

МЕТОДЫ ОПТИМИЗАЦИИ

МЕТОД УРОВНЕЙ, ЕГО ОБОБЩЕНИЯ И ПРИЛОЖЕНИЯ*

© 1995 г. Гольштейн Е.Г., Немировский А.С., Нестеров Ю.Е.

(Москва)

Описан так называемый метод уровней, предназначенный для минимизации негладких выпуклых функций. Метод перенесен на задачи отыскания седловой точки выпукло-вогнутой функции и выпуклого программирования с ограничениями. Построена двойственная декомпозиционная схема анализа задач линейного программирования специальной структуры, основанная на методе уровней. Приведены результаты вычислительных экспериментов.

1. ВВЕДЕНИЕ

Метод уровней относится к семейству так называемых bundle-алгоритмов негладкой выпуклой минимизации. Эти алгоритмы предназначены для решения задачи

$$f(x) \rightarrow \min, x \in G, \quad (1)$$

где G — выпуклый многогранник диаметра D в n -мерном евклидовом пространстве E^n , а f — выпуклая липшицева с константой L функция на G .

Специфической особенностью bundle-методов в отличие от более простых (и соответственно менее эффективных на практике) методов субградиентного спуска является использование кусочно-линейной модели

$$f_i(x) = \max_{1 \leq j \leq i} [f(x_j) + (x - x_j)^T f'(x_j)] \quad (2)$$

минимизируемой функции ($x_j \in G$ — итеративно формируемые *поисковые точки*, $f(x_j)$; $f'(x_j)$ — вычисляемые в этих точках значения и субградиенты целевого функционала). Достоинство модели состоит в том, что она аккумулирует всю полученную к соответствующему моменту информацию о f . Для сравнения укажем, что в методах субградиентного типа на итерации i используется лишь последний (отвечающий $j = i$) элемент модели ($f(x_i)$; $f'(x_i)$).

Исторически первым bundle-методом и, пожалуй, первым методом негладкой выпуклой минимизации вообще был метод секущей плоскости Келли [1], задаваемый весьма естественной рекурсией

$$x_{i+1} \in \operatorname{Arg} \min_{x \in G} f_i(x). \quad (3)$$

Оказалось, однако, что сходимость метода Келли может быть — и достаточно часто оказывается на практике — весьма медленной. Дело в том, что в многомерном случае точка минимума модели f_i крайне неустойчива ("почти случайна"). Существенный

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 93-012-499).

прогрессе в качестве методов был связан с предложенным К. Лемарешалем ([2, 3]) переходом от "слишком оптимистического" правила (3) к более осторожному правилу

$$x_{i+1} = \arg \min_{x \in G} \left[f_i(x) + \frac{d_i}{2} |x - \bar{x}_i|^2 \right], \quad (4)$$

в котором вычислительный процесс стабилизируется за счет введения квадратичного члена, штрафующего большие отклонения от $\bar{x}_i$ — "наиболее перспективной" из пройденных точек. Заметим, что (4) восходит к так называемому *prox*-алгоритму минимизации, предложенному Б. Мартине [4] на основе идей Ж. Моро [5] и задаваемом соотношением

$$x_{i+1} = \arg \min_{x \in G} \left[f(x) + \frac{d_i}{2} |x - x_i|^2 \right],$$

где f — минимизируемая на G функция. Алгоритмы типа (4) (при надлежаще выбранных правилах подстройки штрафного коэффициента d_i и пересчета центра регуляризации $\bar{x}_i$) весьма эффективны на практике. Вычислительный опыт показывает, однако, что *bundle*-методы обладают уязвимой точкой: их практическая эффективность весьма существенно зависит от применяемых правил подстройки d_i , причем правила, успешно применяемые для одних задач, могут быть неподходящими для других.

Один из возможных подходов к синтезу "робастных" *bundle*-методов связан с некоторым неявным контролем штрафного коэффициента d_i , предложенным в [6]. Соответствующий метод, получивший название "метод уровней", и является предметом настоящей работы. Ниже излагается базовый вариант метода, который затем переносится на задачу отыскания седловой точки выпукло-вогнутых функций и задачу выпуклого программирования с ограничениями. Для всех трех вариантов метода уровней устанавливается скорость сходимости. Далее на основе описанного метода строится декомпозиционная схема анализа задач линейного программирования (ЛП) специальной структуры. В последнем разделе приводятся результаты численного тестирования метода уровней, из которых вытекает его высокая практическая эффективность.

2. МЕТОД УРОВНЕЙ ДЛЯ МИНИМИЗАЦИИ НА МНОГОГРАННИКЕ

Метод уровней, обозначаемый далее Lev_λ , определяется единственным параметром $\lambda \in (0,1)$ и решает задачу (1) следующим образом.

К началу итерации i в нашем распоряжении имеются сформированные ранее поисковые точки $x_j \in G$, $j = 1, \dots, i-1$, с уже вычисленными $f(x_j)$ и субградиенты $f'(x_j)$.

Итерация i начинается с $f(x_i)$, $f'(x_i)$ и построения модели (2) целевого функционала f . Из выпуклости f следует, что модель f_i оценивает f снизу: $f_i(x) \leq f(x)$, $x \in G$. Кроме того, ясно, что $f_i(x)$ не убывает по i , причем f и f_i совпадают в точках x_j , $j \leq i$. Далее решается задача ЛП $f_i(x) \rightarrow \min$, $x \in G$, и вычисляется минимальное значение f_i i -й модели f_i на многограннике G . В силу отмеченных свойств моделей f_i является нижней оценкой оптимального значения f^* задачи (1), причем эти оценки не убывают по i . Затем вычисляются текущее рекордное значение $\bar{f}_i$ целевого функционала f и оценка ошибки Δ_i : $\bar{f}_i = \min_{j \leq i} f(x_j)$; $\Delta_i = \bar{f}_i - f_i$.

С учетом сказанного $\Delta_i \geq \bar{f}_i - f^*$ оценивает сверху отклонение от минимума по целевому функционалу лучшей (с минимальным значением f) из сформированных поисковых точек x_j , $j \leq i$, причем Δ_i не возрастает по i .

Наконец, определяется новая поисковая точка x_{i+1} как проекция x_i на множество

уровня $l_i = \underline{f}_i + \lambda \Delta_i$ модели f_i (отсюда и название метода):

$$x_{i+1} = \arg \min \{ |x - x_i|^2 \mid x \in G, f_i(x) \leq l_i \},$$

после чего происходит переход к следующей итерации.

Приближенным решением задачи, найденным после i итераций, считается рекордная точка $\bar{x}_i$, определенная условием: $f(\bar{x}_i) = \bar{f}_i$. Метод останавливается после достижения неравенства: $\Delta_i \leq \varepsilon$, где $\varepsilon > 0$ – заранее фиксированная точность решения задачи (1).

Скорость сходимости описанного метода устанавливает следующее утверждение.

Теорема 1. При любом $\varepsilon > 0$ число шагов метода Lev_λ до построения ε -решения задачи (т.е. до момента первого выполнения неравенства $f(\bar{x}_i) - f^* \leq \varepsilon$) не превосходит

$$N(\varepsilon) = \left\lceil c(\lambda) \frac{L^2 D^2}{\varepsilon^2} \right\rceil, \quad c(\lambda) = ((1-\lambda)^2 \lambda (2-\lambda))^{-1}, \quad (5)$$

где $\lfloor z \rfloor$ – наименьшее целое число, большее либо равное вещественного числа z ; L – константа Липшица f ; D – диаметр G .

Доказательство. Поскольку $f(\bar{x}_i) - f^* \leq \Delta_i$, достаточно доказать, что условие $\Delta_i > \varepsilon$ не может выполняться более чем на $N(\varepsilon)$ шагах.

Итак, допустим N таково, что $\Delta_N > \varepsilon$, и оценим N сверху.

1°. Разобьем множество $I = \{1, \dots, N\}$ номеров итераций на группы $I_1, \dots, I_k$ следующим образом. Первая группа I_1 заканчивается номером $i(1) \equiv N$ и содержит все номера $i \leq i(1)$ такие, что $\Delta_i \leq (1-\lambda)^{-1} \Delta_N \equiv (1-\lambda)^{-1} \Delta_{i(1)}$. Если I_1 отличается от I , определим $i(2)$ как наибольший из номеров $i \in I$, не принадлежащих I_1 , и обозначим через I_2 множество всех номеров $i \leq i(2)$, для которых $\Delta_i \leq (1-\lambda)^{-1} \Delta_{i(2)}$. Если объединение I_1 и I_2 все еще не совпадает с I , определим $i(3)$ как наибольший из номеров в I , не принадлежащих $I_1 \cup I_2$, и положим $I_3 = \{i : \Delta_i \leq (1-\lambda)^{-1} \Delta_{i(3)}\}$.

Описанный процесс продолжается до полного исчерпания множества I . Поскольку Δ_i не возрастает по i , $I_s = \{i : i(s-1) < i \leq i(s)\}$, $s = 1, \dots, k$ ($i(0) = 0$).

Итак, множество I разбивается на последовательные сегменты $I_1, \dots, I_k$ (I_s следует за I_{s+1}). Наибольший индекс в I_s есть $i(s)$, причем по построению

$$\Delta_{i(s+1)} > (1-\lambda)^{-1} \Delta_{i(s)}, \quad s = 1, \dots, k-1. \quad (6)$$

2°. Проверим, что множества $Q_i = \{x \in G \mid f_i(x) \leq l_i\}$ уровня моделей, отвечающих одной группе итераций I_s , имеют общую точку – точку минимума $u_s = \underline{x}_{i(s)}$ последней, $i(s)$ -й, модели группы.

В самом деле, модели не убывают по i , а наилучшие найденные значения $\bar{f}_i$ не возрастают по i , так что при $i \in I_s$ имеем: $f_i(u_s) \leq f_{i(s)}(u_s) = \bar{f}_{i(s)} - \Delta_{i(s)} \leq \bar{f}_i - \Delta_{i(s)} \leq \bar{f}_i - (1-\lambda)\Delta_i \equiv l_i$ (заключительное неравенство следует из того, что $i \in I_s$, т.е. $\Delta_i \leq (1-\lambda)^{-1} \Delta_{i(s)}$).

3°. Теперь мы в состоянии оценить число N_s итераций в группе I_s . В самом деле, x_{i+1} – проекция x_i на выпуклый многогранник Q_i , содержащий точку u_s ($i \in I$), так что по стандартным свойствам проекции $|x_{i+1} - u_s|^2 \leq |x_i - u_s|^2 - |x_i - x_{i+1}|^2$, $i \in I_s$, откуда, обозначая через $j(s)$ первый номер в I_s , получаем

$$\sum_{i \in I_s} |x_i - x_{i+1}|^2 \leq |x_{j(s)} - u_s|^2 \leq D^2. \quad (7)$$

Оценим теперь снизу смещения $|x_i - x_{i+1}|$. В точке x_i i -я модель f_i равна $f(x_i)$ и поэтому не меньше, чем $\bar{f}_i$, тогда как в точке x_{i+1} та же модель, по построению, не превосходит $l_i = \bar{f}_i - (1-\lambda)\Delta_i$. Таким образом, при переходе от x_i к x_{i+1} i -я модель

изменяется по меньшей мере на $(1 - \lambda)\Delta_i$, что, в свою очередь, не меньше $(1 - \lambda)\Delta_{i(s)}$. Так как функция f_i очевидно, липшицева с константой L , то $|x_i - x_{i+1}|^2 \geq (1 - \lambda)\Delta_{i(s)}L^{-1}$. Отсюда и из (7) следует, что число N_s итераций группы I_s допускают оценку $N_s \leq (1 - \lambda)^{-2}L^2D^2\Delta_{i(s)}^{-2}$.

Так как $\Delta_{i(1)} \equiv \Delta_N > \varepsilon$ и $\Delta_{i(s)} > (1 - \lambda)^{-s+1}\Delta_{i(1)}$ при $s > 1$ (см. (6)), мы получаем из предыдущей оценки, что $N_s < (1 - \lambda)^{-2}L^2D^2(1 - \lambda)^{2(s-1)}\varepsilon^{-2}$, откуда $N = \sum_s N_s < c(\lambda)L^2D^2\varepsilon^{-2}$. Теорема 1 доказана.

Заметим, что Lev_λ на самом деле принадлежит семейству *bundle*-методов: в этом методе текущий центр регуляризации всегда совпадает с последней поисковой точкой, а штрафной коэффициент на каждой итерации выбирается так, чтобы точка x_{i+1} , определяемая соотношением (4), принадлежала множеству $\{x \in G \mid f_i(x) = l_i\}$. Подчеркнем, однако, что реализация метода уровней никоим образом не требует явного вычисления соответствующего d_i . Укажем еще, что при $\lambda \rightarrow 0$ метод уровней превращается в метод Келли.

Известно ([7]), что оценка (5) не допускает равномерного по размерности n задачи улучшения более чем в абсолютную константу раз. Таким образом, теорема 1 утверждает, что метод уровней является оптимальным (в некотором точном смысле) методом решения выпуклых задач большой размерности.

3. МЕТОД УРОВНЕЙ ДЛЯ ОТЫСКАНИЯ СЕДЛОВЫХ ТОЧЕК ВЫПУКЛО-ВОГНУТЫХ ФУНКЦИЙ

Идею метода уровня можно перенести на задачи поиска седловой точки функции f , определенной на прямом произведении G выпуклых многогранников $P \subset E^{n_x}$ и $Q \subset E^{n_y}$: найти

$$z^* = (x^*, y^*) \in G = P \times Q : f(x, y^*) \geq f(x^*, y^*) \geq f(x^*, y), (x, y) \in G, \quad (8)$$

где f — выпукло-вогнутая (по $x \in P$ и $y \in Q$ соответственно) функция, удовлетворяющая на G условию Липшица с константой L .

Сформулированная задача, как известно, эквивалентна двойственной паре выпуклых задач

$$\varphi(x) \equiv \max_{y \in Q} f(x, y) \rightarrow \min, \quad x \in P,$$

$$\psi(y) \equiv \min_{x \in P} f(x, y) \rightarrow \max, \quad y \in Q$$

с равными оптимальными значениями, или, что то же, единственной оптимизационной задаче

$$v(z) = \varphi(x) - \psi(y) \rightarrow \min, \quad z = (x, y) \in G \quad (9)$$

с нулевым оптимальным значением.

Погрешность точки $(x, y) \in G$ в качестве решения указанной задачи естественно характеризовать величиной $v(x, y) = \varphi(x) - \psi(y) = \left\{ \varphi(x) - \min_{x \in P} \varphi(x) \right\} +$

$+ \left\{ \max_{y \in Q} \psi(y) - \psi(y) \right\}$ — суммой погрешностей по функционалу точек x и y , рассматри-

ваемых как приближенные решения задачи минимизации φ и максимизации ψ соответственно.

Заметим, что если мы уже сформировали последовательность поисковых точек

$z_j = (x_j, y_j) \in G, j = 1, \dots, i$, и вычислили в этих точках значения $f(z_j)$ и частные суб/суперградиенты $f'_x(z_j), f'_y(z_j)$ функций $f(\cdot, y), f(x, \cdot)$ (предполагается, что норма указанных векторов не превосходят L), то мы в состоянии построить модели

$$\begin{aligned}\varphi_i(x) &= \max_{j \leq i} [f(z_j) + (x - x_j)^T f'_x(z_j)], \\ \psi_i(y) &= \min_{j \leq i} [f(z_j) + (y - y_j)^T f'_y(z_j)]\end{aligned}\quad (10)$$

функций φ, ψ . Ясно, что φ_i оценивает φ снизу, а ψ_i оценивает ψ сверху, так что модель

$$v_i(x, y) = \varphi_i(x) - \psi_i(y), \quad (11)$$

как и в ситуации разд. 2, оценивает снизу целевой функционал подлежащей решению выпуклой задачи (9). Единственное отличие нынешней ситуации от описанной в разд. 2 состоит в том, что модель (11), вообще говоря, не точна и в поисковых точках. Это осложнение, однако, не мешает реализации метода уровней, i -я итерация которого теперь такова.

Имея информацию $f(z_j), f'_x(z_j), f'_y(z_j)$ в поисковых точках $z_j \in G, j < i$, и очередную поисковую точку z_i , вычисляем $f(z_i), f'_x(z_i), f'_y(z_i)$ и строим модели φ_i, ψ_i, v_i .

Далее вычисляем минимальное значение $-\Delta_i$ модели v_i на G , решая пару задач ЛП: $\varphi_i(x) \rightarrow \min, x \in P, \psi_i(y) \rightarrow \max, y \in Q$. Решив указанные задачи, находим соответствующие двойственные множители $\lambda_j, \mu_j, i = 1, \dots, i$, которые удовлетворяют условиям

$$\lambda_j \geq 0, \sum_{j=1}^i \lambda_j = 1, \min_{x \in P} \varphi_i(x) = \min_{x \in P} \sum_{j=1}^i \lambda_j [f(z_j) + (x - x_j)^T f'_x(z_j)], \quad (12)$$

$$\mu_j \geq 0, \sum_{j=1}^i \mu_j = 1, \max_{y \in Q} \psi_i(y) = \max_{y \in Q} \sum_{j=1}^i \mu_j [f(z_j) + (y - y_j)^T f'_y(z_j)]. \quad (13)$$

С помощью этих множителей формируем i -е приближенное решение задачи — точку

$$z_i^* = \left(x_i^* = \sum_{j=1}^i \mu_j x_j, y_i^* = \sum_{j=1}^i \lambda_j y_j \right).$$

Новая поисковая точка определяется из соотношения

$$z_{i+1} = (x_{i+1}, y_{i+1}) = \arg \min \{ |z_i - z| \mid z \in G, v_i(z) \leq -(1 - \lambda) \Delta_i \},$$

где $\lambda \in (0, 1)$ — фиксированный параметр метода. Итерация закончена. Обозначим описанный вариант метода Lev_λ через Lev'_λ .

Теорема 2. Для метода Lev'_λ при любом i выполняется неравенство

$$v(z_i^*) \leq \Delta_i. \quad (14)$$

Кроме того, этот метод при любом $\varepsilon > 0$ находит ε -решение задачи (8) (т.е. обеспечивает неравенство $v(z_i^*) \leq \varepsilon_i$) за не более чем

$$N'(\varepsilon) = \left\lceil c'(\lambda) \frac{L^2 D^2}{\varepsilon^2} \right\rceil, \quad c'(\lambda) = 2((1 - \lambda)^2 \lambda (2 - \lambda))^{-1}, \quad (15)$$

шагов, где D — диаметр G и L — константа Липшица функции f .

Доказательство. 1°. Проверим (14). Фиксируем $(x, y) \in G$. Имеем

$$f(x_j, y) \leq f(z_j) + (y - y_j)^T f'_y(z_j),$$

$$f(x, y_j) \geq f(z_j) + (x - x_j)^T f'_x(z_j).$$

Беря суммы указанных неравенств с весами λ_j, μ_j соответственно и вычитая результирующие друг из друга, получаем

$$\begin{aligned} \sum_{j=1}^i \{ \lambda_j f(x, y_j) - \mu_j f(x_j, y) \} &\geq \sum_{j=1}^i \lambda_j \{ f(z_j) + (x - x_j)^T f'_x(z_j) \} - \\ &- \sum_{j=1}^i \mu_j \{ f(z_j) + (y - y_j)^T f'_y(z_j) \}. \end{aligned} \quad (16)$$

Левая часть в (16) ввиду выпукло-вогнутости f не больше $f(x, y_i^*) - f(x_i^*, y)$. Следовательно, $f(x, y_i^*) - f(x_i^*, y) \geq \sum_{j=1}^i \lambda_j \{ f(z_j) + (x - x_j)^T f'_x(z_j) \} - \sum_{j=1}^i \mu_j \{ f(z_j) + (y - y_j)^T f'_y(z_j) \}.$

Минимизируя обе части полученного неравенства по $(x, y) \in G$ и принимая во внимание (12), (13), мы приходим к следующему: $-v(z_i^*) \geq \min_{x \in P} \varphi_i(x) - \max_{y \in Q} \psi_i(y)$, как и утверждается в (14).

2°. Остается проверить, что Δ_i заведомо $\leq \varepsilon$ для всех $i \geq N(\varepsilon)$ (см. (15)). В самом деле, пусть целое N таково, что $\Delta_N > \varepsilon$. Докажем, что в этом случае N меньше $N'(\varepsilon)$.

В силу конструкции наших моделей ясно, что v_i поточечно не убывают с ростом i , так что величины Δ_i не возрастают по i . С учетом этого мы, как и в доказательстве теоремы 1, можем разбить множество $I = \{1, \dots, N\}$ на группы $I_1, \dots, I_k$ так, чтобы группа I_{s+1} предшествовала I_s , и выполнялись неравенства

$$\Delta_i \leq (1 - \lambda)^{-1} \Delta_{i(s)}, \quad i \in I_s, \quad \Delta_{i(s+1)} > (1 - \lambda)^{-1} \Delta_{i(s)}, \quad (17)$$

где $i(s)$ — последний номер в группе I_s . Мы утверждаем, что множества $G_i = \{z \in V_i(z) \leq l_i \equiv -(1 - \lambda)\Delta_i\}$, отвечающие итерациям $i \in I_s$, имеют общую точку, именно, минимизатор w_s последней, i_s -й, модели группы. В самом деле, модели поточечно не убывают по i , откуда

$$v_i(w_s) \leq v_{i(s)}(w_s) = -\Delta_{i(s)} \leq -(1 - \lambda)\Delta_i, \quad i \in I_s$$

(мы учли (17)), так что $w_s \in G_i$ при $i \in I_s$.

Теперь доказательство завершается точно так же, как и в случае теоремы 1. Именно, заметим, что $v_i(z_i) \geq 0$ (ибо по очевидным причинам $\varphi_i(x_i) \geq f(z_i) \geq \psi_i(y_i)$, так что $z_i \notin G_i$; поэтому $v_i(z_{i+1}) = -(1 - \lambda)\Delta_i$ и

$$v_i(z_i) - v_i(z_{i+1}) \geq (1 - \lambda)\Delta_i \geq (1 - \lambda)\Delta_{i(s)}. \quad (18)$$

Так как, очевидно, функция v_i липшицева с константой L по x и по y , она липшицева с константой $L\sqrt{2}$ по z и (18) дает

$$|z_i - z_{i+1}| \geq (1 - \lambda)\Delta_{i(s)} L^{-1} / \sqrt{2}. \quad (19)$$

В то же время, поскольку $w_s \in G_i, i \in I_s$, с учетом известного свойства операции проектирования имеем

$$|z_{i+1} - w_s|^2 \leq |z_i - w_s|^2 - |z_i - z_{i+1}|^2 \leq |z_i - w_s|^2 - \frac{1}{2L^2} (1 - \lambda)^2 \Delta_{i(s)}^2.$$

Следовательно, число N_s номеров в I_s удовлетворяет неравенству: $N_s \leq \leq 2D^2L^2(1-\lambda)^{-2}\Delta_{i(s)}^{-2}$.

В силу (17) и предположения о том, что $\Delta_{i(s)} \equiv \Delta_N > \varepsilon$

$$\Delta_{i(s)} \geq (1-\lambda)^{-(s-1)} \Delta_{i(s)} > (1-\lambda)^{-(s-1)} \varepsilon.$$

Итак, $N_s < 2D^2L^2(1-\lambda)^{-2}(1-\lambda)^{2(s-1)}\varepsilon^{-2}$, откуда

$$N = \sum_s N_s < \frac{2}{(1-\lambda)_\lambda^2(2-\lambda)} D^2L^2\varepsilon^{-2}.$$

Теорема 2 доказана.

4. МЕТОД УРОВНЕЙ ДЛЯ ВЫПУКЛЫХ ЗАДАЧ С ОГРАНИЧЕНИЯМИ

Этот раздел посвящен распространению метода уровней на общую задачу выпуклого программирования

$$f(x) \rightarrow \min, \quad g^i(x) \leq 0, \quad i = 1, \dots, m, \quad x \in G, \quad (20)$$

где $G \subset E^n$ — выпуклый компакт; f и $g^i, i = 1, \dots, m$, выпуклые функции на G , удовлетворяющие условию Липшица. Предполагается, что задача (20) разрешима.

Для дальнейшего удобно перейти от задачи с несколькими функциональными ограничениями к аналогичной с одним ограничением. Для этой цели достаточно задаться произвольными масштабными коэффициентами $\lambda_i > 0, i = 1, \dots, m$, и переписать (20) в эквивалентном виде

$$f(x) \rightarrow \min, \quad g(x) \equiv \max_{1 \leq i \leq m} \lambda_i g^i(x) \leq 0, \quad x \in G. \quad (21)$$

Обозначим через $L(f)$ и $L(g)$ константы Липшица функций f, g соответственно.

Допустим, что вычислены значения функций f и g вместе с их субградиентами в точках $x_1, \dots, x_i \in G$. Полученная информация дает возможность построить i -е модели $f_i(x) = \max_{1 \leq j \leq i} [f(x_j) + (x - x_j)^T f'(x_j)]$ и $g_i(x) = \max_{1 \leq j \leq i} [g(x_j) + (x - x_j)^T g'(x_j)]$ целевого функционала f и функции ограничения g . Эти модели оценивают соответствующие функции снизу и не убывают с ростом i , т.е.

$$f_1(x) \leq \dots \leq f(x), \quad g_1(x) \leq \dots \leq g(x). \quad (22)$$

По моделям f_i и g_i строится i -я аппроксимация задачи (21)

$$f_i(x) \rightarrow \min, \quad g_i(x) \leq 0, \quad x \in G. \quad (23)$$

В силу (22) допустимое множество задачи (23) шире, чем у задачи (21), а ее целевой функционал не больше, чем у (21). Следовательно, оптимальное значение $\underline{f}_i = \min\{f_i(x) | x \in G, g_i(x) \leq 0\}$ задачи (23) не больше оптимального значения f^* исходной задачи (21). С ростом i допустимое множество задачи (23) может только сужаться, а ее целевой функционал — лишь расти, поэтому

$$\underline{f}_1 \leq \dots \leq f^*. \quad (24)$$

Введем i -е множество достижимости Γ_i , положив

$$\Gamma_i = \text{con } v\{(u, v) \in E^2 | \exists j \leq i: (f(x_j), g(x_j)) \leq (u, v)\}. \quad (25)$$

Определим "расстояние" d_i от точки $z_i = (\underline{f}_i, 0)$ до множества Γ_i соотношением

$$d_i = \min_{(u, v) \in \Gamma} \max(u - \underline{f}_i, v).$$

Из (25) вытекает, что d_i – оптимальное значение задачи ЛП

$$\omega \rightarrow \min, \sum_{j=1}^i \mu_j f(x_j) - \omega \leq \underline{f}_i, \quad (26)$$

$$\sum_{j=1}^i \mu_j g(x_j) - \omega \leq 0, \quad \sum_{j=1}^i \mu_j = 1, \quad \mu_j \geq 0, \quad j = 1, \dots, i,$$

с переменными $\mu_1, \dots, \omega$. Пусть $(\mu^i, \omega^i) = (\mu_1^i, \dots, \mu_i^i, d_i)$ – решение задачи (26). Положим $\bar{x}_i = \sum_{j=1}^i \mu_j^i x_j$. Из (24), (26) и выпуклости функций f, g следует, что $\bar{x}_i$ является d_i -решением задачи (21), т.е.

$$f(x_i) \leq f^* + d_i, \quad g(\bar{x}_i) \leq d_i, \quad \bar{x}_i \in G. \quad (27)$$

Из (27), в частности, вытекает, что $d_i \geq 0$ для любого i . Двойственная к (26) задача ЛП может быть записана в виде

$$h_i(\alpha) \equiv \min_{1 \leq j \leq i} [\alpha(f(x_j) - \underline{f}_i) + (1 - \alpha)g(x_j)] \rightarrow \max, \quad 0 \leq \alpha \leq 1. \quad (28)$$

Ясно, что $h_i(\alpha)$, введенная в (28), – вогнутая кусочно-линейная функция $\alpha \in [0, 1]$, удовлетворяющая условию Липшица с константой $L = (L(f) + L(g)) D$, где D – диаметр множества G .

Из (28) можно заключить, что точки x_i следует выбирать так, чтобы гарантировать стремление к нулю величин $d_i = \max_{\alpha \in [0, 1]} h_i(\alpha)$ при $i \rightarrow \infty$. Заметим, что в силу (28) функции h_i не возрастают по i , т.е.

$$h_1(\alpha) \geq \dots, \quad \alpha \in [0, 1]. \quad (29)$$

Кроме того, можно считать, что $d_i = \max_{\alpha \in [0, 1]} h_i(\alpha) > 0$, ибо в ситуации $d_i = 0$ точка $\bar{x}_i$, согласно (27), является решением исходной задачи.

Предлагаемый ниже алгоритм складывается из серии шагов метода Lev_λ из разд. 2, применяемого для минимизации функции $\varphi_\alpha(x) = \alpha f(x) + (1 - \alpha)g(x)$ на G при различных значениях α .

При обосновании алгоритма будет использовано следующее утверждение.

Лемма 1. Пусть $\varepsilon > 0$ и $y^1, y^2, \dots$ – траектория метода уровней Lev_λ , примененного к задаче $\varphi_\alpha(x) \rightarrow \min, x \in G$. Допустим, что при некоторых i и j точки $y^1, \dots, y^j$ содержатся среди точек $x_1, \dots, x_i$ и при этом $j \geq c(\lambda)(L(f) + L(g)^2)D^2\varepsilon^{-2}$. Тогда

$$h_i(\alpha) \leq \varepsilon. \quad (30)$$

Доказательство. В самом деле, по теореме 1 имеем

$$\min_{k \leq j} \varphi_\alpha(y^k) - \min_{y \in G} \max_{k \leq j} \{\varphi_\alpha(y^k) + (y - y^k)\varphi'_\alpha(y^k)\} \leq \varepsilon. \quad (31)$$

Далее, учитывая определение функций h, φ_α и формулировку леммы, получаем

$$\begin{aligned} \min_{k \leq j} \varphi_\alpha(y^k) &\geq \min_{k \leq i} \varphi_\alpha(x_k) = \min_{k \leq i} \{\alpha f(x_k) + (1 - \alpha)g(x_k)\} = \\ &= \min_{k \leq i} \{\alpha(f(x_k) - \underline{f}_i) + (1 - \alpha)g(x_k)\} + \alpha \underline{f}_i = h_i(\alpha) + \alpha \underline{f}_i \end{aligned} \quad (32)$$

и кроме того

$$\begin{aligned} \min_{y \in G} \max_{k \leq j} \{\varphi_\alpha(y^k) + (y - y^k)\varphi'_\alpha(y^k)\} &\leq \min_{y \in G} \max_{k \leq i} \{\varphi_\alpha(x_k) + \\ &+ (y - x_k)^T \varphi'_\alpha(x_k)\} \leq \min_{y \in G} \{\alpha f_i(y) + (1 - \alpha)g_i(y)\} \leq \alpha \underline{f}_i \end{aligned} \quad (33)$$

(заключительное неравенство немедленно следует из определения $\underline{f}_i$). Соединяя (31)–(33), получаем (30). Лемма доказана.

Таким образом, имеется возможность стремиться к нулю с некоторой фиксированной скоростью значение функции $h_i(\alpha)$ в любой заданной точке α . Чтобы добиться стремления к нулю интересующей нас величины $d_i = \max_{\alpha \in [0,1]} h_i(\alpha)$, поступим следующим образом.

Пусть $\delta_i = \{\alpha \in [0, 1] | h_i(\alpha) \geq 0\}$. Так как h_i вогнута, δ_i – некоторый отрезок; а поскольку h_i не возрастает с ростом i ((29))

$$\delta_1 \supset \delta_2 \supset \dots \quad (34)$$

Зададимся некоторым параметром $\chi \in (0, 1/2)$ и условимся говорить, что точка $c \in [a, b]$ χ -центрирована по отношению к отрезку $[a, b]$ вещественной прямой, если $\min\{|c - a|, |c - b|\} \geq \chi |b - a|$.

Пусть $h(\alpha)$ – вогнутая на $[0, 1]$ и неотрицательная на отрезке $\delta \supset [0, 1]$ функция, точка $\alpha \in \delta$. Если эта точка χ -центрирована по отношению к δ , то легко проверить, что

$$\max_{\alpha \in [0, 1]} h(\alpha) \leq \chi^{-1} h(\alpha). \quad (35)$$

В частности, если $h_i(\alpha)$ – последовательность вогнутых функций на $[0, 1]$, отрезки $\delta_i = \{\alpha \in [0, 1] | h_i(\alpha) \geq 0\}$ непусты, $\alpha_i \in \delta_i$ и χ -центрированы по отношению к δ_i , то, согласно (35), из $h(\alpha_i) \rightarrow 0$ при $i \rightarrow \infty$ следует, что $\max_{\alpha \in [0,1]} h(\alpha) \rightarrow 0$ при $i \rightarrow \infty$.

Теперь мы подготовлены к описанию метода уровней $\text{Lev}_{\lambda, \chi}$ для решения задачи (21). Метод определяется параметрами $\lambda \in (0, 1)$, $\chi \in (0, 1/2)$ и работает следующим образом. Полагаем $\alpha^1 = 1/2$ и на первом этапе начинаем применять метод Lev_{λ} к задаче минимизации на G функции $\varphi_{\alpha}(x) = \alpha f(x) + (1 - \alpha)g(x)$ при $\alpha = \alpha^1$. В ходе работы метода строим описанные выше объекты $\underline{f}_i$, Γ_i , $h_i(\alpha)$, δ_i и проверяем на каждом шаге, является ли α^1 χ -центрированной по отношению к отрезку $\delta_i = \{\alpha | h_i(\alpha) \geq 0\}$. На первом шаге (шаг $i(1)$), на котором это условие нарушается, определяем α^2 как середину отрезка $\delta_{i(1)}$ и начинаем минимизировать методом Lev_{λ} функцию $\varphi_{\alpha}(x)$, отвечающую $\alpha = \alpha^2$. Второй этап продолжается до тех пор, пока точка α^2 остается χ -центрированной относительно δ_i . Когда последнее условие нарушается (пусть это случилось на шаге $i(2)$), приступаем к третьему этапу, на котором определяем α^3 как середину отрезка $\delta_{i(2)}$ и применяем Lev_{λ} к функции $\varphi_{\alpha}(x)$, при $\alpha = \alpha^3$, и т.д. В ходе работы метода на каждом шаге i формируется приближенно решение $\bar{x}_i$ задачи (21) (см. (27)).

Свойства сходимости описанного метода дает следующее утверждение.

Теорема 3. Каково бы ни было $\varepsilon > 0$, для описанного метода $\text{Lev}_{\lambda, \chi}$ $0 < \lambda < 1$, $0 < \chi < 1/2$, условие $d_N \leq \varepsilon$ заведомо выполнится (и, следовательно, x_N окажется ε -решением (21)) при

$$N \geq c(\varepsilon, \chi) c(\lambda) (L(f) + L(g)D)^2 (\chi\varepsilon)^{-2}, \quad (36)$$

где

$$c(\varepsilon, \chi) = 1 + \left[\ln \frac{2}{2\chi + 1} \right]^{-1} \ln[1 + (L(f) + L(g)D)\varepsilon^{-1}],$$

$$c(\lambda) = ((1 - \lambda)^2 \lambda (2 - \lambda))^{-1}.$$

Доказательство. Пусть, напротив, N удовлетворяет (36) и $d_N > \varepsilon$; $N_1, \dots, N_k$ — число шагов метода до момента N включительно на последовательных этапах (k — число этапов, начавшихся не позднее N). На всех шагах i этапа $j \leq k$ точка a^i χ -центрирована по отношению к отрезку δ_j , поэтому на этих шагах

$$h_i(a^i) > \chi \varepsilon. \quad (37)$$

В самом деле, в противном случае ввиду (35) было бы $d_N \leq d_i = \max_{\alpha} h_i(\alpha) \leq \chi^{-1} h_i(a^i) \leq \varepsilon$, что противоречит условию $d_N > \varepsilon$ (напомним, что $d_i \geq d_N$ при $i \leq N$ согласно (29)).

Ввиду (37) и леммы 1 справедлива оценка

$$N_j < c(\nu) \chi^{-2} (L(f) + L(g))^2 D^2 \varepsilon^{-2}. \quad (38)$$

Теперь оценим число этапов k . На первом шаге $i = i(j)$ этапа j точка a^i совпадает с серединой отрезка δ_j , а на первом шаге $i' = i(j+1)$ следующего этапа она не χ -центрирована по отношению к отрезку $\delta_{j'} \subset \delta_j$ (включение вытекает из (34)). Очевидно, такое возможно лишь при

$$|\delta_{j'}| \leq \left(\chi + \frac{1}{2} \right) |\delta_j|.$$

Следовательно

$$|\delta_N| \leq \left(\chi + \frac{1}{2} \right)^{k-1}. \quad (39)$$

Проверим, что если $k > 1$, то

$$|\delta_N| \geq \min \left\{ \frac{\varepsilon}{(L(f) + L(g))D}, 1 \right\}. \quad (40)$$

В самом деле, если $|\delta_N| < 1$, то, в силу происхождения δ_N , функция $h_N(\alpha)$ неположительна в одном из концов δ_N . Поскольку она к тому же липшицева с константой $(L(f) + L(g))D$, ее максимум на δ_N (т.е. в точности величина $d_N > \varepsilon$) не превосходит $(L(f) + L(g))D |\delta_N|$, поэтому $(L(f) + L(g))D |\delta_N| > \varepsilon$, как и утверждалось. Соединяя (39), (40), получаем

$$k \leq 1 + \left[\ln \left(\frac{2}{2\chi + 1} \right) \right]^{-1} \ln [1 + (L(f) + L(g))D \varepsilon^{-1}].$$

Вспоминая, наконец, (38), приходим к противоречию с (36). Итак, при N , удовлетворяющем (36), d_N на самом деле не превосходит ε . Теорема 3 доказана.

Из теоремы 3 следует, что теоретическая оценка трудоемкости отыскания ε -решения условно-экстремальной выпуклой задачи (21) описанным методом лишь логарифмическим (относительно $1/\varepsilon$) множителем хуже аналогичной оценки для задач без функциональных ограничений (теорема 1).

5. ПРИЛОЖЕНИЯ К ДЕКОМПОЗИЦИИ

Один из источников негладких выпуклых задач минимизации — схемы декомпозиции оптимизационных задач с ограничениями. Этот раздел посвящен двойственному декомпозиционному подходу к решению задач ЛП, ограничения которых разбиты на два горизонтальных блока. Подход основан на методе уровней Lev_λ , изложенных в разд. 2.

Рассмотрим задачу ЛП

$$P: (c, x) \rightarrow \max, Ax = b, x \in G, \quad (41)$$

где $c, x \in E^n$, $b \in E^m$, матрица A имеет m строк и n столбцов; $G \subset E^n$ — непустой выпуклый многогранник. Предполагается, что система ограничений задачи (41) совместна, т.е. (41) разрешима. В основе декомпозиции лежит предположение о том, что система линейных уравнений и неравенств, задающая G , имеет определенную специфику, облегчающую решение задачи ЛП вида

$$(c', x) \rightarrow \max, x \in G \quad (42)$$

при любом $c' \in E^n$.

Общая схема двойственной декомпозиции состоит в следующем.

Введем функцию $f: E^m \rightarrow E^1$, положив

$$f(y) = \max_{x \in G} (c - y^T A, x) + (y, b). \quad (43)$$

Очевидно, f — выпуклая, кусочно-линейная функция, имеющая конечные значения во всех точках E^m . Решение задачи (41) сводится к минимизации f в том смысле, что если $\bar{y} \in \text{Arg} \min_{y \in E^m} f(y)$, то среди решений задачи (42) при $c' = c - \bar{y}^T A$ содержится оптимальный план исходной — (41). Из определения f вытекает, что при любом $y \in E^m$ ее субдифференциал $\partial f(y) = \{l: l = b - Ax, x \in G_y^*\}$, где G_y^* — множество решений задачи (42) при $c' = c - y^T A$. Поэтому для вычисления $f(y)$ и некоторого субградиента $f'(y) \in \partial f(y)$ достаточно найти одно из решений (42) при $c' = c - y^T A$: если $x_y \in G_y^*$, то

$$f(y) = (c, x_y) + (y, b - Ax_y), f'(y) = b - Ax_y. \quad (44)$$

Из (44) следует, что любой сходящийся алгоритм минимизации первого порядка приводит к итеративной процедуре решения задачи (1), на каждом шаге которой необходимо решать (42) при некотором c' . Такого рода процедуры принято называть декомпозиционными (блочными) методами, причем, учитывая их связь с двойственной функцией f , подчеркивать в названии двойственный характер этих методов. Заметим, что при реализации двойственного блочного метода необходимо иметь удовлетворительный способ решения так называемой проблемы возврата к исходной задаче, т.е. уметь экономно отыскивать приближенное решение задачи (41) по имеющейся приближенной точке минимума функции f . Ниже описывается двойственный блочный метод, связанный с минимизацией f методом уровней. Выбор Lev_λ для реализации двойственной декомпозиции мотивирован его эффективностью (в смысле числа вычисленных $f(y)$ и $f'(y)$, т.е. решенных задач типа (2)) и возможностью решить проблему возврата практически без дополнительных вычислений. Чтобы использовать метод уровней для двойственной декомпозиции задачи (41), перейдем от (41) к семейству задач

$$P(R): (c, x) - R \|Ax - b\|_1 \rightarrow \max, x \in G, \quad (45)$$

где $R > 0$ — штрафной параметр, $\|y\|_1 = \sum_{i=1}^m |y_i|$.

Легко видеть, что задача, двойственная к $P(R)$, имеет вид

$$\tilde{P}(R): f(y) \rightarrow \min, y \in Q(R), \quad (46)$$

где $Q(R) = \left\{ y \in E^m: \max_{1 \leq i \leq m} |y_i| \leq R \right\}$.

Обозначим через $V(R)$ общее экстремальное значение задач (45) и (46). При $R \rightarrow \infty$

задачи (45) и (46) переходят соответственно в (41) и двойственную к (41) задачу

$$f(y) \rightarrow \min, y \in E^m. \quad (47)$$

Общее экстремальное значение задач (41) и (47) обозначим $V = V(\infty)$.

Введение семейств двойственных задач (45) и (46) связано с тем, что метод уровней приспособлен к минимизации функций на компакте. Пусть $y^j \in Q(R)$, $x^j \in G_{y^j}^* \subset G$, $1 \leq j \leq i$. Положим $l^j = b - Ax^j$, тогда, согласно (44), $f(y^j) + (l^j, y - y^j) = (c, x^j) + (b - Ax^j, y)$. Поэтому i -я модель f_i функции f определяется соотношением

$$f_i(y) = \max_{1 \leq j \leq i} [(c, x^j) + (b - Ax^j, y)]. \quad (48)$$

Зафиксировав $R > 0$ и $i = 1, \dots$, рассмотрим двойственную пару задач ЛП

$$P(R, i): (c, x) - R \|Ax - b\|_1 \rightarrow \max, x \in G(i), \quad (49)$$

где $G(i)$ – выпуклая оболочка точек $x^j, j \leq i$, и

$$\tilde{P}(R, i): f_i(y) \rightarrow \min, y \in Q(R). \quad (50)$$

Их общее экстремальное значение обозначим через $V(R, i)$.

Зафиксируем произвольное положительное R . Опишем отдельную (i -ю) итерацию декомпозиционного метода решения задачи $P(R)$, основанного на применении к задаче $\tilde{P}(R)$ метода Левенберга при $\lambda = 1/2$.

К началу итерации i имеем точки $y^j \in Q(R)$, $1 \leq j \leq i$, $x^j \in G$, $1 \leq j \leq i-1$ (в качестве y^1 выбирается любая точка из $Q(R)$). Итерация i начинается с вычисления точки x^i , в качестве которой берется одно из решений задачи (42) при $c' = c - (y^i)^T A$. Затем, используя (48), составляем i -модель f_i функции f , и решаем задачу ЛП $\tilde{P}(R, i)$ (см. (50)). В результате определяется $\underline{y}^i$ – оптимальный план задачи (50) и вектор двойственных множителей $\gamma = (\gamma_1, \dots, \gamma_i)$, $\sum_{j=1}^i \gamma_j = 1$, $\gamma_j \geq 0$, $j \leq i$. С помощью γ вычисляется точка

$$\underline{x}^i = \sum_{j=1}^i \gamma_j x^j \quad (51)$$

– i -е приближенное решение задачи $P(R)$. Одновременно подсчитывается

$$\Delta_i = \min_{1 \leq j \leq i} f(y^j) - V(R, i) \geq 0. \quad (52)$$

Если $\Delta_i \leq \varepsilon$, где ε – заранее фиксированная точность, то итерация i , а вместе с ней и весь процесс заканчивается, выдавая в качестве приближенного решения задачи $P(R)$ точку $\underline{x}^i$. В противном случае решается задача квадратичного программирования

$$\|y - y^i\|^2 \rightarrow \min, f_i(y) \leq V(R, i) + \frac{\Delta_i}{2}, \quad (53)$$

и ее решение y^{i+1} пополняет набор точек y^j . Итерация i на этом заканчивается.

Положим $\rho(x) = \max_{1 \leq s \leq m} |(Ax - b)_s|$, где $(Ax - b)_s$ – s -я координата вектора $Ax - b$, $c_1 = \max_{x \in G} (c, x)$, $c_2 = \min_{x \in G} (c, x)$.

Сходимость двойственного декомпозиционного метода устанавливает следующее утверждение.

Теорема 4. 1°. Точка $\underline{x}^i$ является Δ_i -решением задачи $P(R)$, т.е.

$$(c, \underline{x}^i) - R \|A\underline{x}^i - b\|_1 \geq V(R) - \Delta_i. \quad (54)$$

2°. Пусть $\varepsilon > 0$, $R \geq 1 + \frac{c_1 - c_2}{\varepsilon}$. Если после i итераций блочного метода оказывается выполненным неравенство $\Delta_i < \varepsilon$, то $\underline{x}^i$ — ε -решение задачи (1), т.е.

$$(c, \underline{x}^i) \geq V - \varepsilon, \quad \rho(\underline{x}^i) \leq \varepsilon. \quad (55)$$

Доказательство. 1°. Учитывая (48), свойство вектора двойственных множителей γ и (51), имеем

$$\begin{aligned} V(R, i) &= \min_{y \in Q(R)} f_i(y) = \min_{y \in Q(R)} \max_{1 \leq j \leq i} [(c, \underline{x}^j) + (b - A\underline{x}^j, y)] = \\ &= \min_{y \in Q(R)} \sum_{j=1}^i \gamma_j [(c, \underline{x}^j) + (b - A\underline{x}^j, y)] = \min_{y \in Q(R)} [(c, \underline{x}^i) + \\ &+ (b - A\underline{x}^i, y)] = (c, \underline{x}^i) - R \|b - A\underline{x}^i\|_1, \end{aligned}$$

т.е. точка $\underline{x}^i$ является решением задачи $P(R, i)$.

Неравенство (54) вытекает теперь из (52) и того, что

$$V(R, i) \leq V(R) \leq \min_{1 \leq j \leq i} f(y^j) \quad (56)$$

(левая часть (56) — следствие неравенства $f_i(y) \leq f(y)$, правая имеет место в силу определения $V(R)$ как экстремального значения задачи $\tilde{P}(R)$).

2°. Из определения константы c_1 и неравенства $\rho(x) \leq \|Ax - b\|_1$ при любом $R > 0$

$$c_1 - R\rho(\underline{x}^i) \geq (c, \underline{x}^i) - R \|A\underline{x}^i - b\|_1. \quad (57)$$

Поскольку задача (41) предполагается разрешимой, существует точка $x^\circ \in G$, для которой $Ax^\circ = b$, откуда, учитывая определение константы c_2 , получаем $(c, \underline{x}^i) - R \|A\underline{x}^i - b\|_1 \geq V(R) - \Delta_i \geq (c, x^\circ) - \Delta_i \geq c_2 - \Delta_i$.

Последнее неравенство вместе с (57) и требованиями к R и Δ_i в формулировке теоремы 4 приводит к правому соотношению в (55).

Далее из (54) и требования к Δ_i в формулировке теоремы 4 вытекает, что $(c, \underline{x}^i) \geq (c, \underline{x}^i) - R \|A\underline{x}^i - b\|_1 \geq V(R) - \varepsilon \geq (c, x^*) - \varepsilon$, где x^* — решение задачи (41). Таким образом, $\underline{x}^i$ — ε -решение задачи (41). Теорема 4 доказана.

Из первой части теоремы 4 следует, что при фиксированном штрафном параметре R число итераций декомпозиционного метода, требуемое для получения ε -решения задачи $P(R)$, не превышает числа $N(\varepsilon)$ из теоремы 1, т.е. числа порядка $O(\varepsilon^{-2})$. Прокомментируем вторую часть утверждения теоремы 4. Очевидно, функция f , определяемая в (43), удовлетворяет условию Липшица на E^n с некоторой константой L_f . Поэтому число итераций N декомпозиционного метода, необходимое для ε -решения задачи P , не превышает, согласно теореме 1

$$\bar{N}(\varepsilon) = \left\lceil c \left(\frac{1}{2} \right) L_f^2 n (1 + (c_1 - c_2) \varepsilon^{-1})^2 \varepsilon^{-2} \right\rceil, \quad (58)$$

т.е. числа порядка $O(\varepsilon^{-4})$. В случае разрешимости задачи (41) эта оценка для N допускает усиление. Действительно, нетрудно показать (см. теорему 1 из [8]), что параметр R из второй части теоремы 4 можно выбрать независимо от точности ε , положив его равным $1 + \max_{1 \leq i \leq m} y_i^*$, где y^* — вектор двойственных оценок ограничений $Ax = b$ в задаче P , поэтому N оценивается сверху числом порядка $O(\varepsilon^{-2})$. Заметим,

однако, что при выборе R в соответствии с формулировкой теоремы 4 мы можем применять декомпозиционный метод, не предполагая заранее разрешимость задачи (41). Для этого достаточно, зафиксировав $\varepsilon > 0$, провести не более, чем $\bar{N}(\varepsilon)$ итераций декомпозиционного метода (см. 58)) и получить $\Delta_i \leq \varepsilon$. Возможны два случая: $\rho_i(\underline{x}^i) > \varepsilon$ и $\rho_i(\underline{x}^i) \leq \varepsilon$. В первом задача (41) неразрешима (согласно теореме 4). Во втором вектор является ε -решением задачи (41), если она разрешима (теорема 4). Если же информация о разрешимости (41) отсутствует, можно утверждать, что $\underline{x}^i$ - ε -решение скорректированной задачи (41), в которой вектор b заменен на $b' = b + \Delta b$, где $\Delta b = A \underline{x}^i - b$, причем каждая координата вектора Δb , очевидно, не превышает ε . В самом деле, $\underline{x}^i$ — план скорректированной задачи (41), а для любого другого ее плана x имеем

$$(c, \underline{x}^i) - R \|A \underline{x}^i - b\| \geq V(R) - \varepsilon \geq (c, x) - R \|Ax - b\| - \varepsilon = (c, x) - R \|\Delta b\| - \varepsilon,$$

т.е. $(c, \underline{x}^i) \geq (c, x) - \varepsilon$.

6. ВЫЧИСЛИТЕЛЬНЫЙ ОПЫТ

Здесь речь пойдет об алгоритмической реализации базового метода уровней Lev_λ , описанного в разд. 2, а также о результатах его численного тестирования.

На каждой итерации Lev_λ решается одна задача линейного и одна задача квадратичного программирования. В качестве решателя в обоих случаях была использована программа [9], основанная на двойственном методе квадратичного программирования, изложенном в [10] (линейная задача сводилась к квадратичной путем добавления к линейному целевому функционалу квадрата расстояния до точки x_i , помноженного на достаточно малый коэффициент). Число ограничений линейной и квадратичной задач на i -й итерации определяется числом k_i аффинных функций, порождающих текущую модель f_i . В описанном варианте метода Lev_λ число $k_i = i$: набор аффинных функций (называемый далее пучком), определяющий модель f_i , увеличивается на единицу с каждой новой итерацией. Размер пучка существенно влияет на трудоемкость отдельной итерации метода, поэтому в процессе алгоритмизации Lev_λ усилия были направлены на разработку приемов "очистки" пучка — сокращения числа составляющих его аффинных функций. При этом очистка должна быть, с одной стороны, по возможности более глубокой, а, с другой — не вести к слишком значительному росту количества итераций метода. После численного опробования ряда приемов очистки пучка мы остановились на следующем простом правиле.

Пусть $i(s)$ — номер итерации, на которой происходит s -я по счету очистка пучка ($i(0) = 1$). Номер $i(s)$ при любом $s \geq 1$ определяется рекуррентным соотношением

$\Delta_{i(s)} \leq \frac{1}{2} \Delta_{i(s-1)}$, $\Delta_j > \frac{1}{2} \Delta_{i(s-1)}$ для $j < i(s)$, где $\Delta_i = \bar{f}_i - \underline{f}_i$ — оценка ошибки на i -й итерации (см. разд. 2). После очистки на $i(s)$ -й итерации в пучке остаются аффинные функции $l_j(x) = f(x_j) + (x - x_j)^T f(x_j)$, удовлетворяющие условию $l_j(\underline{x}_i) = \underline{f}_i$. Таким образом, как правило, $k_{i(s)} \leq n$, где n — число переменных минимизируемой функции f .

Наконец, при практическом использовании метода Lev_λ необходимо было выбрать и зафиксировать значение параметра $\lambda \in (0, 1)$. Если исходить из оценки (5), то наилучшим значением λ можно было бы считать точку максимума функции $c(\lambda)$ на

$[0, 1]$. Такой точкой, как нетрудно проверить, является $\lambda^* = 1 - 2^{-\frac{1}{2}} \approx 0,293$. Однако, как показали вычислительные эксперименты, сходимость Lev_λ при $0,2 \leq \lambda \leq 0,8$ слабо зависит от λ и резко ухудшается при подходе λ к концам $[0, 1]$, поэтому при практической реализации Lev_λ параметр λ зафиксирован значением $\frac{1}{2}$.

Метод уровней был подвергнут интенсивному тестированию. Мы не будем касаться здесь результатов экспериментов, описанных в [6], где Lev_λ использовался главным

образом для минимизации на параллелепипедах, а остановимся на вычислительном опыте применения метода уровней к двум классам задач транспортного типа. В обоих случаях вычисления были основаны на декомпозиционном подходе из разд. 5.

Первый класс составляют трехиндексные транспортные задачи T_1 с аксиальными суммами, имеющие вид

$$\sum_{i_1=1}^p \sum_{i_2=1}^q \sum_{i_3=1}^r c_{i_1 i_2 i_3} x_{i_1 i_2 i_3} \rightarrow \max,$$

$$\sum_{i_2=1}^q \sum_{i_3=1}^r x_{i_1 i_2 i_3} = \alpha_{i_1}, \quad \sum_{i_1=1}^p \sum_{i_3=1}^r x_{i_1 i_2 i_3} = \beta_{i_2},$$

$$\sum_{i_1=1}^p \sum_{i_2=1}^q x_{i_1 i_2 i_3} = \gamma_{i_3}, \quad x_{i_1 i_2 i_3} \geq 0,$$

$$i_1 = 1, \dots, p, \quad i_2 = 1, \dots, q, \quad i_3 = 1, \dots, r.$$

Предполагается, что $p \leq q \leq r$. При применении к T_1 двойственного декомпозиционного подхода роль связывающих ограничений играла система уравнений

$$\sum_{i_2=1}^q \sum_{i_3=1}^r x_{i_1 i_2 i_3} = \alpha_{i_1}, \quad i_1 = 1, \dots, p, \text{ так что функция } f, \text{ минимизируемая методом уровней,}$$

имеет p переменных. Заметим, что в данном случае задача (42), решаемая на каждом шаге метода, легко приводится к единственной транспортной задаче в матричной постановке, для решения которой использовалась программа MEXTR из [11]. Размеры решенных задач типа T_1 определяются условиями $5 \leq p \leq 40$, где p кратно 5; $p \leq q \leq r \leq 100$; q и r кратны 10; $prq \leq 324\,000$ (ограничения программы). Для каждого набора (p, q, r) решалась серия из 15 задач, в которых исходные данные выбирались с помощью датчика случайных чисел. Общее число решенных задач типа T_1 равно 4875.

Второй класс составляют многопродуктовые транспортные задачи на сети, часть дуг которой имеет ограниченную пропускную способность по совокупности продуктов. Вычислительный эксперимент был проведен для подкласса этого класса, состоящего из задач минимизации суммарных издержек при снабжении s продуктами, находившимися в m пунктах, n потребителей (причем снабжение осуществляется через p баз, имеющих ограниченную вместимость). Назовем эти задачи T_2 . Они решались также на основе декомпозиционного подхода, изложенного в разд. 5, причем в качестве связывающих выступали ограничения на вместимость баз, поэтому, как и в задачах типа T_1 , метод уровней применялся для минимизации функции f , имеющей p переменных. В рассматриваемом случае решаемая на каждой итерации задача (42) представляет собой s (по числу продуктов) транспортных сетевых задач. Для решения последних использовалась программа CNEXTR из [12]. Размеры решенных задач типа T_2 таковы: $5 \leq p \leq 50$, p кратно 5; $10 \leq m \leq n \leq 100$, m, n кратно 10; $s \in \{2, 3, 5, 10, 15, 20\}$. Для каждого набора (p, m, n, s) решалась серия из 15 задач, исходные данные для которых выбирались с помощью случайного генератора. Всего было решено 8250 различных задач типа T_2 .

Процесс решения задачи типа T_1 или T_2 прерывался после соблюдения условия:

$$\frac{\Delta_k}{|f_{-k}|} \leq \theta, \text{ т.е. после получения приближенного решения задачи с относительной точ-}$$

ностью по функционалу, равной θ . Вычислительные эксперименты проводились при $\theta = 10^{-g}$, где $g \in \{3, 4, 5\}$. Расчеты велись на персональных компьютерах IBM 386 и 486 (с сопроцессором, тактовая частота 33 МГц, объем оперативной памяти 4 Мб) с помощью программ, написанных на языке FORTRAN (транслятор NDP FORTRAN, версия 2.0). Вся информация хранилась в оперативной памяти. Проведенные расчеты,

s	t_l	t_s	t_s/t_l	s	t_l	t_s	t_s/t_l
3	33	345	10,4	25	102	6611	64,8
10	54	1203	22,3	30	118	9822	83,2
15	73	2571	35,2	35	138	13255	96,1
20	80	4582	57,3	40	175	17732	101,3

подробное описание которых содержится в [13, 14], показали высокую эффективность и надежность метода уровней при решении всех тестовых задач.

Остановимся на некоторых выводах из вычислительного эксперимента. Назовем $\delta_i = \frac{f(\bar{x}_i) - f^*}{LD}$ относительной погрешностью рекорда $\bar{x}_i$, полученного после i шагов метода Lev_λ , где L – константа Липшица функции f , заданной на компакте G , D – диаметр G . Согласно теореме 1, имеет место оценка

$$\delta_i \leq c(\lambda) i^{-\frac{1}{2}}, \quad i = 1, \dots, \quad (59)$$

из которой, как отмечалось, следует оптимальность метода уровней при невысокой относительной погрешности $\delta_i (n^{-\frac{1}{2}} < \delta_i < 1$, где n – размерность $E^n \supset Q$). Это теоретически установленное свойство Lev_λ не имеет, конечно, большого прикладного значения: если бы скорость сходимости определялась правой частью (59) при высоких требованиях к относительной погрешности $\delta_i, 0 < \delta_i < m^{-1}$, то он никогда не нашел бы практического применения. Однако вычислительные эксперименты, а также опыт, описанный в [6], где число переменных минимизируемых функций доходило до 400, показали, что метод уровней ведет себя как оптимальный (в смысле [7]) при всех значениях относительной погрешности.

Положим $\bar{\delta}_i = (f(\bar{x}_i) - f^*) / (\max_{1 \leq j < i} f(x_j) - \min_{1 \leq j \leq i} f(x_j))$. Очевидно, $\delta_i \leq \bar{\delta}_i$, причем $\bar{\delta}_i$, в отличие от δ_i , легко определяется в ходе счета по методу уровней. Во всех вычислительных экспериментах $\bar{\delta}_i$ убывала со скоростью геометрической прогрессии, знаменатель которой зависит от размерности пространства p так же, как и у оптимального метода при высоких требованиях к точности. Именно $\bar{\delta}_i \leq \exp\left(-\frac{i}{\mu p}\right)$, $i = 1, \dots$, где μ , характеризующий "трудность" задачи, всегда меньше единицы и, как правило, лежит в пределах между 0,2 и 0,7.

Вычислительный эксперимент позволил также получить представление о трудоемкости отдельной итерации метода уровней. Во всех решенных задачах типа T_1 и T_2 средняя трудоемкость отдельной итерации метода оказалась величиной порядка p^2 . В частности, в процессе эксперимента с задачами типа T_2 для метода уровней была выявлена следующая зависимость времени t_l его работы, необходимого для достижения относительной погрешности $\bar{\delta}$, от p

$$t_l = cp^3 \ln(1/\bar{\delta}), \quad (60)$$

где $c \in [2,7 \times 10^{-4}; 9,6 \times 10^{-4}]$, причем время t_l и концы отрезка изменения параметра c измеряются в секундах. Отрезок изменения c в (60) отвечает скорости персонального компьютера IBM 386 (33).

В заключение приведем результаты сравнительных расчетов задачи типа T_2 при $m = n = p = 15$ и различных s с помощью двойственного декомпозиционного метода, основанного на Lev_1 , и симплекс-алгоритма (пакет MINOS). В четырех столбцах

таблицы записаны соответственно значения s (количество продуктов), t_l (время в

секундах решения задачи на основе метода уровней), t_s (время в секундах решения задачи симплекс-методом), t_s/t_l (выигрыш во времени, раз.) Как видно из таблицы, t_s/t_l меняется от 10 до 100 в зависимости от числа продуктов, что подтверждает целесообразность применения двойственного декомпозиционного подхода из разд. 5 для решения многопродуктовых транспортных задач.

ЛИТЕРАТУРА

1. Kelley J.E. The Cutting-Plane Method for Solving Convex Programs // J. Soc. Industr. Appl. Math. 1960, № 4.
2. Lemarechal C. Nonsmooth Optimization and Descent Methods // Research Report 78-4. IASA Laxenburg, 1978.
3. Lemarechal C., Strodiot J.J., Bihain A. On a Bundle Algorithm for Nonsmooth Optimization // Nonlinear Programming V. 4. L.: Acad. Press. 1981.
4. Martinet B. Régularisation d'inéquations variationnelles par approximations successives // Rev. Française Inf. Rech. Oper. 1970. № 2.
5. Moreau J.J. Proximité et dualité dans un espace hilbertien // Bull. Soc. Math. France. 1965. V. 93, № 3.
6. Lemarechal C., Nemirovskii A., Nesterov Yu. New Variants of Bundle Methods // Research Report N 1508, Institut National de Recherche en Informatique et en Automatique (INRIA). Paris, 1991.
7. Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. М.: Наука, 1979.
8. Гольштейн Е.Г. Об одном двойственном блочном методе линейного программирования // Автоматика и телемеханика. 1994. № 12.
9. Powell M.J.D. A FORTRAN Subroutine for Convex Programming // Report NA17. Cambridge: Univ. Cambridge, 1983.
10. Goldfarb D., Idnani A. Numerically Stable Dual Method for Solving Strictly Convex Quadratic Problems // Math. Program. 1983. V. 27. № 1.
11. Программа MEXTR решения матричной транспортной задачи / Пакет анализа оптимизационных моделей ППП ПАОЭМ ЕС, задачи транспортного типа. Калинин: НПО Центр ПрограммСистем, 1982.
12. Программа CNEXTR решения сетевой транспортной задачи с ограничениями пропускной способности // Пакет анализа оптимизационных моделей ППП ПАОЭМ ЕС, задачи транспортного типа. Калинин: НПО Центр ПрограммСистем, 1982.
13. Гольштейн Е.Г., Немировский А.С., Соколов Н.А. Новый алгоритм двойственной декомпозиции с приложением к задачам транспортного типа // Экономика и мат. методы. 1994. Т. 30. Вып. 2.
14. Гольштейн Е.Г., Соколов Н.А. Новый метод решения многопродуктовой транспортной задачи // Экономика и мат. методы. 1995. Т. 31. Вып. 2.

Поступила в редакцию
20 XII 1994