

НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ. ИГРЫ

ЭФФЕКТИВНЫЕ ИТЕРАТИВНЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ С МОНОТОННЫМИ ОПЕРАТОРАМИ

НЕМИРОВСКИЙ А. С.

(Москва)

1. ВВЕДЕНИЕ

1. Многие важные задачи математической экономики (некоторые примеры даются ниже) можно сформулировать в терминах уравнений с монотонными операторами. Напомним, что монотонным оператором на подмножестве G n -мерного евклидова пространства E_n (ограничимся пока только этим случаем) называется отображение $F(x)$, сопоставляющее каждому $x \in G$ непустое множество $F(x) \subset E_n$, причем выполняется условие монотонности: $(\xi - \eta, x - y) \geq 0$ для всех $x, y \in G, \xi \in F(x), \eta \in F(y)$. Уравнения с монотонными операторами, рассматриваемые далее, условно записываются в виде $x \in G, 0 \in F(x)$. Решением такого уравнения, по определению, называется всякая точка $x^* \in G$, такая, что $(\xi, x - x^*) \geq 0$ для всех $x \in G$ и $\xi \in F(x)$. В литературе по монотонным операторам решение часто понимается в более «сильном» смысле, например, так: $x \in G$ — решение, если $\exists \xi \in F(x): (\xi, x - y) \leq 0 \quad \forall y \in G$. Решение уравнения в смысле любого из существующих определений является и решением в смысле определения, исходного для данной статьи. Мы предпочитаем иметь дело со «слабейшим» определением потому, что для него имеется самая сильная теорема существования (см. разд. 2); в то же время для приложений обычно вполне достаточно уметь находить и «слабое» решение, которое к тому же при естественных предположениях об операторе оказывается и «сильным».

В настоящей работе описывается несколько итеративных методов приближенного решения уравнений с монотонными операторами.

2. Приведем примеры задач, сводящихся к уравнениям рассматриваемого типа.

I. Минимизация выпуклой функции. Пусть f — непрерывная выпуклая функция на E_n ; $\partial f(x)$ — множество опорных функционалов f в точке x ; G — выпуклое замкнутое подмножество E_n . Оператор $F(x) = \partial f(x)$ — монотонный на E_n , и, как легко видеть, решения уравнения $0 \in F(x), x \in G$ суть в точности точки минимума f на G .

II. Решение выпукло-вогнутой антагонистической игры двух лиц. Как известно, в этой задаче имеются два игрока, выбирающие ходы соответственно в выпуклых замкнутых множествах $G_1 \subset E_{n_1}$ и $G_2 \subset E_{n_2}$. При выборе первым игроком хода x_1 , а вторым x_2 , иными словами, в партии $x = (x_1, x_2)$, первый платит второму сумму $f(x_1, x_2)$. Функция f предполагается непрерывной на $E_{n_1+n_2} = E_{n_1} \times E_{n_2}$, выпуклой по x_1 и вогнутой по x_2 . Решением игры называется всякая седловая точка $x^* = (x_1^*, x_2^*)$ функции

f на $G=G_1 \times G_2$, т. е. такая точка G , в которой f достигает максимума по $x_2 \in G_2$ и минимума по $x_1 \in G_1$. Отыскание решения игры можно свести к решению уравнения с монотонным оператором: пусть $F(\bar{x}) = \{(\xi, \eta) \in E_{n_1+n_2} \mid \xi \text{ опорно к } f(x_1, \bar{x}_2) \text{ как функции } x_1 \text{ в точке } x_1 = \bar{x}_1, -\eta \text{ опорно к } f(\bar{x}_1, x_2) \text{ как функции } x_2 \text{ в точке } x_2 = \bar{x}_2\}$. Проверяется, что $F(x)$ монотонно, а решения игры суть в точности решения уравнения $x \in G, 0 \in F(x)$.

III. Решение выпукло-вогнутой бескоалиционной игры N лиц. Эта задача обобщает задачу II. Имеются N игроков, i -й из которых может выбирать свой ход x_i в пределах выпуклого замкнутого множества $G_i \in E_{n_i}$. При выборе игроками набора ходов $x_1, \dots, x_N$ — в партии $x = (x_1, \dots, x_N) \in G \equiv G_1 \times \dots \times G_N \subset E_{n_1+\dots+n_N}$ i -й игрок теряет сумму $f_i(x)$. Предполагается, что $f_i(x)$ — непрерывная функция на $E_{n_1+\dots+n_N}$, выпуклая по x_i и вогнутая по совокупности остальных переменных, $i=1, \dots, N$, причем функция

$$\varphi(x) = \sum_{i=1}^N f_i(x) \text{ выпукла. Решением рассматриваемой игры называется}$$

всякая ее точка равновесия, т. е. такая точка $x^* = (x_1^*, \dots, x_N^*) \in G$, в которой каждая из функций f_i достигает минимума по переменной $x_i \in G_i$. Иными словами, точка равновесия x^* характеризуется тем, что каждому игроку i невыгодно менять ход x_i^* на любой другой ход из G_i , если только все прочие игроки j придерживаются ходов x_j^* . Как и в случае игр двух лиц (сводящемуся к данному при $N=2$ и $f_2 = -f_1$), решение рассматриваемой игры сводится к решению уравнения $0 \in F(x)$, $x \in G$ при выборе $F(\bar{x}) = \{(\xi_1, \dots, \xi_N) \in E_{n_1+\dots+n_N} \mid \forall i \ \xi_i \text{ — опорный функционал к } f_i(\bar{x}_1, \dots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \dots, \bar{x}_N) \text{ как функции } x_i \text{ в точке } x_i = \bar{x}_i\}$. Оператор $F(x)$ оказывается монотонным, а решения указанного уравнения суть в точности точки равновесия рассматриваемой игры.

Из сказанного ясно, что методы решения уравнений с монотонными операторами могут найти существенные приложения, в том числе и в экономических задачах.

3. В настоящее время теория монотонных операторов значительно продвинута как в плане изучения самих этих операторов [1], так и в направлении разработки и исследования численных методов решения уравнений с этими операторами (например, [2–5]). Однако последнее изучалось в основном применительно к операторам, удовлетворяющим некоторым условиям гладкости и (или) «сильной» монотонности; более общий случай исследован относительно хуже (подробнее см. в разд. 4). В данной работе внимание сосредоточено на ситуации, когда от оператора не требуется почти ничего сверх монотонности; наша цель состоит в распространении на этот случай методов решения выпуклых экстремальных задач из [6]. Будет показано, что это действительно можно сделать, причем не ухудшая по порядку скорости сходимости методов. Более того, построенные методы окажутся в некотором смысле оптимальными — их трудоемкость нельзя существенно снизить, не ухудшив точности и не сужая класс рассматриваемых уравнений.

2. ОПРЕДЕЛЕНИЯ И ОБОЗНАЧЕНИЯ

1. Для дальнейшего изложения целесообразно несколько расширить и уточнить постановку задачи.

Пусть $E, \|\cdot\|$ — вещественное рефлексивное банахово пространство (при желании можно считать, что $E=R^n$, а $\|\cdot\|$ — норма на R^n вида

$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$, $1 \leq p < \infty$, или вида $\|x\|_\infty = \max_i |x_i|$. Пространство R^n

с нормой $\|\cdot\|_p$, $1 \leq p \leq \infty$, будет обозначаться l_p^n . Ясно, что l_2^n — просто евклидово пространство E_n . Пусть E^* , $\|\cdot\|$ — пространство, сопряженное к E , $\|\cdot\|$. Для $\varphi \in E^*$ и $x \in E$ (φ, x) — значение функционала φ на векторе x (для E , $\|\cdot\| = l_p^n E^*$, $\|\cdot\|_* = l_{p^*}^n$, где $p^* = p/(p-1)$); при этом

$(\varphi, x) = \sum_{i=1}^n \varphi_i x_i$ — обычное скалярное произведение в R^n). Пусть G — выпук-

лое замкнутое подмножество E и $F(x)$ для каждого $x \in G$ есть непустое подмножество E^* ; назовем $F(x)$ оператором на G . Такой оператор называется монотонным (на G), если

$$(\eta - \xi, y - x) \geq 0 \text{ для всех } x, y \in G, \eta \in F(y), \xi \in F(x). \quad (1)$$

Будем решать уравнения вида

$$0 \in F(x), \quad x \in G. \quad (2)$$

По определению, решением (2) называется всякая точка $x^* \in G$, такая, что

$$(\xi, x - x^*) \geq 0 \quad \forall x \in G \quad \forall \xi \in F(x). \quad (3)$$

Известно, что в сделанных предположениях относительно E , G при ограниченном G и монотонном F уравнение (2) необходимо имеет решения.

Скажем, что F ограничено, если конечна величина $\|F\|_G = \sup \{\|\xi\|_* \mid \xi \in \bigcup_{x \in G} F(x)\}$. Везде далее для (2) предполагается, что G ограничено (и, как говорилось выше, выпукло и замкнуто), E рефлексивно и F ограничено и монотонно на G . Положим $r_{\|\cdot\|}(F) = d_{\|\cdot\|}(G) \|F\|_G$, где $d_{\|\cdot\|}(G) = \sup_{x, y \in G} \|x - y\|$ — диаметр G в норме $\|\cdot\|$. Допустим, что $r_{\|\cdot\|}(F) > 0$; только этот случай нетривиален.

2. Рассмотрим методы приближенного решения уравнений (2), полагая доступными значения F в точках G (наблюдаемые точно или с погрешностями). Чтобы говорить о приближенных методах, надо научиться измерять погрешность точек $x \in G$ в качестве приближенных решений (2). Введем соответствующую меру следующим образом. Пусть $\varepsilon(x, F) = \sup \{(\xi, x - y) \mid y \in G, \xi \in F(y)\}$. Ясно, что при $x \in G$ $0 \leq \varepsilon(x, F) \leq r_{\|\cdot\|}(F)$, причем $\varepsilon(x, F) = 0$ тогда и только тогда, когда x — решение (2). Поэтому величину $\varepsilon(x, F)$ естественно принять за погрешность точки $x \in G$ в качестве приближенного решения (2); назовем эту меру погрешности абсолютной. Отметим, что она не вполне традиционна; обычно погрешность измеряют расстоянием от x до множества точных решений (2). Последний способ в отличие от нашего неудобен тем, что задача отыскания хорошего в этом смысле приближенного решения (2) оказывается некорректно поставленной, если не делать относительно F дополнительных предположений типа сильной монотонности (ср. обсуждение аналогичного вопроса для задач оптимизации в [6, стр. 23–24]).

С математической точки зрения удобно определенным образом нормировать абсолютную погрешность — перейти к относительной погрешности $v_{\|\cdot\|}(x, F) = \varepsilon(x, F) / r_{\|\cdot\|}(F)$.

Из сказанного следует, что $0 \leq v_{\|\cdot\|}(x, F) \leq 1$ при $x \in G$. Удобство относительной погрешности состоит в том, что задача отыскания приближенного

решения (2) с относительной погрешностью $v_{\parallel\parallel}(x, F) \leq v$ ($v > 0$ задано) оказывается, как мы увидим, корректной в том смысле, что трудоемкость (число итераций) ее решения строящимися далее методами оценивается сверху величиной, зависящей лишь от v , G и E , $\|\cdot\|$, но не от конкретного уравнения (2).

Важно заметить, что и (2), и абсолютная погрешность $\varepsilon(x, F)$ описываются в терминах, среди которых не фигурирует явно никакая конкретная норма на E (из числа задающих топологию E), тогда как относительная погрешность $v_{\parallel\parallel}(x, F)$ связана с конкретной нормой. К этому вопросу мы еще вернемся в п. 2 разд. 4.

3. Введенные меры погрешности, как указывалось (и как будет видно из дальнейшего), естественны и удобны в математическом отношении; это, однако, не снимает вопроса об их содержательной полезности. Последнюю нельзя оценить однозначно — по уравнению (2) невозможно понять, какой содержательной задаче оно соответствует и приемлемы ли с точки зрения этой задачи «хорошие» в смысле мер $\varepsilon(x, F)$ и $v_{\parallel\parallel}(x, F)$ приближенные решения (2) (это, впрочем, относится и к любой другой мере погрешности). Тем не менее в конкретных ситуациях этот вопрос можно решить; рассмотрим его для уравнений, отвечающих задачам I—III Введения. Положим $\varepsilon_F(x) = 2v_{\parallel\parallel}(x, F) \ln(2/v_{\parallel\parallel}(x, F))r_{\parallel\parallel}(F)$; $\varepsilon_F(x) \rightarrow 0$ при $v_{\parallel\parallel}(x, F) \rightarrow 0$, причем $\varepsilon_F(x)$ имеет «почти» тот же порядок малости, что и $v_{\parallel\parallel}(x, F)$. Нетрудно видеть, что в задаче I для $\bar{x} \in G$: $f(\bar{x}) - \min_G f \leq \varepsilon_F(\bar{x})$ — малость $v_{\parallel\parallel}(\bar{x}, F)$ означает, что $\bar{x}$ — хорошее приближение к точке минимума f на G «по функционалу». В задаче II $f(\bar{x}) - \min_{x_1 \in G_1} f(x_1, \bar{x}_2) \leq \varepsilon_F(\bar{x})$ и $\max_{x_2 \in G_2} f(\bar{x}_1, x_2) - f(\bar{x}) \leq \varepsilon_F(\bar{x})$. Отход каждого игрока от его хода в партии $\bar{x}$ не принесет ему выгоды, большей $\varepsilon_F(\bar{x})$, если другой игрок не откажется от своего хода из этой партии (такие точки $\bar{x}$ в теории игр называются точками ε -равновесия [7]; в нашем случае $\bar{x}$ есть точка $\varepsilon_F(\bar{x})$ -равновесия). Таким образом, чем меньше $v_{\parallel\parallel}(\bar{x}, F)$, тем меньше стимулов у каждого из игроков отказываться от отвечающего партии $\bar{x}$ хода.

В задаче III интерпретация нашей меры погрешности та же, что и в задаче II. Именно, при любом $i \leq N$, как оказывается,

$$f_i(\bar{x}) - \min_{x_i \in G_i} f_i(\bar{x}_1, \dots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \dots, \bar{x}_N) \leq \varepsilon_F(\bar{x}).$$

Мы видим, что погрешность $v_{\parallel\parallel}(x, F)$ в задачах I—III допускает естественную интерпретацию в соответствующих содержательных терминах; уже это служит достаточным неформальным основанием для использования этой меры погрешности.

Перейдем к описанию некоторых методов приближенного решения (2).

3. МЕТОДЫ С ЛИНЕЙНОЙ СХОДИМОСТЬЮ

1. Распространим на уравнения (2) методы выпуклой оптимизации из [6, гл. II], приходя к методам, сходящимся со скоростью геометрической прогрессии.

Ниже E , $\|\cdot\|$ предполагается конечномерным пространством размерности n . В этом случае, не ограничив общности, можно считать, что внутренность $\text{int } G$ множества G в (2) непуста; далее это предполагается.

Опишем два итеративных метода решения (2), сходящихся со скоростью геометрической прогрессии со знаменателем $1 - O(1/n)$ для первого метода и $1 - O(1/n^2)$ — для второго. Причина, по которой рассматриваются оба метода, а не только первый — более быстрый, будет указана ниже.

2. Наметим идею методов. Первое соображение состоит в следующем. Пусть $G_0 \supset G$ и x_1 — внутренняя точка G_0 , причем G_0 выпукло, замкнуто и ограничено. Если $x_1 \notin \text{int } G$, то через точку x_1 можно провести гиперплоскость $\Pi_1 = \{x \mid (\xi_1, x - x_1) = 0\}$ такую, что G содержится в одной из двух частей, на которые G_0 делится плоскостью Π_1 — в части $\bar{G}_1 = \{x \in G_0 \mid (\xi_1, x - x_1) \leq 0\}$. Если же $x_1 \in \text{int } G$, то, узнав какой-нибудь элемент $\xi_1 \in F(x_1)$, построим $\Pi_1, \bar{G}_1$ по тем же формулам, что и выше, причем из монотонности F следует, что все решения x^* уравнения (2) лежат в $\bar{G}_1$. Итак, во всех случаях можно (узнав значение F в x_1) построить строго меньшую, чем G_0 , область $\bar{G}_1$, содержащую все решения (2). Ясно, что и любая область $G_1 \supset \bar{G}_1$ обладает тем же свойством. Выбрав какую-нибудь область G_1 такого типа и $x_2 \in \text{int } G_1$, повторим процесс, и т. д. В результате получим последовательность областей $G_0, G_1, \dots, G_k, \dots$, каждая из которых есть область локализации решений в том смысле, что она содержит все решения (2). Если бы удалось придумать такие правила выбора x_i в G_{i-1} и правила «раздувания» $\bar{G}_i$ до G_i , при которых области G_i неограниченно убывали в каком-нибудь разумном смысле, можно было бы надеяться на сходимость нашего процесса к решению (2). Разумной мерой величины областей в этой ситуации оказывается n -мерный лебегов объем $|G_i|$ областей G_i ; при такой мере их величины действительно можно добиться быстрого убывания областей. Наиболее эффективная здесь стратегия, впервые предложенная в [8], состоит в выборе $G_0 = G, G_i = \bar{G}_i$ и центра тяжести G_i в качестве x_{i+1} . Эта стратегия, однако, безнадёжна в смысле реализации уже для сколько-нибудь заметных n (из-за катастрофического роста сложностей отыскания центров тяжести при росте n). С целью обойти эту трудность в [9] предложено брать в качестве G_i эллипсоид наименьшего объема, содержащий $\bar{G}_i$, а в качестве x_{i+1} — его центр; при этом G_0 — какой-нибудь содержащий G эллипсоид. При таком способе действий объемы G_i убывают с ростом i в геометрической прогрессии, хотя и с меньшей, чем для первоначальной стратегии, скоростью; зато осуществление итерации теперь не составляет заметной сложности.

Обе эти стратегии будут реализованы ниже. Отметим, что сама идея последовательной локализации решений (2) в областях быстро убывающего объема целиком заимствуется из оптимизационной ситуации; на возможность такого заимствования впервые, насколько нам известно, указал Н. З. Шор [10].

Однако локализация точного решения (2) в области малого объема никоим образом не решает вопроса. Дело в том, что эта область не обязана быть малой в смысле линейных размеров, и вовсе не все ее точки суть сколько-нибудь хорошие приближенные решения (2). Надо еще при локализации решения в области малого объема уметь находить в ней хорошее приближенное решение. Эта задача в оптимизационной ситуации решается довольно просто (см. [9] или [6, гл. II, § 5]); в случае уравнений (2) приходится применять совершенно другой прием (см. ниже). В конечном счете все же удастся построить методы, сходящиеся со скоростью геометрической прогрессии в смысле «настоящей» погрешности, а не только в не имеющей ясной интерпретации в терминах (2) «сходимости к нулю объемов очередных областей локализации решения».

3. Перейдем теперь к описанию предлагаемых методов. Оба они устроены до некоторой степени одинаково. Каждый строит две последовательности точек $E - \{x_i\}$ и $\{y_i\}$, а также некоторую вспомогательную последовательность выпуклых областей $\{G_i\}$. Точка x_i является «поисковой»; если $x_i \in \text{int } G$, то на i -м шаге задается вопрос о значении $F(x_i)$; предполагается, что в ответ на этот вопрос сообщается некий вектор

$\varphi(x_i) \in F(x_i)$. Строящаяся на i -м шаге точка $y_i \in \text{int } G$ является i -м последовательным приближением к решению.

Схема работы обоих методов такова. К i -му шагу имеется некоторая выпуклая замкнутая ограниченная область G_{i-1} , $\text{int } G_{i-1} \neq \emptyset$. Для первого метода $G_{i-1} \subset G$ ($G_0 = G$); для второго G_{i-1} — эллипсоид, пересекающийся с $\text{int } G$; в качестве G_0 выбирается какой-нибудь эллипсоид, содержащий G . На i -м шаге проводятся следующие операции.

Построение G_i . Строится центр тяжести x_i области G_{i-1} . При $x_i \in \text{int } G$ в точке x_i задается вопрос о значении $F(x_i)$; пусть $\varphi(x_i)$ — сообщенный вектор из $F(x_i)$. Если $\varphi(x_i) = 0$, то работа метода прекращается; полагаем $y_j = x_i$, $j \geq i$; отметим, что тогда y_j , $j \geq i$ — точные решения (2).

При $\varphi(x_i) \neq 0$ строим область G_i следующим образом: пусть $\bar{G}_i = \{x \in G_{i-1} \mid (\varphi(x_i), x - x_i) \leq 0\}$.

Для первого метода берется $G_i = \bar{G}_i$; для второго в качестве G_i используется эллипсоид минимального объема, содержащий $\bar{G}_i$.

При $x_i \notin \text{int } G$ (из сказанного выше следует, что такая ситуация может возникнуть лишь во втором методе) пусть φ_i — ненулевой функционал, опорный к функции $\rho(x, G) = \min_{y \in G} \|x - y\|$ в точке $x = x_i$. Полагаем

$\bar{G}_i = \{x \in G_{i-1} \mid (\varphi_i, x - x_i) \leq 0\}$ и, как и выше, выбираем в качестве G_i эллипсоид минимального объема, содержащий $\bar{G}_i$.

Замечание. С вычислительной точки зрения описанные операции, требуемые первым из методов, крайне сложны при сколько-нибудь заметных n . Напротив, реализация этих операций для второго метода не встречает затруднений, так как здесь мы имеем дело с телами простой формы — эллипсоидами и их «половинами». Именно это обстоятельство и приводит к необходимости рассматривать оба метода, а не только первый, лучший по скорости сходимости. По поводу алгоритмической реализации второго из методов см. [6, гл. II, § 5], где он описан применительно к решению выпуклых экстремальных задач; с точки зрения организации метода эта ситуация ничем не отличается от рассматриваемой в данной работе. В [6] показано, что реализация операций для этого метода требует (при памяти $O(n^2)$) $O(n^2)$ арифметических действий на каждом шаге.

Построение y_i . Пусть $\bar{y}$ — раз и навсегда фиксированная точка G . Для каждого i пусть $I_i = \{j \mid 1 \leq j \leq i, x_j \in \text{int } G\}$. Если $I_i = \emptyset$ (это возможно лишь для второго метода), то положим $y_i = \bar{y}$.

В противном случае пусть m_i — число элементов в I_i , а j_s , $1 \leq s \leq m_i$ — s -й по порядку элемент в I_i . Пусть $\Lambda_i = \left\{ \lambda = (\lambda_1, \dots, \lambda_{m_i}) \mid \lambda_s \geq 0, \sum_{s=1}^{m_i} \lambda_s = 1 \right\}$.

Рассмотрим вогнутую функцию

$$\theta_i(\lambda) = \min_{y \in G} \sum_{s=1}^{m_i} \lambda_s (\varphi(x_{j_s}), y - x_{j_s}) \equiv \min_{y \in G} \bar{\theta}_i(\lambda, y),$$

максимизируем ее по $\lambda \in \Lambda_i$. Пусть $\bar{\lambda} = (\bar{\lambda}_1, \dots, \bar{\lambda}_{m_i})$ — соответствующее λ . Полагаем

$$y_i = \sum_{s=1}^{m_i} \bar{\lambda}_s x_{j_s}. \quad (4)$$

Замечание. Операции, описанные выше, нельзя назвать простыми, но и принципиальных затруднений с ними нет. Отыскание y_i требует реше-

ния билинейной игры на $\Lambda_i \times G$; если G — многогранник, то эта задача сводится к задаче линейного программирования. К тому же не обязательно выполнять предлагаемые предписания на каждой итерации — достаточно сделать это, лишь когда требуется остановить метод.

4. Теорема 1. *Описанные методы формируют последовательности точек y_i , сходящиеся к решению уравнения (2): $v_{\|\cdot\|}(y_i, F) \xrightarrow{i \rightarrow \infty} 0$. При этом для всякого v , $0 < v < 1$, неравенство $v_{\|\cdot\|}(y_i, F) \leq v$ выполняется: 1) в случае первого метода — при всех i , удовлетворяющих условию*

$$i \geq]2, 2n \ln 1/v[, \quad (5)$$

2) в случае второго метода — при всех i , удовлетворяющих условию

$$i \geq]cn^2 \ln \beta/v[. \quad (6)$$

Здесь $\beta = (|G_0|/|G|)^{1/n}$, $|A|$ — объем тела A , $]t[= [-t]$ — наименьшее целое, $\geq t$, c — подходящая абсолютная константа.

Доказательство теоремы 1 вынесено в п. 7.

5. Проиллюстрируем работу методов на простейшем примере. Пусть $E = E_2$, $G = \{(x^1, x^2) \mid -1 \leq x^1 \leq 2, -1 \leq x^2 \leq 1\}$, $F(x^1, x^2) = (x^2, -x^1)$ (фактически мы находимся в ситуации задачи II с $f(x) = x^1 x^2$). Удобнее рассматривать первый метод. Первая поисковая точка $x_1 = (0,5; 0)$; область G_1 есть часть G , высекаемая неравенством $x^2 \geq 0$. Нетрудно видеть, что последующие области G_i суть пересечения углов с вершиной в 0 с прямоугольником G , причем одна из сторон этих углов — отрицательный луч оси абсцисс, а другая приближается к этому лучу со скоростью геометрической прогрессии; при этом $x_i \xrightarrow{i \rightarrow \infty} (-0,5; 0)$. Таким образом, G_i стягиваются к отрезку $[-1, 0]$ вещественной оси, а x_i ни в каком смысле не сходятся к искомому решению $x^* = 0$. Тем не менее, при построении y_i точки x_i усредняются так, что результаты усреднений сходятся к 0, хотя, напомним еще раз, наш метод ни разу не «опрашивал» решаемое уравнение вблизи 0.

6. Скажем несколько слов об эффективности построенных методов. Будем понимать под методом решения уравнений (2) любой детерминированный процесс построения последовательных приближений y_i^F к решению (2) на основе информации о значениях F в тех или иных точках G ; считается, что y_i^F строится на основе первых i из этих вычислений (более четкое определение вводимых здесь понятий см. [11]). Трудоемкостью метода $\mathfrak{B}$ на классе уравнений (2) с данными G и E , $\|\cdot\|$ назовем функцию $M_{G, E, \|\cdot\|}^{\mathfrak{B}}(v)$, определяемую как минимальное i , такое, что $v_{\|\cdot\|}(y_i^F, F) \leq v$ для всех уравнений (2) рассматриваемого класса. В этих терминах для первого из методов этого раздела $\mathfrak{B}_1$ имеем $M_{G, R^n, \|\cdot\|}^{\mathfrak{B}_1}(v) \leq]2, 2n \ln 1/v[$, для второго — $M_{G, R^n, \|\cdot\|}^{\mathfrak{B}_2}(v) \leq]cn^2 \ln \beta/v[$. Оказывается, что в асимптотике по $v \rightarrow 0$ трудоемкость этих методов в принципе, т. е. переходом к любому другому методу, нельзя существенно снизить (более чем в абсолютную константу раз для первого метода и более чем в $O(n)$ раз для второго). Это следует из результатов [6, гл. IV], согласно которым для некоторой абсолютной константы $d > 0$ при некотором $v(\|\cdot\|, G) > 0$ для трудоемкости любого метода $\mathfrak{B}$ справедлива нижняя оценка *

$$M_{G, R^n, \|\cdot\|}^{\mathfrak{B}}(v) \geq dn \ln 1/v \quad \text{при } 0 < v < v(\|\cdot\|, G). \quad (7)$$

* Предполагается, что $\text{int } G \neq \emptyset$; оценка (7) набегаает уже на уравнениях, отвечающих задачам выпуклой оптимизации.

Отметим, что если $\|\cdot\|$ есть $\|\cdot\|_\infty$, а G — единичный шар R^n по этой норме, то $v(\|\cdot\|, G) \geq 10^{-1}$.

7. Доказательство теоремы 1. Достаточно проверить второе утверждение теоремы.

1) Вначале проверяется, что в условиях (5) или (6) (в последнем случае — при правильном выборе константы c) для соответствующей версии метода имеем

$$|G_i| \leq v^n |G|, \quad (8)$$

что доказывается так же, как и аналогичное неравенство для оптимизационных «прародителей» рассматриваемых методов (см. [6, гл. II]).

2) В силу (8) $|G_i| < |G|$ (i удовлетворяет (5), соответственно (6)). Но тогда I_i непусто; в самом деле, для первого метода I_i непусто автоматически. Если же применяется второй метод, то из его описания следует, что $G_i \cap G = G_{i-1} \cap G$ при $x_i \notin \text{int } G$; стало быть, при пустом I_i имеем $G_i \supset G$, так что неравенство $|G_i| < |G|$ при таких i невозможно. Итак, I_i непусто. Пусть $y \in G$ и $\xi \in F(y)$. Имеем

$$(\xi, y - y_i) = \sum_{s=1}^{m_i} \bar{\lambda}_s (\xi, y - x_{j_s}) \geq \sum_{s=1}^{m_i} \bar{\lambda}_s (\varphi(x_{j_s}), y - x_{j_s})$$

(мы использовали (1)). Далее, $\sum_{s=1}^{m_i} \bar{\lambda}_s (\varphi(x_{j_s}), y - x_{j_s}) \geq \theta_i(\bar{\lambda})$. Итак,

$(\xi, y - y_i) \geq \theta_i(\bar{\lambda})$ при всех $y \in G$ и $\xi \in F(y)$, т. е.

$$\varepsilon(y_i, F) \leq -\theta_i(\bar{\lambda}). \quad (9)$$

По лемме фон Неймана

$$\begin{aligned} \theta_i(\bar{\lambda}) &= \max_{\lambda \in \Delta_i} \min_{y \in G} \sum_{s=1}^{m_i} \lambda_s (\varphi(x_{j_s}), y - x_{j_s}) = \min_{y \in G} \max_{\lambda \in \Delta_i} \sum_{s=1}^{m_i} \lambda_s (\varphi(x_{j_s}), y - x_{j_s}) = \\ &= \min_{y \in G} \gamma_i(y), \end{aligned}$$

где $\gamma_i(y) = \max_{1 \leq s \leq m_i} (\varphi(x_{j_s}), y - x_{j_s})$.

3) Для завершения доказательства в силу (9) достаточно проверить, что в условиях (8)

$$\min_{y \in G} \gamma_i(y) \geq -v r_{\|\cdot\|}(F). \quad (10)$$

Действительно, пусть y^* — точка минимума $\gamma_i(y)$ на G . Возможно $y^* \notin G_i$. Из описания методов ясно, что для каждого $y \notin G_i$ при некотором $j \in I_i$ выполнено условие $(\varphi(x_j), y - x_j) \geq 0$, т. е. в $G \setminus G_i$ имеем $\gamma_i(y) \geq 0$. Поэтому при $y^* \notin G_i$ (10) выполнено.

Остается разобрать случай $y^* \in G_i$. Пусть $\bar{G}$ — результат подобного сжатия G к y^* в $1/v$ раз. Тогда $|\bar{G}| = v^n |G| > |G_i|$, т. е. $\bar{G}$ не может содержаться в G_i . Последнее означает, что для некоторого $u \in G$ точка $y^* + v(u - y^*) = u$ не лежит в G_i . Раз $y \in G \setminus G_i$, то, как указывалось выше, при некотором $j \in I_i$ имеем $(\varphi(x_j), y - x_j) \geq 0$. Но тогда $(\varphi(x_j), y^* - x_j) + v(\varphi(x_j), u - y^*) = (\varphi(x_j), y - x_j) \geq 0$, т. е. $(\varphi(x_j), y^* - x_j) \geq -v \|\varphi(x_j)\| \|u - y^*\| \geq -v r_{\|\cdot\|}(F)$, и (10) верно по определению $\gamma_i(y)$. Теорема доказана.

4. МЕТОДЫ ЗЕРКАЛЬНОГО СПУСКА

1. Методы из предыдущего раздела не могут нас вполне удовлетворить по двум причинам. Во-первых, они не слишком просты в смысле реализации. Во-вторых (в теоретическом плане это более важно), их «существенная неулучшаемость» по трудоемкости, о которой шла речь в п. 6 разд. 3, имеет место лишь в асимптотике по $\nu \rightarrow 0$, причем выход на эту асимптотику может происходить при весьма малых значениях ν . Первое обстоятельство не запрещает попытки усовершенствовать методы в «доасимптотическом» диапазоне значений ν , а второе позволяет надеяться, что этот диапазон достаточно широк для того, чтобы усовершенствованные методы могли рассчитывать на практическое приложение. Построение такого рода методов и является целью данного раздела.

2. Мы построим, в частности, семейство методов $\mathfrak{B}_n^p$, $n=2, 3, \dots, 1 \leq p < \infty$, решения уравнений (2) на пространствах l_p^n , с оценками трудоемкости

$$M_{G, l_p^n}^{\mathfrak{B}_n^p}(\nu) \leq \begin{cases} b(p)(1/\nu)^{\max(2, p)}, & 1 < p < \infty, \\ b(1) \ln n (1/\nu)^2, & p=1. \end{cases} \quad (11)$$

Здесь $b(p)$ не зависит от n .

Прокомментируем этот результат. Прежде всего, может показаться, что он противоречит (7) — нижняя оценка (7) трудоемкости любого метода линейно увеличивается с ростом n , а верхняя оценка трудоемкости (11) ограничена равномерно по n ($p > 1$) или растет очень медленно — логарифмически ($p=1$). Никакого противоречия между (7) и (11), разумеется, нет, ибо (7) — асимптотическая оценка; условия $\nu < \nu(\|\cdot\|_p, G)$ и $b(p)(1/\nu)^{\max(2, p)} < dn \ln 1/\nu$ просто не могут выполняться одновременно.

Более серьезные недоумения может вызвать тот факт, что строится целое семейство методов, параметризованное с помощью p . Между тем, к каждому из уравнений (2) на R^n можно применить любой из методов $\mathfrak{B}_n^p$ и, казалось бы, вместо целого семейства методов достаточно рассмотреть лишь один из них, лучший в смысле оценки трудоемкости (имеется в виду рассуждение обманчиво. Дело в том, что погрешность каждого из методов $\mathfrak{B}_n^p$ измеряется в своей шкале, определяемой нормой $\|\cdot\|_p$. Если имеется содержательная задача, сводящаяся к уравнению (2), и заданы сформулированные на языке этой задачи требования к погрешности решения, то при применении метода $\mathfrak{B}_n^p$ эти требования надо пересчитать в соответствующую шкалу и относительная погрешность, на которую надо будет настраивать $\mathfrak{B}_n^p$, окажется зависящей от p . Поэтому сравнение правых частей (11), отвечающих различным p при одном и том же ν , не позволяет априорно высказаться в пользу одного из методов семейства. Напротив, меняя параметр p , мы оказываемся в состоянии приспособить до известной степени применяемый метод к специфике задачи.

Приведем пример. Пусть решается матричная игра двух лиц с $n_1 \times n_2$ — матрицей платежей $A = \|A_{ij}\|_{1 \leq i \leq n_1, 1 \leq j \leq n_2}$, а матрица A нормирована

условием $0 \leq A_{ij} \leq 1 \forall i, j$. Пусть используются смешанные стратегии и ищется точка ε -равновесия, $\varepsilon > 0$ задано. В соответствующем уравнении (2) $G = G_1 \times G_2 = \{(p, q) \mid p \in R^{n_1}, p \geq 0, \|p\|_1 = 1, q \in R^{n_2}, q \geq 0, \|q\|_1 = 1\}$, $F = F_A(p, q)$ — однозначный оператор со значениями из $R^{n_1+n_2}$ с координатами

тами значений

$$(F_A(p, q))_i = \begin{cases} \sum_j A_{ij} q_j, & l \leq n_1, \\ -\sum_i A_{il-n_1} p_i, & l > n_1. \end{cases}$$

В силу сказанного в п. 3 разд. 2 для отыскания точки ε -равновесия игры достаточно найти $\bar{x} \in G$ с

$$2v_{\|\cdot\|}(\bar{x}, F_A) \ln(2/v_{\|\cdot\|}(\bar{x}, F_A)) r_{\|\cdot\|}(F_A) \leq \varepsilon. \quad (12)$$

Нетрудно указать такое (удовлетворяющее условию нормировки) $A = \bar{A}$, что неравенство (12) при $\varepsilon < c_0$ (здесь и далее c_i — абсолютные положительные константы) для любой нормы $\|\cdot\|$ дает $v_{\|\cdot\|}(\bar{x}, F_{\bar{A}}) \leq c_1 \varepsilon \ln^{-1}(1/\varepsilon)$, а для $\|\cdot\|_2$ — даже $v_{\|\cdot\|_2}(\bar{x}, F_{\bar{A}}) \leq c_1 \varepsilon \ln^{-1}(\sqrt{n_1+n_2}/\varepsilon)/\sqrt{n_1+n_2}$. Вместе с тем, для любого нормированного A при $\varepsilon \leq c_0$ (12) выполнено при $v_{\|\cdot\|}(\bar{x}, F_A) \leq c_3 \varepsilon \ln^{-1}(1/\varepsilon)$. Таким образом, применение для решения нашей задачи лучшего (по теоретически определенной в п. 6 разд. 3 трудоемкости) из рассмотренных в разд. 3 — метода $\mathfrak{B}_1$ — требует, вообще говоря, не менее $c_4(n_1+n_2) \ln 1/\varepsilon$ вычислений $F_{\bar{A}}$; применение для той же цели метода $\mathfrak{B}_{n_1+n_2}^2$, на первый взгляд лучшего по оценкам (11) из семейства $\{\mathfrak{B}_{n_1+n_2}^p\}$, требует, в силу (11), не менее $c_4(n_1+n_2) \ln^2(\sqrt{n_1+n_2}/\varepsilon)/\varepsilon^2$ таких вычислений и, наконец, применение $\mathfrak{B}_{n_1+n_2}$ требует (опять-таки по (11)) не более $c_5 \ln(n_1+n_2) \ln^2(1/\varepsilon)/\varepsilon^2$ вычислений F_A . Сопоставление этих оценок показывает, что при фиксированном $\varepsilon < c_0$ и достаточно большом n_1+n_2 (между прочим, типичная для экономики ситуация) из наших трех методов лучшим является $\mathfrak{B}_{n_1+n_2}^1$, так что и наличие методов семейства $\{\mathfrak{B}_{n_1+n_2}^2\}$, и свобода в выборе p — обстоятельства, не бесполезные для приложений.

Стоит еще отметить, что методы $\mathfrak{B}_n^p$ с точки зрения реализации относительно просты (для каждого из них построение очередного приближения обходится в $O(n)$ арифметических операций, не считая не зависящих от структуры методов операций, требуемых для вычислений $F(x)$).

По нашему мнению, сказанное — достаточная мотивировка для приводимых далее построений.

3. Опишем теперь конструкцию рассматриваемых в этом разделе методов решения уравнений (2). Они естественным образом обобщают методы зеркального спуска решения выпуклых экстремальных задач и выпукло-вогнутых игр из [6] на случай уравнений (2) *.

Пусть $E, \|\cdot\|$ — пространство (не обязательно конечномерное), удовлетворяющее условиям разд. 2 и, кроме того, такое, что существует функция $V: E^* \rightarrow R$, имеющая на E^* равномерно-непрерывную и ограниченную производную и такая, что (a) $V(0) = 0$,

$$V(\varphi) \geq 0 \quad \forall \varphi; \quad (b) \frac{1}{2} + V(\varphi) \geq \|\varphi\| \cdot V\varphi, \quad (c) \frac{d}{dt} V(t\varphi) \geq 1$$

при $t \geq 1$ и $\|\varphi\| \geq 1$. По поводу запаса таких пространств и способов выбора соответствующих V см. [6, гл. III, § 2].

Сопоставим функции V с указанными свойствами метод $\mathfrak{B}_V$ решения уравнений (2) на подмножестве $G \subset E$ (разумеется, выпуклом, замкнутом

* Автор признателен Б. Т. Поляку, указавшему на возможность непосредственного перенесения результатов [6] на рассматриваемые в данной работе задачи.

и ограниченном). При описании метода будем считать, что G нормировано условиями $0 \in G$, $d_{\|\cdot\|}(G) = 1$ (этого всегда можно достичь с помощью простых преобразований подобия).

Пусть $\pi_G(x)$ при $x \in E$ — ближайшая к x точка G , $\mu_G(x)$ — опорный функционал к выпуклой функции $\rho(x, G) = \|x - \pi_G(x)\|$ в точке x и для $\psi \in E^*$

$$\pi_*(\psi) = \begin{cases} \psi, & \|\psi\|_* \leq 1, \\ \psi/\|\psi\|_*, & \|\psi\|_* > 1. \end{cases}$$

Пусть $\varphi(x)$ — сообщаемый в ответ на вопрос о значении F в x элемент $F(x)$.

Работа метода определяется заранее выбранной последовательностью смещений $\rho_i > 0$ и «функцией памяти» $j(i)$; $j(i)$ натурально и $j(i) \leq i$ для всех i . На i -м шаге работы метода строятся точки $x_i, \bar{x}_i, y_i, \psi_i, \varphi_i$ ($x_i \in E, \bar{x}_i, y_i \in G, \varphi_i, \psi_i \in E^*, \|\psi_i\|_* \leq 1$) в соответствии со следующими рекуррентными правилами; если $i=1$ или $\varphi_j \neq 0, 1 \leq j < i$, то $x_i = V'(\psi_{i-1})$; $\bar{x}_i = \pi_G(x_i)$; $\varphi_i = \varphi(\bar{x}_i)$, если $\varphi_i \neq 0$, то полагаем $\psi_i = \pi_*(\psi_{i-1} - \rho_i(\varphi_i/\|\varphi_i\|_* + \mu_G(x_i)))$, $\rho_i^* = \rho_i/$

$$\|\varphi_i\|_*, y_i = \left(\sum_{j(i)}^i \rho_s^* \bar{x}_s \right) \left(\sum_{j(i)}^i \rho_s^* \right)^{-1}. \text{ Если же } \varphi_i = 0, \text{ то } y_i = \bar{x}_i = x_i = x_i, \psi_i =$$

$$= \psi_{i-1}, \varphi_j = \varphi_i \text{ для всех } j \geq i.$$

В качестве начального условия в описанных правилах выбирается $\psi_0 = 0$.

4. Теорема 2. Пусть F ограничено и выполнены условия $\rho_i \rightarrow 0, \rho_i > 0$,

$$\sum_{j(i)}^i \rho_s \xrightarrow{i \rightarrow \infty} \infty. \text{ Тогда:}$$

$$1) v_{\|\cdot\|}(y_i, F) \xrightarrow{i \rightarrow \infty} 0,$$

2) справедлива оценка

$$v_{\|\cdot\|}(y_i, F) \leq v_i = \frac{2 \left(c_v + \sum_{j(i)}^i \rho_j \omega_v(2\rho_j) \right)}{\sum_{j(i)}^i \rho_j}, \quad (13)$$

где $c_v = \sup_{\|\varphi\|_* \leq 1} V(\varphi) + 1$ и $\omega_v(t) = \sup \{ \|V'(\varphi) - V'(\psi)\| \mid \|\varphi - \psi\|_* \leq t \}$ — модуль непрерывности $V'(\cdot)$.

Доказательство теоремы помещено в п. 10.

5. Обсудим сформулированный результат. Прежде всего, рассмотрим вопрос о рациональном выборе смещений ρ_i и «памяти» $j(i)$ в описанном методе. Ввиду оценки (13) он определяется поведением функции ω_v . Пусть известно, что $\omega_v(t) \leq a_v t^\sigma, 0 < \sigma \leq 1$. Тогда разумно взять $j(i) = [\theta i]$, $0 < \theta < 1$, и $\rho_i = \mu i^{-(1/(1+\sigma))}$, $\mu > 0$, что обеспечивает $\bar{v}_i \leq a(a_v, \sigma, \theta, \mu, c_v) i^{-\sigma/(1+\sigma)}$; иными словами, при любом $v, 0 < v < 1$, для выполнения неравенства $v_{\|\cdot\|}(y_i, F) \leq v$ достаточно брать $i \geq b(a_v, \sigma, \theta, \mu, c_v) (1/v)^{(1+\sigma)/\sigma}$. Выбором параметров θ и μ можно минимизировать множители $a(a_v, \sigma, \theta, \mu, c_v)$, $b(a_v, \sigma, \theta, \mu, c_v)$ в указанных оценках.

Пусть, в частности, $E, \|\cdot\|$ — пространство l_p^n , $1 \leq p < \infty$. Для $\varphi \in R^n$ положим

$$V_\tau(\varphi) = \begin{cases} 1/2 \|\varphi\|_\tau^2, & \|\varphi\|_\tau \leq 1 \\ \|\varphi\|_\tau^{-1/2}, & \|\varphi\|_\tau > 1 \end{cases},$$

а $V_{p,n}(\varphi) = V_{\tau(p,n)}(\varphi)$, где

$$\tau(p,n) = \begin{cases} p/(p-1), & p > 1, \\ 2 \ln n, & p = 1. \end{cases}$$

Тогда, как показано в [6, гл. III, § 2], $V_{p,n}(\varphi)$ удовлетворяет условиям (а)–(с) по отношению к пространству l_p^n , а $\omega_{V_{p,n}}$ явно оценивается. В принятых нами терминах отвечающее $V_{p,n}$ значение $\sigma = \sigma_p$ есть $\min\{1, 1/(p-1)\}$, а $a_{V_{p,n}}$ не зависит от n при $p > 1$ и $a_{V_{p,n}} \leq c_6 \ln n$ при $p = 1$. Обозначим через $\mathfrak{B}_n^p$ метод п. 3, ассоциированный с $V_{p,n}$ и имеющий смещения $\rho_i = \mu i^{-1/(1+\sigma_p)}$, $j(i) = [\theta i]$ с оптимально выбранными по предлагаемой схеме параметрами θ и μ . Из сказанного нетрудно вывести, что для методов $\mathfrak{B}_n^p$ справедливы приведенные в (11) оценки трудоемкости.

Отметим, что методы $\mathfrak{B}_n^p$ оказываются «оптимальными по трудоемкости на l_p^n -шарах большой размерности». Именно, из результатов [6, гл. IV] следует, что если G — единичный шар в l_p^n , $1 \leq p < \infty$, $v > 0$ ($v < 10^{-1}$) фиксировано и n достаточно велико ($n > n_p(v)$), то для любого метода $\mathfrak{B}$ решения уравнений (2) на G справедлива нижняя оценка трудоемкости

$$M_{G, \|\cdot\|_p, R^n}^{\mathfrak{B}}(v) \geq d_n(p) (1/v)^{\max(2,p)}, \quad (14)$$

где $d_n(p) > 0$ не зависит от n при $p > 1$ и $d_n(1) \geq d(1) \ln n$, $d(1) > 0$. Таким образом, в сделанных выше предположениях о G , n и v трудоемкость $\mathfrak{B}_n^p$ в принципе не может быть снижена более, чем в $a(p)$ раз.

6. В гильбертовом случае ($E, \|\cdot\|$ гильбертово, $E^*, \|\cdot\|$ канонически отождествляется с $E, \|\cdot\|$) в качестве V можно выбирать функцию

$$V_2(x) = \begin{cases} 1/2 \|x\|^2, & \|x\| \leq 1, \\ \|x\|^{-1/2}, & \|x\| > 1. \end{cases}$$

Отвечающий ей метод $\mathfrak{B}_{V_2}$ после несущественных упрощений приобретает вид: строятся последовательности $x_i, y_i \in G$ по следующим правилам ($x_1 = 0$): пусть $\varphi_i \in F(x_i)$. Если $\varphi_i = 0$, то $x_j = y_j = x_i$ при всех $j \geq i$. Если же $\varphi_i \neq 0$, $1 \leq s \leq i$, то $x_{i+1} = p_G(x_i - \rho_i^* \varphi_i)$, где $\rho_i^* = \rho_i / \|\varphi_i\|$, и

$$y_i = \left(\sum_{j(i)} \rho_j^* x_s \right) \left(\sum_{j(i)} \rho_j^* \right)^{-1}. \quad (15)$$

Метод (15) предложен в [4, 5], тогда как общая конструкция п. 3, насколько нам известно, нова. Рационально выбирать смещения в (15) в виде $\rho_i = c/\sqrt{i}$, а память — как $j(i) = [\theta i]$ с $c = \ln^{-1/2}(1/\theta)$ и θ , определяемым из уравнения $\ln \frac{1}{\theta} = \frac{1}{\sqrt{\theta}} - 1$ (т. е. $\theta \approx 0,0066$ и $c \approx 0,45$). При таком выборе ρ_i

и $j(i)$ для (15) будет

$$v_{\|\cdot\|}(y_i, F) \leq \frac{0,97}{\sqrt{i}} \quad \text{при } i \geq 10. \quad (16)$$

7. Конструкция методов из п. 3 пригодна и тогда, когда наблюдения значений F происходят при наличии помех, именно, если при наблюдении значения $F(x)$ сообщается реализация случайной величины $\xi(x, \omega)$ со

значениями в E^* , зависящей от x , как от параметра. Здесь ω — точка пространства Ω элементарных событий, снабженного вероятностной мерой Φ_ω . Предполагается, что E сепарабельно, Ω — полное сепарабельное метрическое пространство, Φ_ω — регулярная борелева мера. Функцию $\xi(x, \omega)$ предположим борелевой по (x, ω) и будем считать, что при i -м наблюдении значения $F(x)$, проводимом в точке x_i , методу сообщается реализация случайной величины $\xi(x_i, \omega_i)$, где «шумы источника информации» ω_i , $i=1, 2, \dots$, распределены по Φ_ω и независимы в совокупности.

Пусть выполнены условия несмещенности: $M_{\Phi_\omega} \xi(x, \omega) = \varphi(x) \in F(x)$, $x \in G$ и, кроме того условие $\sup_{x \in G} M_{\Phi_\omega} \|\xi(x, \omega)\|^r = (L_F)^r < \infty$, где $r, 2 \geq r > 1$ —

параметр, а M означает среднее по соответствующему распределению.

Опишем метод $\mathfrak{B}_V$ решения уравнения (2), использующий указанную «зашумленную» информацию о значениях F . Пусть, как и выше, V — функция на E^* , удовлетворяющая (а) — (с), $0 \in G$, $d_{\|\cdot\|}(G) = 1$. Описываемый метод строит точки $x_i \in E$, $\bar{x}_i, y_i \in G$, $\psi_i \in E^*$ по следующим рекуррентным правилам (обозначения те же, что и в п. 3, как и там, $\rho_i > 0$ и $j(i) \leq i$ — заранее выбранные смещения и память; $\omega^i = (\omega_1, \dots, \omega_i)$; $\psi_0 = 0$) $x_i \equiv x_i(\omega^{i-1}) = V'(\psi_{i-1}(\omega^{i-1}))$, $\bar{x}_i \equiv \bar{x}_i(\omega^{i-1}) = \pi_G(x_i)$; $\psi_i \equiv \psi_i(\omega^i) = \pi_*(\psi_{i-1}(\omega^{i-1}) - \rho_i \{\xi(\bar{x}_i, \omega_i) + \|\xi(\bar{x}_i, \omega_i)\|_* \mu_G(x_i)\})$, $y_i \equiv y_i(\omega^{i-1}) = \left(\sum_{j(i)}^i \rho_j \bar{x}_j \right) / \left(\sum_{j(i)}^i \rho_j \right)$, где ω^{i-1} —

начальный фрагмент набора ω^i , а ω_i — его последний элемент.

Допустим (это не ограничит общности, см. [6, гл. V]), что $\pi_G(x)$ и $\mu_G(x)$ борелевы по x . Тогда $\psi_i(\cdot)$, $x_i(\cdot)$, $\bar{x}_i(\cdot)$, $y_i(\cdot)$ — борелевы функции своих аргументов, так что такова же и (случайная) погрешность $\varepsilon_i(\omega^{i-1}) = \varepsilon(y_i(\omega^{i-1}), F)$ i -го результата y_i в качестве решения уравнения $F(x) \ni 0$, $x \in G$ и можно говорить о ее среднем (по распределению $\Phi_{\omega^i} = \underbrace{\Phi_\omega \times \dots \times \Phi_\omega}_{i \text{ раз}}$ набора ω^i) значении $\bar{\varepsilon}_i = M_{\Phi_{\omega^i}} \varepsilon(y_i(\omega^{i-1}), F)$. Следующая теорема устанавливает сходимость (в среднем) описанного метода.

Теорема 3. В сделанных предположениях относительно E , $\|\cdot\|$, V , F и $\xi(\cdot, \cdot)$ для описанного выше метода справедливо следующее утверждение:

пусть $\rho_i > 0$, $\rho_i \rightarrow 0$, $\sum_{i \rightarrow \infty}^i \rho_j \rightarrow \infty$. Тогда:

$$1) \quad \bar{\varepsilon}_i \rightarrow 0, \quad i \rightarrow \infty$$

$$2) \quad \bar{\varepsilon}_i \leq \frac{2c_V + 1/2 + \sum_{j(i)}^i 6L_F \rho_j \omega_{V,r} (6L_F \rho_j)}{\sum_{j(i)}^i \rho_j}. \quad (17)$$

Здесь $\omega_{V,r}(t) = \sup \int_0^t \tau \omega_V(t\tau) d\Phi_\tau$, где верхняя грань берется по распре-

делениям вероятностей Φ_τ на полуоси $\{\tau \geq 0\}$, таким, что $\int_0^\infty \tau^r d\Phi_\tau \leq 1$. О ме-

тоде доказательства теоремы будет сказано в п. 10.

8. Относительно выбора смещений в методе п. 7 можно сделать пример-но те же замечания, что и в п. 4. Пусть $\omega_v(t) \leq \min \{b_v, a_v t^{\sigma_v}\}$, $0 < \sigma_v \leq 1$. Тогда $\omega_{v,r}(t) \leq a(a_v, b_v, r) t^{\min(\sigma_v, r-1)}$. В этих условиях рационален выбор $j(i) = [\theta i]$, $0 < \theta < 1$, и $\rho_i = O(i^{-1/(1+\sigma_r)})$, где $\sigma_r = \min \{\sigma_v, r-1\}$.

Пусть априорно известно не только r , но и такое L , что $L_F \leq L$, F — опе-ратор из решаемого уравнения (2). Тогда следует брать $\rho_i = \frac{c}{L} i^{-1/(1+\sigma_r)}$;

при этом правая часть (17) будет убывать с ростом i , как $O\left(L i^{-\min\{\frac{r-1}{r}, \frac{\sigma_v}{1+\sigma_v}\}}\right)$. В частности, если $r \leq 1 + \sigma_v$ (например, $\sigma_v = 1$), то правая часть (17) при указанном выборе ρ_i и $j(i)$ убывает, как $LQ(a_v, b_v, c_v, \sigma_v, r, \theta, c) i^{-(r-1)/r}$; при этом для обеспечения средней погрешности νL требуется не более $R(a_v, b_v, c_v, \sigma_v, r, \theta, c) (1/\nu)^{r/(r-1)}$ шагов. Например, $E, \|\cdot\| = l_p^n$, $1 < p \leq 2$, и мы заранее знаем $L \geq L_F$. В качестве V возьмем $\tilde{V}_{p,n}$, что дает $\sigma_v = 1$ и $\sigma_r = r-1$; при этом a_v, b_v, c_v будет зависеть лишь от p , но не от n . При вы-боре $j(i) = [\theta i]$ и $\rho_i = \frac{c}{L} i^{-1/r} = \frac{c}{L} i^{-1/(1+\sigma_r)}$ (получившийся метод обозна-

чим $\mathfrak{B}_n^{p,r}$) мы имеем решение уравнения (2) с $L_F \leq L$ средней погрешности νL не позднее, чем после $a(p, r) (1/\nu)^{r/(r-1)}$ шагов. При этом методы $\mathfrak{B}_n^{p,r}$ оказываются существенно неупрощаемыми. Именно, нетрудно показать (см. [6, гл. V, § 3]), что для некоторого $d(r) > 0$ при всех $\nu > 0$ для всех $G \subset l_p^n$, $1 < p \leq 2$, справедливо следующее утверждение. Никакой метод $\mathfrak{B}$ не в состоянии решить все уравнения (2) с $L_F \leq L$ ($L > 0$ фиксировано) и средней (абсолютной) погрешностью $\leq \nu L$, затратив на это число шагов, меньшее $d(r) (1/\nu)^{r/(r-1)}$. Таким образом, при $1 < p \leq 2$ методы $\mathfrak{B}_n^{p,r}$ улучша-ются по трудоемкости не более, чем в $q(p, r)$ раз, независимо от размер-ности и геометрии G .

9. В случае гильбертова $E, \|\cdot\|$ метод $\mathfrak{B}_V$, отвечающий $V = V_2$ (см. п. 6), можно упростить, приведя к виду ($x_i = 0$)

$$x_{i+1} = \pi_G(x_i - \rho_i \xi(x_i, \omega_i)), \quad (18)$$

$$y_i = \sum_{j(i)}^i \rho_j x_j / \left(\sum_{j(i)}^i \rho_j \right).$$

Для этого метода оценка (17) превращается в

$$\bar{\varepsilon}_i \leq \frac{1 + 1,5 \sum_{j(i)}^i L_F^2 \rho_j^2}{\sum_{j(i)}^i \rho_j}. \quad (19)$$

Рационален выбор $\rho_i = \frac{c}{L_F \sqrt{i}}$, $j(i) = [\theta i]$ с тем же значением θ , что и

в п. 6, и $c=1/\sqrt{1,5 \ln(1/\theta)}$ ($c \approx 0,36$). При этом правая часть (19) при $i \geq 10$ не больше $2,75L_F/\sqrt{i}$.

10. Доказательство теоремы 2. Так как $V'(\varphi)$ равномерно непрерывна, то $\omega_V(t) \xrightarrow{t \rightarrow 0} 0$. Поэтому 1) следует из 2), так что доказывать следует лишь 2) Если при данном i и некотором $j \leq i$ было $\varphi_j = 0$, то y_j — точное решение (2), а (13) очевидно. Пусть теперь при данном i и всех $j \leq i$ было $\varphi_j \neq 0$. Положим $\Delta_j = \varphi_j / \|\varphi_j\|_* + \mu_G(x_j)$, $1 \leq j \leq i$, и заметим, что $\|\Delta_j\|_* \leq 2$ (ибо $\|\mu_G(x_j)\|_* \leq 1$ ввиду липшицевости с константой 1 функции $\rho(x, G)$). Пусть $x \in G$ и $V_x(\psi) = V(\psi) - (\psi, x)$. Тогда в силу условия нормировки на G имеем $\|x\| \leq 1$; в свою очередь, ввиду (с) последнее дает $V_x(\varphi) \geq V_x(\pi_*(\varphi))$ для всех $\varphi \in E^*$. В частности, $V_x(\psi_j) = V_x(\pi_*(\psi_{j-1} - \rho_j \Delta_j)) \leq V_x(\psi_{j-1}) - (\Delta_j, V'_x(\psi_{j-1})) \rho_j + 2\rho_j \omega_V(2\rho_j)$ (формула конечных приращений). Итак, ввиду $V'_x(\psi_{j-1}) = x_{j-1} - x$

$$V_x(\psi_j) \leq V_x(\psi_{j-1}) + \rho_j(\Delta_j, x - x_j) + 2\rho_j \omega_V(2\rho_j). \quad (20)$$

Полагая $\xi_j = \varphi_j / \|\varphi_j\|_* = \Delta_j - \mu_G(x_j)$, получим $(\xi_j, x - \bar{x}_j) \geq (\Delta_j, x - x_j)$. В самом деле, $(\Delta_j, x - x_j) - (\xi_j, x - \bar{x}_j) = (\mu_G(x_j), x - x_j) + (\xi_j, \bar{x}_j - x_j) \leq -\rho(x_j, G) + \|\xi_j\|_* \|\bar{x}_j - x_j\| = 0$. Итак,

$$(\xi_j, x - \bar{x}_j) \geq (\Delta_j, x - x_j). \quad (21)$$

Подставляя (21) в (20), найдем

$$V_x(\psi_j) \leq V_x(\psi_{j-1}) + \frac{\rho_j}{\|\varphi_j\|_*} (\varphi_j, x - \bar{x}_j) + 2\rho_j \omega_V(2\rho_j). \quad (22)$$

Пусть $\xi \in F(x)$. В силу (1) $(\xi, x - \bar{x}_j) \geq (\varphi_j, x - \bar{x}_j)$, т. е. (22) дает

$$\rho_j^* (\xi, x - \bar{x}_j) \geq -V_x(\psi_{j-1}) + V_x(\psi_j) - 2\rho_j \omega_V(2\rho_j). \quad (23)$$

Просуммировав (23) по $j = j(i), \dots, i$, найдем

$$(\xi, x - y_i) \leq \left\{ \sum_{j(i)}^i \rho_j^* \right\}^{-1} \left(V_x(\psi_i) - V_x(\psi_{j(i)-1}) - 2 \sum_{j(i)}^i \rho_j \omega_V(2\rho_j) \right). \quad (24)$$

Заметим, что при $\|\varphi\|_* \leq 1$ и $\|x\| \leq 1$ имеем $|V_x(\varphi)| \leq c_V$ по определению c_V . Кроме того, $\|\varphi_j\|_* \leq r_{\|\cdot\|}(F)$ ввиду $d_{\|\cdot\|}(G) = 1$, так что

$$\left(\sum_{j(i)}^i \rho_j^* \right)^{-1} \leq r_{\|\cdot\|}(F) \left(\sum_{j(i)}^i \rho_j \right)^{-1}. \text{ Итак, (24) дает}$$

$$(\xi, x - y_i) \leq -2r_{\|\cdot\|}(F) \frac{\sum_{j(i)}^i \rho_j \omega_V(2\rho_j)}{\sum_{j(i)}^i \rho_j}. \quad (25)$$

Так как (25) верно при всех $x \in G$ и $\xi \in F(x)$, то

$$\varepsilon(y_i, F) \leq \frac{2 \left(c_v + \sum_{j(i)}^i \rho_j \omega_v(2\rho_j) \right)}{\sum_{j(i)}^i \rho_j} r_{\Pi\Pi}(F),$$

что и требуется в (13). Теорема доказана.

Теорема 3 доказывается по этой же схеме, наличие помех учитывается в точности так же, как и в доказательстве теоремы 3.1 из [6, гл. VI, § 3].

ЛИТЕРАТУРА

1. Вайнберг М. М. Вариационный метод и метод монотонных операторов в теории нелинейных уравнений. М.: Наука, 1972.
2. Бакушинский А. Б., Поляк Б. Т. О решении вариационных неравенств.— Докл. АН СССР, 1974, т. 219, вып. 5.
3. Гольштейн Е. Г. Метод модификации монотонных отображений.— Экономика и матем. методы, 1975, т. XI, вып. 6.
4. Bruck R. On Weak Convergence of an Ergodic Iterations for the Solution of Variational Inequalities for Monotone Operators In Hilbert Space.— J. Math. Anal. Appl., 1977, v. 61, № 1.
5. Lions P. L. Une méthode iterative de resolution d'une inéquation variationnelle.— I. J. Math., 1978, v. 31, № 2.
6. Немировский А. С., Юдин Д. Б. Сложность задач и эффективность метода оптимизации. М.: Наука, 1979.
7. Бергс К. Общая теория игр нескольких лиц. М.: Физматгиз, 1961.
8. Левин А. Ю. Об одном алгоритме минимизации выпуклых функций.— Докл. АН СССР, 1965, т. 160, вып. 6.
9. Юдин Д. Б., Немировский А. С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач.— Экономика и матем. методы, 1976, т. XII, вып. 1.
10. Шор Н. З. Новые направления в развитии методов негладкой оптимизации.— Кибернетика, 1977, № 6.
11. Юдин Д. Б., Немировский А. С. Эффективность случайного поиска в задачах управления.— Изв. АН СССР. Техн. кибернетика, 1977, № 3.

Поступила в редакцию
16 X 1979