовались свойствами A1–A3 и особенностями вектора y. В результате перед нами оказалось квадратное неравенство

$$\lambda\omega t^2 - \|y\|\,t - K_1(y,\lambda,\omega) \leq 0$$

применительно к числовой переменой $t = \|(Ax-b)^+\|$, где в силу сделанных предположений относительно K и теоремы слабой двойственности $K_1(y,\lambda,\omega) = -K+\bar{v}+\sum_{j=1}^m \mathcal{C}_1(y_j,\lambda,\omega) > 0$. Отсюда вытекает искомая оценка сверху для

$$\|(Ax-b)^+\| = t \leq K_2(y,\lambda,\omega) := \frac{\|y\| + \sqrt{\|y\|^2 + 4\lambda\omega K_1(y,\lambda,\omega)}}{2\lambda\omega}. \tag{4.7}$$

Это позволяет нам определить подходящее Δb для выполнения включения $\Omega_K^{\lambda,\omega}\subseteq\bar{M}_{\Delta b}^{\delta}$, а именно $\Delta b = K_2(y,\lambda,\omega)\,e$.

Перейдем к подбору подходящего δ. Для этого построим оценку снизу для величины (c,x). Во-первых, возьмем ее грубую оценку вида

$$(c,x) = \mathcal{H}(x,\lambda,\omega) + H(Ax-b,\lambda,\omega) \geq K + \lambda\omega\big\|\bar{\Delta b}\big\|^2 + \sum_{j\in J_-(x)}\omega h\big((l_j(x),\lambda\big).$$

Во-вторых, по свойствам вектора y имеем равенство

$$(c,x)=(y,Ax-b)+(y,b)=(b,y)+\sum_{j\in J_+(x)} y_j l_j(x)+\sum_{j\in J_-(x)} y_j l_j(x).$$

Умножим первое соотношение на $(\lambda+1)/\lambda>1$, а второе — на $-1/\lambda$, после чего сложим их. С учетом свойства A3 получим

$$(c,x)\ge\frac{\lambda+1}{\lambda}\big(K+\lambda\omega\|\bar{\Delta b}\|^2\big)$$

$$-\frac{\lambda+1}{\lambda}\sum_{j\in J_-(x)}\Big(\frac{1}{\lambda+1}y_j l_j(x)-\omega h\big(l_j(x),\lambda\big)\Big)-\frac{1}{\lambda}\sum_{j\in J_+(x)} y_j l_j(x)-\frac{1}{\lambda}(y,b)$$

$$\ge\frac{\lambda+1}{\lambda}\big(K+\lambda\omega\|\bar{\Delta b}\|^2\big)-\frac{\lambda+1}{\lambda}\sum_{j=1}^{m}\mathcal{C}_2(y_j,\lambda,\omega)-\frac{1}{\lambda}\|y\|K_2(y,\lambda,\omega)e-\frac{1}{\lambda}(y,b). \tag{4.8}$$

Согласно найденным ранее оценкам для Δb и для $\|(Ax-b)^+\|$ выводим формулу определения параметра

$$\delta=\bar{v}-\frac{\lambda+1}{\lambda}\Big[K+\lambda\omega\|\bar{\Delta b}\|^2-\sum_{j=1}^{m}\mathcal{C}_2(y_j,\lambda,\omega)\Big]-\frac{\|y\|\|b\|}{\lambda}+\frac{1}{\lambda}\|y\|K_2(y,\lambda,\omega),$$

при котором $(c,x)\ge\bar{v}-\delta$ и, следовательно, $\Omega_K^{\lambda,\omega}\subseteq\bar{M}^{\delta}_{\Delta b}$.

Итак, множества уровня $\Omega_K^{\lambda,\omega}$ оказались частью ограниченных множеств вида (4.6) и в силу непрерывности и строгой вогнутости функции $\mathcal{H}(x,\lambda,\omega)$ верхняя граница в (1.3) конечна и достигается всякий раз в единственной точке $\bar{x}(\lambda,\omega)$, зависящей от λ, ω.

Этап 3. Осталось показать сходимость таких точек к оптимальному множеству задачи (4.1) при условии $\omega\to+0,\ \omega\lambda\to+\infty.$

Пусть выбраны последовательности $\omega_k\to+0,\ \omega_k\lambda_k\to+\infty$, и пусть $\bar{x}_k$ — отвечающая им последовательностей точек максимума функций $\mathcal{H}(x,\lambda_k,\omega_k)$. По первым двум последовательностям выстроим последовательность уровней

$$N_k=\bar{v}-\omega_k\|\bar{\Delta b}\|\|e\|-\lambda_k\omega_k\|\bar{\Delta b}\|^2-\varepsilon_k,$$

где ε_k — произвольная последовательность положительных чисел, $\varepsilon_k\to+0$.

Для завершения доказательства заметим, что в силу (4.4) все $\bar{x}_k\in\Omega_{N_k}^{\lambda_k,\omega_k}$, и потому исходя из соотношений (4.7) и (4.8) имеем

$$\|(A\bar{x}_k-b)^+\|\le\frac{\|y\|}{2\lambda_k\omega_k}+\sqrt{\Big(\frac{\|y\|}{2\lambda_k\omega_k}\Big)^2+\frac{1}{\lambda_k\omega_k}\Big[(y,b)-N_k+\sum_{j=1}^{m}\mathcal{C}_1(y_j,\lambda_k,\omega_k)\Big]}\ \to\|\bar{\Delta b}\|,$$

$$(c,\bar{x}_k)\ge\frac{\lambda_k+1}{\lambda_k}\Big[N_k+\lambda_k\omega_k\|\bar{\Delta b}\|^2-\sum_{j=1}^{m}\mathcal{C}_2(y_j,\lambda_k,\omega_k)\Big]-\frac{\|y\|\big(\|b\|+\|(A\bar{x}_k-b)^+\|\big)}{\lambda_k}\to\bar{v}.$$

Утверждение доказано полностью.

5. Результаты численных экспериментов

Эксперименты проводились на вычислительном комплексе "УРАН" Института математики и механики им. Н. Н. Красовского УрО РАН в среде MATLAB.

Целью экспериментов было сравнение эффективности применения метода Ньютона к стандартной логарифмической функции барьера (алгоритм 0) с его применением к предлагаемым в статье гладким штрафным функциям $H_1(\cdot, \lambda)$, $H_2(\cdot, \lambda)$ и $H_3(\cdot, \lambda)$ (соответственно алгоритмы 1, 2 и 3). Работа алгоритмов сравнивалась на тестовых задачах ЛП с ограничениями типа неравенств. Матрицы ограничений генерировались при помощи датчика случайных чисел с равномерным распределением от -1 до 1. Размерности задач колебались от 100×300 до 1000×3000. Заполненность матриц ненулевыми элементами составляла 3—5%. Правые части ограничений и коэффициенты целевой функции подбирались таким образом, чтобы решение задачи было единственным и совпадало с некоторым заданным заранее. Также обеспечивалось наличие внутренней точки у допустимой области прямой задачи. Последнее условие необходимо для применения алгоритма 0.

Таким образом, каждая задача решалась четыре раза. Стартовые приближения во всех случаях брались одинаковыми (это внутренняя точка допустимой области прямой задачи). Величина штрафа не менялась на протяжении одного испытания.

Типовые результаты расчетов представлены в табл. 1, 2. В первой колонке приведен десятичный порядок штрафного параметра λ. Во второй — показан десятичный порядок нормы градиента оптимизируемой функции, при достижении которого процесс останавливался. В остальных колонках указаны показатели эффективности четырех исследуемых методов.

Показатели эффективности алгоритмов оценивались тремя числами, первое и второе из которых показывают десятичные порядки точности полученных решений прямой и двойственной задач (этот порядок, как правило, у обоих решений совпадал), а третье — потребовавшееся

Таблица 1

Результаты экспериментов с задачами ЛП размерности 100×300

Штраф*	Норма $\nabla\mathcal{H}^*$	Алгоритм 0	Алгоритм 1	Алгоритм 2	Алгоритм 3
1	−15	0, 0, 13	−1, −1, 14	0, 0, 13	1, 1, 18
2	−14	−1, −1, 15	−2, −2, 15	−1, −1, 15	0, 0, 21
3	−14	−2, −2, 17	−3, −3, 18	−2, −2, 18	−1, 0, 26
4	−13	−3, −3, 18	−4, −4, 22	−3, −3, 21	−2, −1, 27
5	−12	−4, −4, 23	−5, −5, 25	−4, −4, 21	−3, −2, 28
6	−11	−5, −5, 27	–	−5, −5, 21	−3, −2, 31
7	−10	−6, −6, 29	–	−6, −6, 23	−4, −3, 32

∗ Приведен десятичный порядок

Таблица 2

Результаты экспериментов с задачами ЛП размерности 1000×3000

Штраф*	Норма $\nabla\mathcal{H}^*$	Алгоритм 0	Алгоритм 1	Алгоритм 2	Алгоритм 3
1	−12	0, 0, 22	−1, −1, 24	0, 0, 23	2, 3, 25
2	−12	0, 0, 28	−1, −1, 39	0, 0, 26	1, 1, 29
3	−11	−1, −1, 34	−2, −2, 50	−1, −1, 37	0, 0, 41
4	−11	−2, −2, 45	−3, −3, 35	−2, −2, 41	−1, 0, 43
5	−10	−3, −3, 51	−4, −4, 39	−3, −3, 49	−2, −1, 53
6	−10	−4, −4, 54	–	−5, −5, 50	−3, −2, 55
7	−9	−5, −5, 57	–	−5, −5, 53	−4, −3, 59

∗ Приведен десятичный порядок

для этого число итераций метода Ньютона. Следует отметить, что трудоемкость одной итерации каждого из методов примерно одинакова. Укажем также, что слишком локальную степень сходимости при росте штрафного параметра продемонстрировал только алгоритм 1.

Как видно из приведенных данных, первые две штрафные конструкции практически не уступают логарифмическим барьерным функциям. Третья функция имеет более широкий радиус квадратичной скорости сходимости, но из-за необходимости уменьшать масштабный множитель до нуля уступает первым двум функциям по точности получаемого решения, особенно двойственного.

Заключение

Подведем итоги. Приведенные в статье принципы построения композитных штрафных функций путем гладкой “склейки” графиков различных вариантов внешних и внутренних штрафов позволили объединить положительные качества двух названных направлений. В частности оказалось, что новые конструкции применимы не только к регулярным, но и к некорректно поставленным задачам, для которых они находят их обобщенные решения, несущие информацию об обнаруженных в модели противоречиях и о степени их остроты. При этом проведенный автором дополнительный вычислительный эксперимент подтвердил отмечавшуюся ранее в [20] эффективность обсуждаемого подхода, вполне сопоставимую с эффективностью обычного метода логарифмическихо барьеров, послужившего, как известно, прообразом современных методов центрального пути.

СПИСОК ЛИТЕРАТУРЫ

1. **Еремин И.И.** Метод штрафов в выпуклом программировании // Докл. АН СССР. 1967. Т. 173, № 4. С. 748–751.
2. **Зангвилл У.И.** Нелинейное программирование. Единый подход. М.: Сов. радио, 1973. 312 с.
3. **Osborne M.R., Ryan D.M.** On penalty function methods for nonlinear programming problems // J. Math. Anal, and Appl. 1970. Т. 31, № 3. P. 559–579. https://doi.org/10.1016/0022-247X(70)90009-0
4. **Фиакко А., Мак-Кормик Г.** Нелинейное программирование. Методы последовательной безусловной минимизации. М.: Мир, 1972. 240 с.
5. **Полак Э.** Численые методы оптимизации. М.: Мир, 1974. 376 с.
6. **Еремин И.И., Астафьев Н.Н.** Введение в теорию линейного и выпуклого программирования. М.: Наука, 1976. 192 с.
7. **Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М.** Методы оптимизации. М.: Наука, 1978. 351 с.
8. **Васильев Ф.П.** Численные методы решения экстремальных задач. М.: Наука, 1980. 520 с.
9. **Евтушенко Ю.Г.** Методы решения экстремальных задач и их применение в системах оптимизации. М.: Наука, 1982. 432 с.
10. **Поляк Б.Т.** Введение в оптимизацию. М.: Наука, 1983. 384 с.
11. **Еремин И.И., Мазуров Вл.Д., Астафьев Н.Н.** Несобственные задачи линейного и выпуклого программирования. М.: Наука, 1983. 336 с.
12. **Еремин И.И.** Противоречивые модели оптимального планирования. М.: Наука, 1988. 160 с.
13. **Гольштейн Е.Г., Третьяков Н.В.** Модифицированные функции Лагранжа. Теория и методы оптимизации. М.: Наука, 1989. 400 с. (Серия "Экономико-математическая библиотека").
14. **Скарин В. Д.** Метод штрафных функций и регуляризация в анализе несобственных задач выпуклого программирования // Тр. Ин-та математики и механики УрО РАН. 2018. Т. 24, № 3. С. 187–199. https://doi.org/10.21538/0134-4889-2018-24-3-187-199
15. **Бирюков А.Г., Чернов А.В., Чернова Ю.Г., Шароватова Ю.И.** Штрафные, барьерные, квазибарьерные функции и функции, обратные к ним // Интеллектуальные системы. Теория и приложения. 2018. Т. 22, № 4. С. 31–50.
16. **Renegar J.** A polynomial-time algorithm, based on Newton’s method, for linear programming // Mathematical Programming. 1988. Vol. 40. P. 59–93. https://doi.org/10.1007/BF01580724

17. **Gonzaga C.C.** Interior point algorithms for linear programming problems with inequality constraints // Mathematical Programming. 1991. Vol. 52. P. 209–225. https://doi.org/10.1007/BF01582888
18. **Roos C., Terlaky T., Vial J.-Ph.** Theory and algorithms for linear optimization. Chichester: John Wiley & Sons Ltd, 1997. 484 p.
19. **Попов Л.Д.** Барьеры и симметричная регуляризация функции Лагранжа при анализе несобственных задач линейного программирования // Тр. Ин-та математики и механики УрО РАН. 2023. Т. 29, № 3. С. 138–155. https://doi.org/10.21538/0134-4889-2023-29-3-138-155
20. **Popov L.D.** On the application of composite penalty functions obtained by “gluing” of external penalties with barrier ones in linear programming // Optimization and Applications: Proc. 15th Internat. Conf. (OPTIMA 2024) / eds. Nicholas Olenev, Yuri Evtushenko, Milojica Jaćimović, Michael Khachay, Vlasta Malkova. Cham: Springer, 2025. P. 43–57. (Ser. Lecture Notes in Computer Science; vol. 15218). https://doi.org/10.1007/978-3-031-79119-2_4
21. **Гилл Ф., Мюррей У., Райт М.** Практическая оптимизация. М.: Мир, 1985. 509 с.
22. **Kanzow C., Qi H., Qi L.** On the minimum norm solution of linear program // J. Optimi. Theory Appl. 2003. Vol. 116. P. 333–345.https://doi.org/10.1023/A:1022457904979
23. **Mangasarian O.L.** A Newton method for linear programming // J. Optim. Theory Appl. 2004. Vol. 121. P. 1–18. https://doi.org/10.1023/B:JOTA.0000026128.34294.77

Поступила 23.04.2025
После доработки 16.06.2025
Принята к публикации 23.06.2025

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

REFERENCES

1. Eremin I.I. The “penalty” method in convex programming. *Sov. Math., Dokl.*, 1967, vol. 8, pp. 459–462.
2. Zangwill W. *Nonlinear programming: a unified approach*, Englewood Cliffs, Prentice-Hall, 1969, 356 p. ISBN: 978-0136235798 . Translated to Russian under the title *Nelineinoe programmirovanie: edinyi podkhod*, Moscow, Soviet Radio Publ., 1973, 312 p.
3. Osborne M.R., Ryan D.M. On penalty function methods for nonlinear programming problems. *J. Math. Anal. Appl.*, 1970, vol. 31, no. 3, pp. 559–579. https://doi.org/10.1016/0022-247X(70)90009-0
4. Fiacco A.V., McCormick G.P. *Nonlinear programming: sequential unconstrained minimization techniques.* Classics in applied mathematics, vol. 4. Philadelphia, SIAM, 1987, 226 p. https://doi.org/10.1137/1.9781611971316 . Translated to Russian under the title *Nelineinoe programmirovanie. Metody posledovatel’noi bezuslovnoi minimizatsii.* Moscow, Mir Publ., 1972, 240 p.
5. Polak E. *Computational methods in optimization: a unified approach.* Mathematics in Science and Engineering Ser., vol. 77. NY, London, Acad. Press, 1971, 329 p. ISBN-10: 0125593503 . Translated to Russian under the title *Chislennyye metody optimizatsii: yedinyy podkhod*. Moscow, Mir Publ., 1974, 376 p.
6. Eremin I.I., Astaf’ev N.N. *Vvedenie v teoriyu lineinogo i vypuklogo programmirovaniya* [Introduction to the theory of linear and convex programming]. Moscow, Nauka Publ., 1976, 196 p.
7. Moiseev N.N., Ivanilov Yu.P., Stolarova E.M. *Metody optimizacii* [Optimization methods]. Moscow, Nauka Publ., 1978, 351 p.
8. Vasil’ev F.P. *Metody resheniya ekstremal’nykh zadach* [Methods for solving extremal problems]. Moscow, Nauka Publ., 1980, 520 p.
9. 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.
10. Polyak B.T. *Introduction to optimization.* New York, Division Publ., 1987, 438 p. ISBN-10: 0911575146 . Original Russian text published in Polyak B. T. *Vvedeniye v optimizatsiyu*, Moscow, Nauka Publ., 1983, 384 p.

11. Eremin I.I., Mazurov V.D., Astaf'ev N.N. *Nesobstvennye zadachi lineinogo i vypuklogo programmirovaniya* [Improper problems of linear and convex programming], Moscow, Nauka Publ., 1983, 336 p.
12. Eremin I.I. *Protivorechivye modeli optimal'nogo planirovaniya* [Contradictory models of optimal planning], Moscow, Nauka Publ., 1988, 160 p. ISBN: 5-02-013773-1 .
13. Golshtein E.G., Tret'yakov N.V. *Modifitsirovannyye funktsii Lagranzha. Teoriya i metody optimizatsii* [Modified Lagrange functions. Theory and methods of optimization]. Moscow, Nauka Publ., 1989, 400 p. ISBN: 5-02-013965-3 .
14. Skarin V.D. Penalty function method and regularization in the analysis of improper convex programs. *Proc. Steklov Inst. Math.*, 2019, vol. 305, suppl. 1, pp. S166–S177. https://doi.org/10.1134/S0081543819040175
15. Birjukov A.G., Chernov A.V., Chernova Yu.G., Sharovatova Yu.I. Penalty, barrier, quasi-barrier functions and functions inverse to them. *Intellektual'nyye sistemy. Teoriya i prilozheniya*, 2018, vol. 22, no. 4, pp. 31–50 (in Rusian).
16. Renegar J. A polynomial-time algorithm, based on Newton's method, for linear programming. *Math. Programm.*, 1988, vol. 40, pp. 59–93. https://doi.org/10.1007/BF01580724
17. Gonzaga C.C. Interior point algorithms for linear programming problems with inequality constraints. *Math. Programm.*, 1991, vol. 52, pp. 209–225. https://doi.org/10.1007/BF01582888
18. Roos C., Terlaky T., Vial J.-Ph. *Theory and algorithms for linear optimization: an interior point approach*. Chichester, NY, Wiley, 1997, 482 p.
19. Popov L.D. Barriers and symmetric regularization of the lagrange function in the analysis of improper linear programming problems. *Trudy Inst. Mat. Mekh. UrO RAN*, 2023, vol. 29, no. 3, pp. 138–155 (in Russian). https://doi.org/10.21538/0134-4889-2023-29-3-138-155
20. Popov L.D. On the application of composite penalty functions obtained by "gluing" of external penalties with barrier ones in linear programming. In: *Proc. 15th Inter. Conf. "Optimization and applications"* (*OPTIMA 2024*), Ser. Lecture Notes in Comp. Sci., vol. 15218, Cham, Springer, 2025, pp. 43–57. https://doi.org/10.1007/978-3-031-79119-2_4
21. Gill P., Murray W., Wright M. *Practical optimization*. London, NY, Toronto, Sydney, San Francisco, Academic Press, 1981, 401 p. Translated to Russian under the title *Prakticheskaya optimizatsiya*, Moscow, Mir Publ., 1985, 509 p.
22. Kanzow C., Qi H., Qi L. On the minimum norm solution of linear program. *J. Optim. Theory Appl.*, 2003, vol. 116, pp. 333–345. https://doi.org/10.1023/A:1022457904979
23. Mangasarian O.L. A Newton method for linear programming. *J. Optim. Theory Appl.*, 2004, vol. 121, pp. 1–18. https://doi.org/10.1023/B:JOTA.0000026128.34294.77

Received April 23, 2025
Revised June 16, 2025
Accepted June 23, 2025

Funding Agency: The work was performed as part of research conducted in the Ural Mathematical Center with the financial support of the Ministry of Science and Higher Education of the Russian Federation (Agreement number 075-02-2025-1549) .

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: e-mail: popld@imm.uran.ru .

Cite this article as: L. D. Popov. Principles of construction and properties of penalty functions of composite type (on the example of linear programming problems). *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2025, vol. 31, no. 3, pp. 215–232 .