ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2011, том 51, № 5, с. 1–11

УДК

ФЕЙЕРОВСКИЕ АЛГОРИТМЫ С АДАПТИВНЫМ ШАГОМ[1)]

© 2011 г. Е. А. Нурминский

(690041 Владивосток, ул. Радио, 5, Ин-т автоматики и процессов управления ДВО РАН)

e-mail: nurmi@dvo.ru

Поступила в редакцию 15.06.2010 г.
Переработанный вариант 14.09.2010 г.

Для фейеровских процессов с аттрактантами предложена общая адаптивная схема управления шаговыми множителями и доказана сходимость подобного класса алгоритмов к стационарным точкам. Приведены результаты вычислительных экспериментов, демонстрирующих в основном линейную скорость сходимости. Библ. 12. Фиг. 3.

Ключевые слова: фейеровские процессы; выпуклая оптимизация; вариационные неравенства, декомпозиция; регулировка шага.

ВВЕДЕНИЕ

В работах [1], [2] в качестве моделей декомпозиционных алгоритмов решения оптимизационных и других задач были рассмотрены фейеровские процессы с убывающими возмущениями. Такой подход представляет большую свободу в конструировании алгоритмов распределенных и параллельных вычислений, позволяет обойти многие трудности, связанные с учетом сложных ограничений. Однако в этих работах рассматривались возмущения с контролируемыми шаговыми множителями, которые удовлетворяли классическим условиям малости и суммарной расходимости. На практике известно, что такие условия приводят к весьма медленной сходимости и задача ускорения подобных методов имеет большое теоретическое и практическое значение. В связи с этим в [3] был предложен адаптивный способ управления шаговыми множителями фейеровских процессов и доказана его сходимость для простейшего случая. Вычислительные эксперименты продемонстрировали существенное ускорение сходимости. В настоящей работе приведено доказательство сходимости подобных алгоритмов для общего случая сильно фейеровских процессов с аттрактантными возмущениями, которые могут порождаться как градиентами целевых функций оптимизационных задач, так и операторами вариационных неравенств.

1. ОСНОВНЫЕ ПОНЯТИЯ И ОБОЗНАЧЕНИЯ

Пусть E обозначает основное конечномерное евклидово пространство векторов со скалярным произведением xy и соответствующей нормой $\|x\| = \sqrt{xx}$. Размерность этого пространства будем обозначать через $\dim(E)$.

Сумму $A + B$, где $A, B \subset E$, будем понимать как $A + B = \{a + b : a \in A, b \in B\}$. Для упрощения обозначений одноэлементное множество $\{a\}$ будет обозначаться просто как a там, где это не приводит к недоразумениям.

Единичный шар $\{x : \|x\| \le 1\}$ будем обозначать через B и под окрестностью нуля будем понимать произвольное открытое подмножество E, содержащее начало координат. Конечно, в качестве таких окрестностей могут быть использованы открытые шары $B_\varepsilon = \{x : \|x\| < \varepsilon\}$, $\varepsilon > 0$. Норму множества A будем определять как $\|A\| = \sup_{a\in A} \|a\|$.

1) Работа выполнена при финансовой поддержке РФФИ (код проекта 09-01-00042-а) и проекта ДВО РАН 09-III-А-01-004.

Выпуклую оболочку конечного семейства векторов $\{a^i : i \in \mathcal{N}\} \subset E$, где $\mathcal{N} = \{1, 2, \dots, N\}$, определим и обозначим традиционным образом:

$$\mathrm{co}\{a^i : i \in \mathcal{N}\} = \left\{a = \sum_{i \in \mathcal{N}} \lambda_i a^i : \sum_{i \in \mathcal{N}} \lambda_i = 1, \lambda_i \geq 0, i \in \mathcal{N}\right\} = \{a = A\lambda : \lambda \in \Delta\},$$

где $\Delta = \left\{\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) : \sum_{i=1}^{N} \lambda_i = 1, \lambda_i \geq 0, i \in \mathcal{N}\right\}$ — стандартный симплекс, а A — $\dim(E) \times N$-матрица, столбцами которой являются векторы a^i, $i \in \mathcal{N}$.

Множество всех подмножеств пространств E будем обозначать как 2^E. Для точечно-множественных отображений $G : E \to 2^E$ будет использоваться понятие полунепрерывности сверху в точке x как существование для любой окрестности нуля U такой окрестности нуля V, что $G(y) \subset G(x) + U$ при $y \in x + V$. Точечно-множественное отображение G полунепрерывно сверху, если оно полунепрерывно сверху в любой точке $x \in E$.

Для выпуклого множества X нормальным конусом $N_X^+(x)$ в точке x будем называть множество $N_X^+(x) = \{p : p(y - x) \geq 0, y \in X\}$. Заметим, что $N_X^+(x) = \{0\}$ для $x \in \mathrm{int}(X)$.

Для краткости опорную функцию множества A будем обозначать через $(A)_p$. Операция $(\cdot)_p$ очевидным образом положительно-линейна: $(\lambda A)_p = \lambda (A)_p$ для $\lambda \geq 0$ и $(A + B)_p = (A)_p + (B)_p$.

2. ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ

Основным объектом исследования этой работы являются фейеровские отображения.

Определение 1. Отображение $F : E \to E$ будем называть фейеровским по отношению к некоторому заданному непустому замкнутому множеству $V \subset E$, если выполняется неравенство

$$\|F(x) - v\| \leq \|x - v\| \tag{1}$$

для любых $v \in V$.

Очевидно, что при этом $F(v) = v$ для всех $v \in V$. Что понимается под множеством V обычно ясно из контекста определения F и обычно мы будем опускать зависимость F от V.

Фейеровские отображения позволяют определить фейеровские процессы

$$x^{k+1} = F(x^k), \quad k = 0, 1, \dots, \tag{2}$$

часто используемые (см., например, [4]) как модели итеративных алгоритмов для решения различных задач. При этом в качестве V выступает искомое множество решений этих задач и основной вопрос обоснования применимости процессов (2) заключается в доказательстве сходимости этого процесса к множеству V.

Как правило, для обоснования сходимости процесса (2) необходимо более сильное свойство, чем просто фейеровость отображения F, и в настоящей работе, как и в [1], мы будем использовать следующее

Определение 2. Фейеровское отображение F будем называть локально сильным, если для любого $\bar{x} \notin V$ существует окрестность нуля U и число $\alpha \in [0, 1)$ такие, что $\|F(x) - v\| \leq \alpha \|x - v\|$ для всех $v \in V$ и $x \in \bar{x} + U$.

Как и выше, в очевидных случаях будем опускать зависимость F от V.

2.1. Сходимость фейеровских процессов с возмущениями

Для локально сильных фейеровских отображений оказывается возможным показать устойчивость процессов (2) по отношению к убывающим аддитивным возмущениям (см. [1]), процесс вида

$$x^{k+1} = F(x^k + z^k), \quad k = 0, 1, \dots, \tag{3}$$

где F — локально-сильное фейеровское отображение, z^k — произвольное убывающее ($z^k \to 0$) возмущение, сходится к множеству V. Сходимость к множеству V понимается в том смысле, что каждая предельная точка последовательности $\{x^k\}$ принадлежит множеству V. Этот тип результатов может быть использован, например, для решения выпуклых задач допустимости, дополняя известные алгоритмы, обзор которых, например, приведен в [5].

Однако, используя специальные возмущающие добавки z^k, от процесса (3) можно добиться дополнительного положительного эффекта сходимости к выделенным подмножествам V. Для описания таких специальных возмущений введем следующее

Определение 3. Точечно-множественное отображение $G : E \to 2^E$ назовем *локальным аттрактантом* $Z \subset V$, если $g(z - x) \geq 0$ для всех $x \in V \setminus Z$, $g \in G(x)$ и $z \in Z$.

Для гарантий уточненных результатов по сходимости процессов вида (3), вообще говоря, необходимо усилить это определение.

Определение 4. Локальный аттрактант G будем называть *сильным локальным аттрактантом*, если для каждого $x' \in V \setminus Z$ существует окрестность нуля U и число $\delta > 0$ такие, что

$$g(z - x) \geq \delta > 0$$

для всех $z \in Z$, $x \in x' + U$, $g \in G(x)$.

Далее будем предполагать, не оговаривая это особо, локальную равномерную ограниченность G: для каждого x пусть существует окрестность нуля U и константа $C < \infty$ такие, что $\|G(y)\| \leq C$ для $y \in x + U$, В этом случае для сильных локальных аттрактантов результат о сходимости процессов вида (3) может быть усилен.

Теорема 1 (см. [1]). *Пусть F — локально сильное фейеровское отображение относительно множества V, $G : E \to 2^E$ — сильный локальный аттрактант множества $Z \subset V$ и последовательность $\{x^s\}$, определяемая рекуррентным соотношением*

$$x^{s+1} = F(x^s + \lambda_s g^s), \quad g^s \in G(x^s), \tag{4}$$

где $\lambda_k \to +0$ и $\sum \lambda_k = \infty$, ограничена. Тогда любая предельная точка этой последовательности принадлежит множеству Z.

Более полезная для развития декомпозиционных алгоритмов версия этой теоремы предоставляет возможность использовать набор фейеровских операторов $\mathcal{F} = \{\phi_i, i \in \mathcal{M} = \{1, 2, \ldots, \mathcal{M}\}\}$, каждый из которых является локально-сильным относительно своего множества V_i, $i \in \mathcal{M}$, пересечение которых дает допустимое множество V.

Справедлива следующая

Теорема 2 (см. [1]). *Пусть $D(\cdot)$ — сильный локальный аттрактант $Z \subset V$ и $\mathcal{F} = \{\phi_i, i \in \mathcal{M}\}$ — конечное семейство непрерывных локально сильных относительно соответствующих V_i, $i \in \mathcal{M}$, фейеровских операторов. Пусть для любого $x \notin V = \cap_{i \in \mathcal{M}} V_i$ существует $k(x) \in \{\mathcal{M}\}$ такой, что $\phi_{k(x)}$ локально сильный в x. Тогда если последовательность $\{x^s\}$, построенная по правилу*

$$\begin{gathered} x^{s+1} = F_s(x^s + \lambda_s d^s), \quad d^s \in D(x^s), \quad F_s = \phi_{k_s}, \\ \text{где индекс} \quad k_s \in \mathcal{M} \quad \text{такой, что} \quad x^s + \lambda_s d^s \notin V_{k_s}, \end{gathered} \tag{5}$$

ограничена, то она сходится к множеству Z при $\lambda_k \to +0$ и $\sum \lambda_k = \infty$.

Эта теорема обосновывает методы типа последовательных проекций градиентов вида

$$x^{k+1} = F_k(x^k - \lambda_k g^k), \quad k = 0, 1, \ldots, \tag{6}$$

для решения экстремальных задач

$$\min_{x \in X} f(x) \quad \text{с} \quad X = \bigcap_{i=1}^{\mathcal{M}} V_i,$$

где фейеровские операторы F_k являются проекциями на соответствующие множества V_i, относительно которых точка $x^k - \lambda_k g^k$ является недопустимой. Во многих случаях множества V_i можно выбрать так, что операция проекции на них достаточно легко реализуема.

Алгоритмы вида (6), но с $g^k = 0$, хорошо известны для решения задачи выпуклой допустимости, т.е. нахождения какого-либо $x \in X$ (см., например, [5], [9]), однако для решения задачи оптимизации их применение до сих пор не было обосновано в такой постановке. Некоторый аналог алгоритма (6) может быть усмотрен в алгоритмах линеаризации ограничений, когда допустимое множество X, заданное в виде выпуклого неравенства $X = \{x : h(x) \le 0\}$, представляется как пересечение в принципе континуального набора полупространств $H_y = \{x : h(y) + g(x - y) \le 0$, $g \in \partial h(y)$, $y \in E\}$ (для решения задач вариационных неравенств см., например, [8]). Работа с такой бесконечной системой линейных ограничений требует, однако, последующего конечного или счетного представительного отбора полупространств, их агрегации и прочих операций, по-видимому отрицательно влияющих на сходимость. Возможно, именно по этой причине они не получили широкого распространения, хотя детальное сравнение с алгоритмами (6) стоит, по-видимому, еще произвести. В ограниченном виде идея раздельной обработки ограничений также использовалась в задаче расщепленной допустимости (split feasibility) (см. [10]). Вычислительную эффективность алгоритма (6) понижают весьма абстрактные и не связанные с конкретной задачей условия на шаговые множители λ_k, использованные также и в [8].

Основной задачей данной работы является уточнение адаптивной схемы регулировки шаговых множителей λ_k в алгоритмах типа (6), первоначально предложенной в [3]. Эта схема продемонстрировала в предварительных вычислительных экспериментах вполне удовлетворительную (линейную) скорость сходимости, однако, строго говоря, была обоснована только для методов градиентного типа в задачах без ограничений.

2.2. Необходимые условия стационарности

Для описания множеств предельных точек сходящихся фейеровских процессов с аттрактантами мы будем применять условия, аналогичные необходимым условиям оптимальности в форме вариационных неравенств. Действительно, поскольку фейеровские операторы имеют много общего с проективными операторами, они особенно удобны для решения проективных уравнений вида

$$x = \Pi_X(x - \lambda g), \quad g \in G(x), \tag{7}$$

где проективный оператор $\Pi_X(\cdot)$ определяется традиционным образом

$$\|x - \Pi_X(x)\| = \min_{z \in X} \|x - z\|, \quad \Pi_X(x) \in X. \tag{8}$$

К решению уравнения (7) могут быть сведены задачи условной оптимизации с допустимым множеством X и вариационные неравенства

$$g(x^* - x) \ge 0, \quad g \in G(x^*) \quad \text{для всех} \quad x \in X, \tag{9}$$

что покрывает значительную часть проблем вычислительной математики.

Для связи проективных уравнений (7) с развиваемой теорией фейеровских операторов с аттрактантами необходимо сопоставить их условия стационарности и определяющие соотношения соответствующих задач вычислительной математики. Для выпуклых условных оптимизационных задач вида

$$\min_{x \in X} f(x), \tag{10}$$

где X — выпуклое замкнутое множество, f — выпуклая собственная замкнутая функция, конечная на некотором открытом множестве $X' \supset X$, необходимые условия экстремума достаточно хорошо изучены (см., например [6, 142]). В терминах нормальных конусов $N_X^+(x)$ множества X справедливо следующее утверждение:

Теорема 3. *Пусть $N_X^+(x)$ — нормальный конус в точке x выпуклого замкнутого множества X, $\partial f(x)$ — субдифференциал функции f в точке x. Точка x^* является решением задачи* (10) *тогда и только тогда*, *когда* $0 \in \partial f(x^*) + N_X^+(x^*)$.

Для дальнейшего потребуется также свойства непрерывности нормальных конусов и субдифференциалов. Опять же, используя традиционный аппарат выпуклого анализа, можно показать, что многозначное отображение $\partial f(x) + N^+(x)$ является полунепрерывным сверху.

Для исследования сходимости с нетривиальными операторами F потребуются некоторые общие условия сходимости, которые кратко будут приведены ниже.

2.3. Условия сходимости

В дальнейшем будут использованы общие условия сходимости итеративных процессов (см. [7]), являющиеся в определенном смысле дискретной версией условий асимптотической устойчивости Ляпунова. В соответствие с этими условиями, для того чтобы у последовательности $\{x^s\}$, порожденной некоторым итеративным процессом, существовала предельная точка, принадлежащая заданному множеству X_* достаточно выполнения следующих условий

А1. Последовательность $\{x^s\}$ ограничена.

А2. Для каждой подпоследовательности $\{x^{n_k}\}$, сходящейся к $x' \notin X_*$, существует $\epsilon > 0$ такое, что для каждого n_k найдется индекс $m_k > n_k$ такой, что $\left\|x^{n_k} - x^s\right\| \le \epsilon$ для $n_k \le s < m_k$ и $\left\|x^{n_k} - x^{m_k}\right\| > \epsilon$.

А3. Существует непрерывная функция $W(x)$ такая что для всякой подпоследовательности $\{x^{n_k}\}$, сходящейся к $x' \notin X_*$ и существующей в силу условия **А2** соответствующей подпоследовательности $\{x^{m_k}\}$ найдется подпоследовательность $\{p_k\}$ с $n_k < p_k \le m_k$ такая, что выполняется неравенство

$$\limsup_{k\to\infty} W(x^{p_k}) < \liminf_{k\to\infty} W(x^{m_k}).$$

Для того чтобы каждая предельная точка последовательности $\{x^s\}$ принадлежала X_*, достаточно выполнения двух дополнительный условий

А4. Если $x^{n_k} \to x^* \in X_*$, то $\left\|x^{n_k+1} - x^{n_k}\right\| \to 0$.

А5. Множество $W(X_*) = \{W(x^*) : x^* \in X_*\}$ таково, что $R \backslash W(X_*)$ всюду плотно.

Содержательно условие **А2** препятствует “застреванию” итеративного процесса $\{x^s\}$ в точках, не принадлежащих X_*. Условие **А3** неявным образом запрещает предельные циклы, не проходящие через точки X_*, и является аналогом отрицательности полной производной функции Ляпунова по траектории динамического процесса в теории устойчивости систем, описываемых обыкновенными дифференциальными уравнениями. Условие **А4** предотвращает “отпрыгивание” точек последовательности $\{x^s\}$ от предельного множества X_*, условие **А5** в совокупности с **А1**–**А4** фактически гарантирует сходимость $\{W(x^s)\}$, что в сочетании с **А3** исключает наличие у $\{x^s\}$ предельных точек, не принадлежащих X_*.

3. МЕТОД АДАПТИВНОЙ СТУПЕНЧАТОЙ РЕГУЛИРОВКИ ШАГА

Идея адаптивного управления шагом в фейеровских алгоритмах с аттрактантами, первоначально предложенная в [3], заключается в следующем. Рассмотрим итеративный процесс

$$x^{k+1} = F(x^k + \lambda_k g^k), \quad g^k \in G(x^k), \tag{11}$$

где $G(x)$ – это точечно-множественный аттрактант, требования к которому будут уточнены далее. Мы будем исследовать сходимость процесса (11) к множеству X_*, определяемому условиями стационарности

$$X_* = \{x^* : 0 \in G(x^*) + N_X^+(x^*)\}, \tag{12}$$

где X – множество стационарных точек фейеровского отображения F.

Процесс (11) может быть представлен в виде

$$x^{k+1} = x^k + \lambda_k d^k, \quad d^k = [(F(x^k + \lambda_k g^k) - x^k]/\lambda_k, \quad g^k \in G(x^k), \tag{13}$$

и d^k можно считать принадлежащими некоторым отображениям D_k. Эквивалентно $d^k = (x^{k+1} - x^k)/\lambda_k$.

Для упрощения обозначений положим $D(p,q) = \mathrm{co}\{d^p, d^{p+1}, \ldots, d^q\}$.

Для заданной последовательности $\theta_m \to +0$, $m = 0,1,\ldots$, определим соответствующую последовательность индексов $\{k_m\}$ и чисел $\{\lambda_k\}$:

положим $m = 0$, $k_0 = 0$ и выберем начальный шаг $\lambda_0 > 0$, пусть $q \in (0,1)$;

для данного m и k_m определим k_{m+1} как индекс, который удовлетворяет условиям

$$0 \notin D(k_m, k) + \theta_m B, \quad k_m \le k < k_{m+1}, \quad 0 \in D(k_m, k_{m+1}) + \theta_m B, \tag{14}$$

с $\lambda_k = \lambda_{k_m}$ для $k_m \le k < k_{m+1}$; положим

$$\lambda_{k_{m+1}} = q\lambda_{k_m}. \tag{15}$$

Далее увеличим $m : m \to m + 1$ и продолжим итерации алгоритма (11).

Содержательно условие (14) состоит в том, чтобы обнаружить тот момент, когда некоторый отрезок последовательности $\{x^k\}$ начинает удовлетворять суррогату условия стационарности (14). При уменьшении шага λ_k к нулю, что гарантируется фактором $q < 1$, этот отрезок последовательности стягивается в сколь угодно малую окрестность, в которой удовлетворяется условие стационарности (12).

В [3] предлагаемая схема регулировки шага обосновывалась для простейшего случая, когда F_k являлись тождественными операторами и по сути дела, $d^k = g^k$, $D_k(x) = G(x)$, хотя вычислительные эксперименты демонстрировали применимость этой регулировки шага и для общей схемы (11). Ниже мы приведем обоснование регулировки (14) для общего случая.

Для доказательства необходимо несколько вспомогательных утверждений.

Лемма 1. *Если последовательность* $\{x^k\}$, *генерированная* (11) *с регулировкой шага* (14) *ограничена, то все ее предельные точки принадлежат равенству* X.

Доказательство. Легко видеть, что из ограниченности $\{x^k\}$ следует, что последовательность индексов $\{k_m\}$, определяемая (14), бесконечна. Действительно, в противном случае для некоторого $\bar{m}$ выполнялось бы $0 \notin D(k_{\bar{m}}, k) + \theta_{\bar{m}} B$ для всех $k > k_{\bar{m}}$ и, следовательно, в силу монотонности по включению $D(k_{\bar{m}}, k)$ по k,

$$\bar{0} \notin \mathrm{ri}\left[\lim_{k\to\infty} D(k_{\bar{m}}, k)\right] + \theta_{\bar{m}} B = \mathrm{ri}(\bar{D}) + \theta_{\bar{m}} B,$$

где предел множеств понимают как $\bar{D} = \bigcap_{k > k_{\bar{m}}} D(k_{\bar{m}}, k)$. Соответственно, существует p с $\|p\| = 1$ такое, что

$$(\bar{D} + \theta_{\bar{m}} B)_p \le 0,$$

или

$$(\bar{D})_p \le -\theta_{\bar{m}} < 0.$$

При этом $\lambda_k = \lambda_{\bar{m}} > 0$ для всех $k > k_{\bar{m}}$ и соответственно

$$-\left\|x^k - x^{k_{\bar{m}}}\right\| \le (x^k - x^{k_{\bar{m}}})p = \sum_{s=k_{\bar{m}}}^{k-1} (x^{s+1} - x^s)p =$$

$$= \sum_{s=k_{\bar{m}}}^{k-1} \frac{d^s}{\lambda_s} p \le \sum_{s=k_{\bar{m}}}^{k-1} \frac{1}{\lambda_s} (\bar{D})_p \le -\frac{\theta_{\bar{m}}}{\lambda_{\bar{m}}} (k - 1 - k_{\bar{m}}) \to -\infty$$

при $k \to \infty$, откуда $\left\|x^k\right\| \to \infty$ при $k \to \infty$, что противоречит ограниченности $\{x^k\}$.

В свою очередь, из неограниченности последовательности $\{k_m\}$ следует, что $\lambda_k \to 0$ при $k \to \infty$ и, следовательно, в силу теоремы 3,

$$\lim_{s\to\infty} x^{k_s} = x' \in V$$

для любой сходящейся подпоследовательности $\{x^{k_s}\}$.

Пусть теперь $\lim_{k\to\infty} x^{n_k} = x'$. Тогда, в силу леммы 1, $x' \in X$. Дальнейшее уточнение поведения последовательности $\{x^k\}$ в окрестности точки x' дается следующей леммой.

Лемма 2. *Если* $x^{n_k} \to x'$ *и* $0 \notin G(x') + N_X^+(x')$, *тогда существует* $\varepsilon > 0$ *такое, что* $\|x^{n_k} - x^{m_k}\| > \varepsilon$ *для некоторого* $m_k > n_k$ *при всех* k.

Доказательство. Для доказательства леммы предположим противное, то есть, что существует подпоследовательность $\{x^{n_k}\} \to x' \notin X_\star$ такая, что для любого $\varepsilon > 0$ существует k такое, что $\|x^s - x^{n_k}\| \le \varepsilon$ для всех $s > n_k$. Вообще говоря, это означает, что вся последовательность $\{x^s\}$ сходится к x'.

Из замкнутости и отделимости от 0 множества $G(x') + N_X^+(x')$ следует, что существуют p и δ такие, что $p \ne 0$, $\|p\| = 1$ и $py \ge \delta$ для всех $y \in G(x') + N_X^+(x') + U$, где U — некоторая окрестность нуля. Тогда при $s > n_k$ и достаточно больших k с $\bar{x}^s = x^s - \lambda_s g^s$ имеем

$$\frac{1}{\lambda_s}(\bar{x}^s - x^s) = g^s \in G(x') + U$$

в силу полунепрерывности сверху G и $\bar{x}^s - F_s(\bar{x}^s) \in N_X^+(x') + U$ в силу фейеровости F.

Поэтому

$$p(x^{s+1} - x^s) = \lambda_s p\left(\frac{F(\bar{x}^s) - \bar{x}^s}{\lambda_s} + \frac{\bar{x}^s - x^s}{\lambda_s}\right) = \lambda_s p w^s,$$

где $w^s \in G(x') + N_X^+(x') + U$ и, следовательно

$$p(x^{s+1} - x^s) \ge \lambda_s \delta > 0, \quad s \ge n_k.$$

Не умаляя общности, можно считать, что $\lambda_s \ge q\lambda_{n_k}$ для $s > n_k$, поскольку в этом случае для $s > n_k$ дробление шага могло произойти не более одного раза. Суммируя последние неравенства по s от n_k до $m - 1$, получаем

$$p(p^m - x^{n_k}) \ge q\lambda_{n_k}(m - n_k) \to \infty$$

при $m \to \infty$, что противоречит ограниченности $\{x^s, s \ge n_k\}$ и, следовательно, доказывает лемму.

С помощью приведенных выше лемм можно доказать окончательное утверждение о сходимости алгоритмов вида (6).

Теорема 4. *Пусть последовательность* $\{x^k\}$ *построена по правилу* (11), *где* F — *локально-сильный липшицево-непрерывный фейеровский оператор относительно* X, G — *ограниченный полунепрерывный сверху аттрактант множества* $X_* = \{x^* : G(x^*) + N_X^+(x^*)\}$, *последовательность шаговых множителей определяется правилом* (14), (15). *Тогда все ее предельные точки принадлежат* X_*.

Доказательство. Леммы 1, 2 гарантировали допустимость предельных точек $\{x^k\}$ и выполнение в каждой из них условия **A2**. Поскольку **A4** выполнены по построению, то остается проверить выполнение условий **A3**, **A5**.

Определим $W(x) = \inf_{x^* \in X_*}\|x - x^*\|^2$, тогда условие **A5** выполнится автоматически. Далее имеем

$$\|x^{s+1} - x^*\|^2 = \|F(x^s - \lambda_s g^s) - x^*\|^2 \le \|\alpha x^s - \lambda_s g^s - x^*\|^2 \le$$
$$\le \alpha(\|x^s - x^*\|^2 - 2q\lambda_{n_k} g^s(x^s - x^*) + \lambda_s^2\|g^s\|^2) \le$$
$$\le \alpha(\|x^s - x^*\|^2 - 2q\lambda_{n_k}\delta + q^2\lambda_{n_k}^2 C) \le \alpha\|x^s - x^*\|^2 - \gamma\lambda_{n_k}$$

с некоторой константой $\gamma > 0$. Вычисляя $\inf$, получаем

$$W(x^{s+1}) \le \alpha W(x^s) - \gamma\lambda_{n_k} \le W(x^s) - \gamma\lambda_{n_k}.$$

Сложив эти неравенства по s от n_k до $m_k - 1$, получим

$$W(x^{m_k}) \le W(x^{n_k}) - \gamma\lambda_{n_k}(m_k - n_k - 1). \tag{16}$$

С другой стороны,

$$\varepsilon < \left\|x^{m_k} - x^{n_k}\right\| \le \sum_{s=n_k}^{m_k-1}\left\|x^{s+1} - x^s\right\| \le Lq\lambda_{n_k}(m_k - n_k) \le \kappa\lambda_{n_k}(m_k - n_k),$$

откуда получалось $\lambda_{n_k}(m_k - n_k) > \varepsilon / \kappa$. Подставляя эту оценку в (16), получаем

$$W(x^{m_k}) \le W(x^{n_k}) - \varepsilon/\kappa + \gamma\lambda_{n_k} \le W(x^{n_k}) - \varepsilon/2\kappa$$

для достаточно больших k, что в пределе дает

$$\limsup_{k\to\infty} W(x^{m_k}) \le \lim_{k\to\infty} W(x^{n_k}) - \varepsilon/2\kappa < W(x').$$

Это доказывает выполнение условия **А3** и, следовательно, окончательно доказывает теорему.

4. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Вычислительные эксперименты, проведенные с алгоритмами, реализующими идеи адаптивной ступенчатой регулировки шага, были направлены в первую очередь на выяснение качественного характера сходимости для задач различного класса. В связи с этим были рассмотрены в первую очередь задачи условной и безусловной выпуклой оптимизации различной структуры, а также вариационное неравенство, не сводящееся к экстремальной задаче. Все это позволило опробовать общую идею регулировки шага на задачах, традиционно рассматривавшихся в рамках отдельных направлений вычислительной математики и попробовать сделать первые выводы о практической применимости развиваемого подхода.

4.1. Выпуклые экстремальные задачи

Первая из тестовых задач представляла собой минимизацию линейной целевой функции на множестве, заданном системой выпуклых квадратичных цилиндрических ограничений:

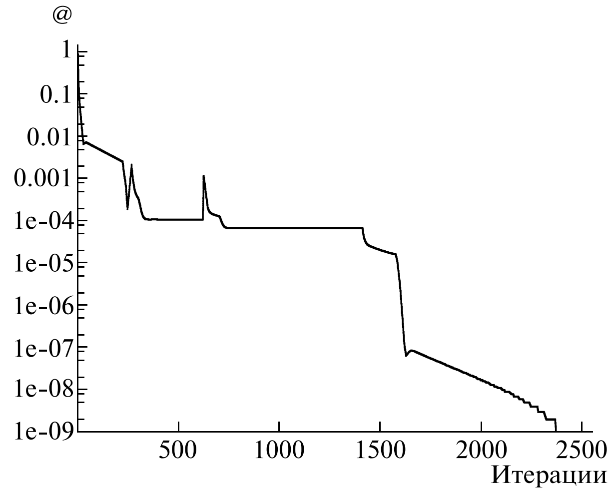
$$\min cx$$
$$\sum_{i=1, i\ne k}^{n} x_i^2 \le 1, \quad k = 1, 2, \ldots, n. \tag{17}$$

Задача решалась для небольшой размерности $n = 4$, однако даже в этом случае нахождение экстремальной точки на границе представляет собой достаточно сложную проблему. На фиг. 1 представлена сходимость градиентного метода решения этой задачи с последовательным проектированием на систему ограничений.

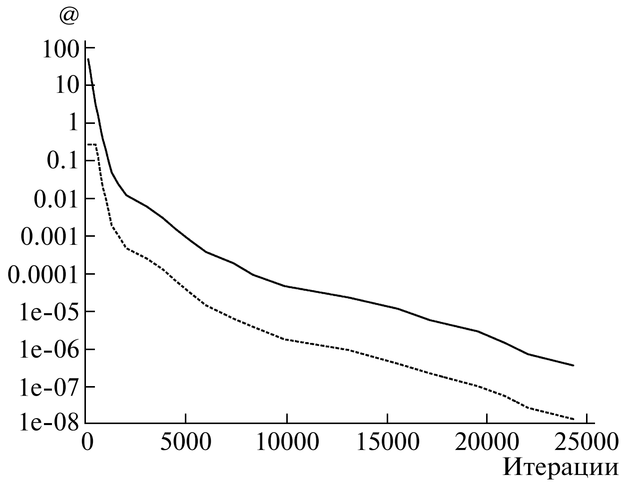
В этом случае из текущей точки итерационной последовательности x^k производился градиентный шаг $\overline{x}^{k+1} = x^k - \lambda_k c$ и затем производилось проектирование на j_k-е ограничение, определяемое соотношением $j_k = 1 + k \bmod n$:

$$\left\|x^{k+1} - \overline{x}^{k+1}\right\| = \min\left\|x - \overline{x}^{k+1}\right\|, \quad x \in C(j_k), \quad C(j_k) = \left\{x : \sum_{i=1, i\ne j_k}^{n} x_i^2 \le 1\right\},$$

Такая политика в задачах выпуклой допустимости называется политикой циклического (round-robin) обхода ограничений. На фиг. 1 представлены только итерации, соответствующие допустимым точкам, что дает возможность оценить тип сходимости алгоритма по отношению к поиску оптимального значения целевой функции. Можно заметить, что несмотря на значительные лакуны, вызванные недопустимыми точками (горизонтальные линии на графике), алгоритм демонстрирует в основном линейную скорость сходимости.



Фиг. 1.

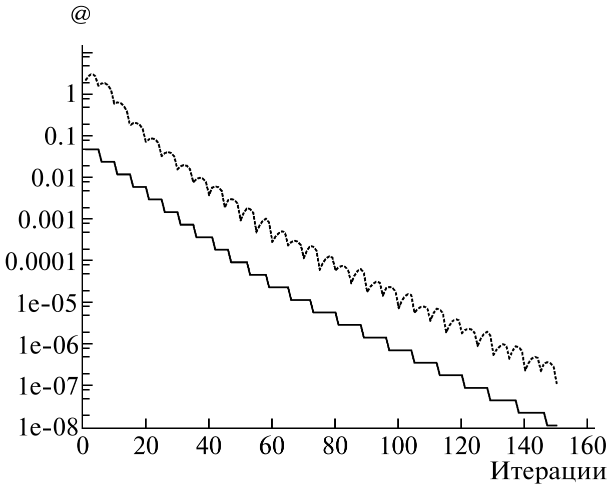


Фиг. 2.

Вычислительные эксперименты были проведены также с задачей TR48, двойственная к которой известна как достаточно сложная проблема для субградиентных алгоритмов. В последней требуется минимизировать кусочно-линейную функцию

$$-f(x) = \sum_{i=1}^{48} s_i x_i + \sum_{j=1}^{48} d_j \min_{i=1,\ldots,48} (a_{ij} - x_i), \tag{18}$$

где 48×48-матрица A и 48-мерные векторы s и d заданы, как в [11]. Фортран и MATLAB-совместимый код для этой задачи вместе с необходимым данными может быть найден на сайте [12] российско-украинского проекта РФФИ (09-01-90413-Укр_ф_а) и ДФФД (Ф28.1.005) в разделе “Тесты” ⟶ “Тестовые функции” ⟶ “medium”. На фиг. 2 продемонстрирована сходимость к нулю относительной ошибки определения оптимального значения целевой функции (сплошная линия) и убывание шагового множителя (пунктир) для решения задачи TR48. Хорошо виден линейный, хотя и не слишком быстрый характер убывания этих величин, что говорит также и о равномерной ограниченности количества итераций между делениями шага, по-видимому характерной для выпуклых полиэдральных функций, как функций с “острым минимумом”.



Фиг. 3.

4.2. Вариационные неравенства

В качестве примера рассмотрим вариационное неравенство $G(x)(y - x) \geq 0,\ x \in X$, где

$$G(x) = (1 + \|x\|)(x - x^*), \quad X = \{x \leq 0\}, \tag{19}$$

с точкой $x^* = (1, 2, \ldots, n)$. Решением этого неравенства является точка $(0, \ldots, 0)$, и на фиг. 3 продемонстрирована сходимость итеративной последовательности, порождаемой алгоритмом вида (5), со ступенчатой регулировкой шага с последовательным учетом ограничений неположительности переменных. Так же, как и выше, по ходу итеративного процесса на k -й итерации производилась проверка j_k-го ограничения $x_{j_k} \leq 0$, где $j_k = 1 + k \bmod n$, и если оно нарушалось, то выполнялась тривиальная проекция на полупространство $H(j_k) = \{x : x_{j_k} \leq 0\}$. На фиг. 3 показано поведение евклидового расстояния до решения (сплошная линия) и убывание длины шага (пунктир). Как и в приведенных выше экспериментах с задачами выпуклой оптимизации, хорошо видна линейная сходимость алгоритма.

ВЫВОДЫ

Как из теоретического обоснования метода адаптивной ступенчатой регулировки шага в фейеровских процессах с аттрактантами, так и из вычислительных экспериментов можно сделать вывод о широкой применимости этого подхода для решений задач различной структуры. Особый прагматический интерес представляет собой экспериментально наблюдаемый линейный характер сходимости, что существенно ускоряет фейеровские процессы по сравнению с классическими условиями малости шаговых множителей и расходимости их суммы, однако данный факт еще нуждается в строгом теоретическом обосновании.

СПИСОК ЛИТЕРАТУРЫ

1. *Нурминский Е.А.* Использование дополнительных малых воздействий в фейеровских моделях итеративных алгоритмов // Ж. вычисл. матем. и матем. физ. 2008. Т. 48. № 12. С. 2121–2128.
2. *Нурминский Е.А.* Фейеровские процессы с малыми возмущениями // Докл. РАН. 2008. Т. 422. Вып. 5. С. 601–604.
3. *Nurminski E.A.* Envelope stepsize control for iterative algorithms based on Fejer processes with attractants // Optmizat. Methods and Software. 2010. V. 25, № 1 . P. 97–108.
4. *Васин В.В.. Еремин И.И.* Операторы и итерационные процессы фейеровского типа. Теория и приложения. Екатеринбург: УрО РАН, 2005/
5. *Bauschke H.H., Borwein J.M.* On projection algorithms for solving convex feasibility problems // SIAM Revs. 1996. V. 38. № 3. P. 367–426.
6. *Пшеничный Б.Н.* Выпуклый анализ и экстремальные задач. М.: Наука, 1980.
7. *Нурминский Е.А.* Численные методы выпуклой оптимизации. М.: Наука, 1991.

8. *Fukushima M.* A relaxed projection method for variational inequalities // Math. Program. 1986. V. 35. P. 58–70.
9. *Byrne C.L.* A unified treatment of some iterative algorithms in signal processing and image reconstruction // Inverse Problems.v2004. V. 20. № 1. P. 103–120.
10. *Yang Q.* The relaxed CQ algorithm solving the split feasibility problem // Inverse Problems, 2004. V. 20. № 4. P. 1261–1266.
11. Nonsmooth optimization // Proc. IASA Workshop, March 28–April 8, 1977 (Internat. Inst. Appl. Systems Analys. IIASA proc. series: Vol. 3) / Eds. Lemarechal C., Mifflin R. Oxford: Pergamon Press, 1978.
12. Совместный российско-украинский проект РФФИ (Российский фонд фундаментальных исследований, Россия) и ДФФД (Державний фонд фундаментальних досліджень, Украіна) “Субградиентные методы ускоренной сходимости в задачах выпуклого программирования” [Электронный ресурс]. Метод доступа http://elis.dvo.ru.