Министерство высшего и среднего специального образования
Р С Ф С Р

Горьковский ордена Трудового Красного Знамени
научно-исследовательский радиофизический институт (НИРФИ)

Препринт № 140

НЕАСИМПТОТИЧЕСКИЕ ОЦЕНКИ СКОРОСТИ СХОДИМОСТИ ПСЕВДОГРАДИЕНТНЫХ ИТЕРАТИВНЫХ АЛГОРИТМОВ

Альбер Я.И.,

Шильман С.В.

Горький 1980

УДК 518.5

Устанавливаются неасимптотические (справедливые с первого шага) аналитические оценки скорости сходимости детерминированных и стохастических итеративных алгоритмов для задач, удовлетворяющих условиям сильной и равномерной псевдоградиентности.

ВВЕДЕНИЕ

Итеративные алгоритмы являются широко распространенным средством решения алгебраических уравнений, задач безусловной и условной оптимизации, оценивания параметров и состояний, проблем адаптации, обучения, планирования и т.д. В связи с этим изучение свойств этих процедур имеет важное практическое значение. В настоящее время наиболее полно изучена проблема сходимости [1 ÷ 9]. Значительно меньше исследован весьма важный вопрос о скорости сходимости этих алгоритмов. Известные в литературе оценки, как правило, носят асимптотический характер [2, 6, 9]. Кроме того, все они получены только для тех задач оптимизации, которые удовлетворяют условию сильной выпуклости или более общему условию сильной псевдоградиентности. В этих рамках получено большинство известных рекомендаций по выбору параметров детерминированных и стохастических алгоритмов.

Ясно, что указанных результатов совершенно недостаточно для определения необходимого числа итераций, обеспечивающего заданную точность вычисления решения. Более того, асимптотические оценки могут оказаться справедливыми при весьма большом (практически не реализуемом) числе итераций. Отсюда следует, что вопрос о возможности использования вытекающих из них рекомендаций при практических вычислениях, вообще говоря, является открытым.

Наряду с этим обстоятельством отметим, что изучение итеративных процессов при условии сильной псевдоградиентности исключает из рассмотрения важный и достаточно широкий класс вырожденных задач. К последним относятся, в частности, задачи поиска экстремума функций, имеющих в точке экстремума вырожденный гессиан (особую матрицу вторых производных).

В настоящей работе устанавливаются неасимптотические, справедливые с первого шага, аналитические оценки скорости сходимости итеративных алгоритмов (детерминированных и стохастических), удовлетворяющих как условию сильной псевдоградиентности, так и более общему условию равномерной псевдоградиентности. В основе этого исследования лежит изучение числовых рекуррентных неравенств специального вида [10].

2. Постановка задачи

Будем рассматривать алгоритмы, описываемые следующими рекуррентными соотношениями

$$x_{n+1} = \pi_G \left[x_n - \alpha_n S_n (x_n) \right], \quad n = 1, 2 \ldots , \qquad (1)$$

где x_n – m – мерный вектор из пространства R^m, $S_n(x_n)$ – вектор-функция, называемая при принятых ниже предположениях псевдоградиентом, α_n – последовательность неотрицательных чисел.

Через π_G в (1) обозначена операция проектирования на выпуклую, замкнутую область $G \subset R^m$. Допускается, что $G = R^m$, тогда $\pi_G = E$, где E – единичная матрица. Псевдоградиент $S_n(x_n)$ может быть как детерминированным, так и случайным вектором.

Соотношение (1) охватывает большое число детерминированных и стохастических итеративных процедур. Сюда относятся градиентные методы и метод обобщенных градиентов для негладких функций, стохастические алгоритмы Роббинса-Монро, Кифера-Вольфовица, случайного поиска и другие.

Обозначим через X^* непустое множество решений задачи, поиск которых осуществляется процедурой (1). Введем показатель точности поиска за n шагов – $\lambda_n = V(x_n)$ и $\lambda_n = M[V(x_n)]$ соответственно в детерминированном и стохастическом случаях. Здесь $V(x), x \in Q$, некоторая неотрицательная функция, которую можно рассматривать как аналог функции Ляпунова, M – символ математического ожидания. Обычно в качестве $V(x)$ используется величина $\rho^2(x, X^*)$, равная квадрату расстояния от x до выпуклого, замкнутого множества X^*. Эта функция будет применяться в данной работе во всех тех случаях, когда Q не совпадает с R^m. Если же $Q = R^m$, то наряду с $\rho^2(x, X^*)$ можно брать, вообще говоря, любые функции, удовлетворяющие:

1) условию $\inf\limits_{x \in Q} V(x) = 0$,

2) условию Липшица-Гельдера для градиента $\nabla V(x)$ функции $V(x)$

$$\|\nabla V(x) - \nabla V(y)\| \le L \|x - y\|^{\mu}, \quad x, y \in Q$$

где $L = const$, $\mu = const$, $0 \le \mu \le 1$. В частности, при исследовании алгоритмов минимизации функции многих переменных $f(x)$ в качестве $V(x)$ часто используется выражение $f(x) - \inf\limits_{x} f(x)$ [7].

Множество алгоритмов и решаемых задач будем характеризовать следующими двумя требованиями, формулируемыми по-разному для стохастических и детерминированных процессов.

1 А. Пусть $S_n(x_n)$ –детерминированный вектор. Алгоритм (1) должен быть равномерно псевдоградиентным степенного типа по отношению к выбранной функции Ляпунова $V(x)$ в следующем смысле

$$(\nabla V(x_n),\ S_n(x_n)) \geq \theta_n V^p(x_n) - h_n - \nu_n V^{q_1}(x_n), x_i \in G, \tag{2}$$

где $\theta_n > 0,\ h_n \geq 0,\ \nu_n \geq 0,\ p > 0,\ 0 \leq q_1 \leq 1.$

Условие (2) будем называть условием сильной псевдоградиентности алгоритма, если $p = 1$.

1 Б. Теперь пусть $S_n(x_n)$ является случайным вектором. Тогда условие, подобное (2), имеет вид

$$(\nabla V(x_n), M[S_n(x_n)/x_1, \ldots x_n]) \geq \theta_n \Psi[V(x_n)] - \tag{3}$$

$$- h_n - \nu_n [V(x_n)]^{q_1},$$

где $x_i \in G$, $\Psi(V)$ выпуклая возрастающая функция, $\Psi(0) = 0$, а остальные параметры носят такой же характер, как и в соотношении (2).

2 А. Введем еще одно условие, которое для детерминированных процессов описывается неравенством

$$\|S_n(x_n)\|^{1+\mu} \leq \bar{\varphi}(V(x_n)), \tag{4}$$

где $\bar{\varphi}(V)$– неубывающая функция от V, $\bar{\varphi}(0) < \infty$.

2Б. Для стохастических алгоритмов аналогичное требование представим в виде

$$M\left[\|S_n(x_n)\|^{1+\mu} / x_1, \ldots, x_n\right] \leq \sigma_n + \tau_n V^q(x_n), \tag{5}$$

где $\sigma_n \geq 0$, $\tau_n \geq 0$, $q > 0$.

Условия (4), (5) должны быть согласованы с соответствующими предположениями (2) и (3).

Отметим, что отрицательные члены в правой части неравенств (2), (3) вводятся, в частности, в тех случаях, когда либо алгоритмы носят поисковый характер, либо учитываются систематические погрешности, обусловленные ошибками вычислений, неточностью исходных данных и другими причинами.

3. Вывод рекуррентных числовых неравенств

А. Детерминированные алгоритмы.

Используя формулу конечных приращений, можно найти

$$V(x_{n+1}) \leq V(x_n) - \alpha_n (\nabla V(x_n), S_n(x_n)) + \frac{\alpha_n^{1+\mu}}{1+\mu} L \|S_n(x_n)\|^{1+\mu} \tag{6}$$

Это неравенство справедливо и для алгоритмов с проектированием, если в качестве $V(x)$ взять $\rho^2(x, X^*)$. При этом следует положить $L = 2$, $\mu = 1$, а

$$\nabla V(x_n) = x_n - \pi_{X^*}(x_n) \cdot$$

С учетом предположений 1А, 2А из (6) следует неравенство вида:

$$\lambda_{n+1} \leq \lambda_n - \alpha_n \theta_n \lambda_n^p + \alpha_n h_n + \alpha_n \nu_n \lambda_n^{q/1} + \frac{\alpha_n^{1+\mu}}{1+\mu} L \bar{\varphi}(\lambda_n), \tag{7}$$

$$\lambda_n = V(x_n).$$

Примем далее, что $\theta_n = \theta = \mathrm{const}, \alpha_n = \frac{\alpha}{(n+n_0)^t}, h_n = \frac{h}{(n+n_0)^{\varepsilon_1}},$

$\nu_n = \frac{\nu}{(n+n_0)^{\varepsilon_2}}$, где $\alpha > 0, h \geq 0, \nu \geq 0, 0 < t \leq 1,$
$\varepsilon_1 > 0, \quad \varepsilon_2 > 0, \quad n_0 \geq 0$ – произвольное число.
Тогда из (7) можно найти, что

$$\lambda_{n+1} \leq \lambda_n - \frac{\alpha\theta}{(n+n_0)^t} \lambda_n^p + \frac{1}{(n+n_0)^s} g(\lambda_n), \tag{8}$$

где $s = t + \min\{\varepsilon_1, \varepsilon_2, t\mu\}$, $g(\lambda_n)$ –
– неотрицательная, неубывающая функция, $g(0) < \infty$.

Для дальнейшего потребуется следующая лемма.

<u>Лемма 1.</u> Пусть последовательность неотрицательных чисел $\lambda_1, \dots, \lambda_n, \dots$ удовлетворяет неравенству

(8). Тогда, если $\lambda_1 \leq R_0$ и

$$n_0 > \overline{\sigma} = \min\left\{K : \frac{1}{\alpha\theta(1+K)^{s-t}}\, g\left[R_0 + \frac{g(R_0)}{(1+k)^s}\right] \leq R_0^p\right\},$$

то при всех $n \geq 1$ $\lambda_n \leq R_1$, где

$$R_1 = R_0 + \frac{1}{(1+n_0)^s}\, g(R_0)$$

<u>Доказательство.</u> Пусть $\lambda_1 \leq R_0$. Заметим, что если $\lambda_i \leq R_0$, то в силу (8)

$$\lambda_{n+1} \leq \lambda_n + \frac{1}{(n+n_0)^s}\, g(\lambda_n) \leq R_1$$

Предположим теперь, что номеру $n = k+1$ соответствует первый член последовательности $\{\lambda_n\}$, значение которого становится больше R_1. Тогда λ_k должно обязательно принадлежать интервалу (R_0, R_1). Для λ_k всегда можно допустить одну из исключающих друг друга возможностей:

$$H_1 : \lambda_k^p \leq \frac{1}{\alpha\theta(n+n_0)^{s-t}}\, g(\lambda_k),$$

$$H_2 : \lambda_k^p > \frac{1}{\alpha\theta(n+n_0)^{s-t}}\, g(\lambda_k).$$

Из H_1 следует, что при значениях n_0, указанных в условиях леммы, $\lambda_k \leq R_0$, что противоречит тому, что $R_0 < \lambda_k < R_1$. Теперь примем, что выполняется альтернатива H_2, тогда в силу (8) получаем $\lambda_{k+1} \leq \lambda_k \leq R_1$

Эта оценка также противоречит исходному предположению. Таким образом, остается утверждение леммы: $\lambda_n \leq R_1$ при всех $n \geq 1$.

При указанных в лемме условиях из (8) для детерминированных алгоритмов получаем неравенство вида

$$\lambda_{n+1} \leq \lambda_n - \frac{\alpha\theta}{(n+n_0)^t}\lambda_n^p + \frac{d}{(n+n_0)^s}, \tag{9}$$

где $d = g(R_1)$, $s > t$.

Б. <u>Стохастические алгоритмы.</u>

Используем в качестве исходной формулу (6). Далее перейдем в (6) к условным относительно $x_1, \ldots, x_n$ математическим ожиданиям. С учетом предположений 1Б, 2Б получим

$$\begin{aligned} M\left[V(x_{n+1}) \middle| x_1, \ldots, x_n\right] \leq V(x_n) - \alpha_n \theta_n \Psi\left[V(x_n)\right] + \\ + \alpha_n \nu_n V^{q_1}(x_n) + (1+\mu)^{-1} \alpha_n^{1+\mu} L \tau_n V^{q}(x_n) + \\ + \alpha_n h_n + (1+\mu)^{-1} L \alpha_n^{1+\mu} \sigma_n . \end{aligned} \tag{10}$$

Теперь задача заключается в получении из (10) неравенства для $\lambda_n = M\left[V(x_n)\right]$. Для этого воспользуемся известными неравенствами

$$M\left[\Psi(V)\right] \geq \Psi\left[M(V)\right], \quad M(V^{q_1}) \leq (M\left[V\right])^{q_1},$$

где $0 \leq q_1 \leq 1$. В силу неравенства Юнга

$$\xi^{1-\tau}\eta^{\tau} \leq (1-\tau)\xi + \tau\eta, \quad \xi > 0, \eta > 0, 0 \leq \tau \leq 1,$$

дополнительно находим, что

$$(MV)^{q_1} \leq \left(1 - \frac{q_1}{p_1}\right)\xi^{\frac{q_1}{p_1}} + \frac{q_1}{p_1}(MV)^{p_1}\xi^{\frac{q_1-p_1}{p_1}},$$

где $p_1 \geq q_1$, ξ – произвольное положительное число. Однако всего этого недостаточно в тех случаях, когда $q > 1$, поэтому рассмотрим следующие случаи.

А. Пусть $q = 1$. Тогда после перехода в (10) к безусловным математическим ожиданиям, найдем

$$\lambda_{n+1} \leq (1+\beta_n')\lambda_n - \alpha_n \theta_n \Psi\left[\lambda_n\right] + \alpha_n x_n' + \tag{11}$$

$$+ \alpha_n \nu_n \xi^{\frac{q_1-p_1}{p_1}} \cdot \frac{q_1}{p_1} \lambda_n^{p_1},$$

где $x_n' = h_n + \alpha_n^{\mu}(1+\mu)^{-1} L\sigma_n + \nu_n(1 - q_1 p_1^{-1})\xi^{\frac{q_1}{p_1}}$,

$$\beta_n' = \alpha_n^{1+\mu}(1+\mu)^{-1} L\tau_n .$$

Здесь возможны два случая.

А1. Пусть $\Psi(\lambda) = \lambda$, т.е. выполняется условие сильной псевдоградиентности и $p_1 = 1$. Тогда неравенство (11) будет линейным, при этом члены, содержащие λ_n, можно объединять в разных комбинациях в зависимости от свойств последовательностей ν_n, τ_n, θ_n. В общем случае неравенство (11) запишем в виде

$$\lambda_{n+1} \leq (1+\beta_n)\lambda_n - \alpha_n \vartheta_n \lambda_r + \alpha_n \varkappa_n . \qquad (12)$$

А2. Пусть $\Psi(\lambda) = \lambda^p$, $p > 1$ при $0 \leq \lambda \leq R$, $R = const$. При этом можно брать $p_1 = p$ или $p_1 = 1$. Сформулируем лемму, из которой следует, что задача получения оценок в ситуации А2 может быть сведена к изучению неравенства вида

$$\lambda_{n+1} \leq (1+\beta_n)\lambda_n - \alpha_n \vartheta_n \lambda_n^p + \alpha_n \varkappa_n . \qquad (13)$$

Лемма 2. Пусть последовательность неотрицательных чисел $\lambda_1, \ldots, \lambda_n \ldots$ удовлетворяет неравенству

$$\lambda_{n+1} \leq (1+\beta_n)\lambda_n - \alpha_n \vartheta_n \Psi(\lambda_n) + \alpha_n \varkappa_n ,$$

где $\beta_n \geq 0$, $\alpha_n \vartheta_n > 0$, $\varkappa_n \geq 0$, $\prod\limits_1^{\infty}(1+\beta_n) \leq \bar{C} < \infty$

Тогда $\lambda_n \leq \max\left[\bar{C}\lambda_1, K\right]$, где $K = \sup_n\left[\bar{C}\Psi^{-1}\left(\frac{x_n}{v_n}\right) + \right.$ $\left.+ \alpha_n x_n\right]$.

Теперь примем, что $\Psi(\lambda) = \lambda^p$ при $\lambda_1 \leq R$, где $R \geq K$. Тогда при $q = 1$ и $\lambda_1 \leq \bar{C}^{-1} R$ последовательность $\{\lambda_n\}$ будет ограниченной $(\lambda_n \leq R)$ и поэтому при условии $v_n > 0$ будет удовлетворять неравенству (13).

Б. Пусть G не только замкнутое, выпуклое, но и ограниченное множество, а $q \geq 1$. В этом случае $V(x) = \rho^2(x, X^*)$ и $V(x) \leq D^2$, где D – диаметр множества G. Теперь за счет ограниченности $V(x)$ степень q можно считать произвольным числом. Если взять $q \leq 1$, то из (10) получим неравенство того же вида, что и (11), и можно рассмотреть ситуации, аналогичные А1, А2. Кроме этого случая, рассмотрим следующий: $\Psi(V) = V^p$, $p > 1$, а $q = p$. Если взять $p_1 = p$ или $p_1 = 1$, объединить члены, содержащие V^p, то при $v_n > 0$ из соотношения (10) нетрудно получить неравенство вида (13).

4. Теоремы о скорости сходимости

Итак, для итеративного процесса (1) можно получить неравенство

$$\lambda_{n+1} \leq (1+\beta_n)\lambda_n - \alpha_n v_n \lambda_n^p + \alpha_n x_n . \qquad (13)$$

Введем обозначения: $\prod\limits_{1}^{\infty}(1+\beta_n) \leq \bar{C}$,

$$c(s,t,p)=\left[\gamma_1(t,p)\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)d\right](n_0+1)^{\frac{t-s}{p}},$$

$$\gamma_1(t,p)=(1+n_0)^{\bar{s}-\frac{s-t}{p}},\quad \gamma_2(t,p)=(n_0+1)^{\bar{s}-s},$$

$$\gamma_3(t,p)=(n_0+1)^{\bar{s}(p-1)+t-1},\ \text{где}\ \bar{s}=\min\left\{\frac{s-t}{p},s\right\}.$$

Теперь изучим три случая $p=1$, $p>1$, $p<1$.

Случай $p=1$

Теорема I. Пусть $\alpha_n \vartheta_n (1+\beta_n)^{-1} \geq b(n+n_0)^{1}$, $b>0$,

$$n_0 \geq 0,\ \alpha_n \varkappa_n \leq d(n+n_0)^{-s},\ d>0\ s>1.$$

1. Если $s-1<b$, то при любых $c_0 \geq \frac{b}{b-s+1}$ (или $ab>s-1$, $a=\frac{c_0-1}{c_0}$) и при $\lambda_1 > c(s,1,1)$ справедливы оценки

$$\lambda_n \leq \lambda_1 \bar{C}\left(\frac{n_0+1}{n+n_0-1}\right)^{ab},\ 2 \leq n \leq \bar{x}=(n_0+1)\left[\lambda_1 c^{-1}(s,1,1)\right]^{\frac{1}{ab-s+1}} - \quad (14)$$

$$-n_0+1,$$

$$\lambda_n \leq \bar{C} d \left(\frac{c_0}{b} + \gamma_2(1,1) \right) \left(\frac{1}{n+n_0-1} \right)^{s-1}, \quad n > \bar{x}, \qquad (15)$$

а при $\lambda_1 \leq c(s,1,1)$ оценка (15) выполняется при всех $n \geq 2$.

2. Если $s-1 < b$ и $1 < c_0 < b(b-s+1)^{-1}$ или же $s-1 \geq b$ и $c_0 > 1$, то при всех $n \geq 2$ имеет место оценка

$$\lambda_n \leq \bar{C} C \left(\frac{n_0+1}{n+n_0-1} \right)^{ab}$$

где $C = \max\left[\lambda_1, c(s,1,1)\right]$.

Теорема 2. Пусть $\alpha_n \vartheta_n (1+\beta_n)^{-1} \geq b(n+n_0)^{-t}, n_0 > 0,$

$$b > 0,\ 0 \leq t < 1,\ \alpha_n \varkappa_n \leq d(n+n_0)^{-s},\ d > 0,\ s > t.$$

Тогда при любых $c_0 > 1$ справедливы оценки

$$\lambda_n \leq \bar{C} C \exp\left\{ -\frac{ab}{1-t} \left[(n+n_0-1)^{1-t} - (n_0+1)^{1-t} \right] \right\}, \qquad (16)$$

$$2 \leq n \leq \bar{x},$$

$$\lambda_n \leqslant \bar{C} d \left(\frac{c_0}{b} + \gamma_2(t,1)\right)\left(\frac{1}{n+n_0-1}\right)^{s-t}, \quad n > \bar{x}, \qquad (17)$$

где $C = \max\left[\lambda_1, c(s,t,1)\right]$, $a = \dfrac{c_0 - 1}{c_0}$,

$\bar{x}$ – есть единственный на интервале $(2,\infty)$ корень уравнения

$$c \exp\left\{-\frac{ab}{1-t}\left[(x+n_0-1)^{1-t} - (n_0+1)^{1-t}\right]\right\} = d\left(\frac{c_0}{b} + \gamma_2(t,1)\right)\left(\frac{1}{x+n_0-1}\right)^{s-t}.$$

При этом, если $c_0 \geqslant (n_0+1)^{1-t} b \left[(n_0+1)^{1-t} b - s + t\right]^{-1}$

и $\lambda_1 \leqslant c(s,t,1)$, то оценка (17) выполняется для всех $n \geqslant 2$.

Случай $p > 1$.

Теорема 3. Пусть $\alpha_n \vartheta_n (1+\beta_n)^{-1} \geqslant b(n+n_0)^{-1}$,

$b > 0,\ n_0 \geqslant 0,\ \alpha_n x_n \leqslant d(n+n_0)^{-s},\ d > 0,\ s > 1.$

Тогда при всех $n \geq 2$

$$\lambda_n \leq \bar{C}C\left[1+(p-1)C^{p-1}\frac{c_0-1}{c_0}b\ln\left(\frac{n+n_0-1}{n_0+1}\right)\right]^{-\frac{1}{p-1}}, \tag{18}$$

где $C=\max\left[\lambda_1, c(s,1,p)\right]$, а $c_0 > 1$ выбирается из условия

$$\frac{c_0-1}{c_0}\left[\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(1,p)d\right]^{p-1}b \leq \frac{s-1}{p}\gamma_3(1,p).$$

Теорема 4. Пусть $\alpha_n \vartheta_n (1+\beta_n)^{-1} \geq b(n+n_0)^{-t}$,

$0 \leq t < 1,\ b > 0,\ n_0 \geq 0,\ \alpha_n x_n \leq d(n+n_0)^{-s},\ s > t,\ d > 0.$

Тогда справедливы следующие утверждения.

1. Если $\frac{s-t}{p} < \frac{1-t}{p-1}$, то при любых $c_0 > 1$

$$\lambda_n \leq \bar{C}C\left\{1+C^{p-1}\frac{p-1}{1-t}\,\frac{c_0-1}{c_0}b\left[(n+n_0-1)^{1-t}-(n_0+1)^{1-t}\right]\right\}^{\frac{1}{p-1}}, \tag{19}$$

$2 \leq n \leq \bar{x}$,

$$\lambda_n \leq \bar{c}\left[\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)d\right]\left(\frac{1}{n+n_0-1}\right)^{\frac{s-t}{p}}, \quad n > \bar{x}, \tag{20}$$

где $C=\max\left[\lambda_1, c(s,t,p)\right]$, $\bar{x}$ – есть единственный на интервале $(2, \infty)$ корень уравнения

$$C\left\{1+C^{p-1}\frac{p-1}{1-t}\,\frac{c_0-1}{c_0}\,b\left[(x+n_0-1)^{1-t}-(n_0+1)^{1-t}\right]\right\}^{-\frac{1}{p-1}}=$$

$$=\left[\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)d\right]\left(\frac{1}{x+n_0-1}\right)^{s-t}.$$

При этом, если c_0 и λ_1 удовлетворяют условиям

$$\frac{c_0-1}{c_0}\left[\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)d\right]b\geq\frac{s-t}{p}\gamma_3(t,p)$$

и $\lambda_1\leq c(s,t,p)$,

то оценка (20) выполняется при всех $n\geq 2$.

2. Если $\frac{s-t}{p}\geq\frac{1-t}{p-1}$ то при любых c_0, удовлетворяющих условиям: $c_0>1$ и

$$\frac{c_0-1}{c_0}\left[\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)\,d\right]^{p-1}b\leq$$

$$\leq\frac{s-t}{p}\gamma_3(t,p),$$

при всех $n\geq 2$ имеет место оценка (19).

Случай $p < 1$

Функция $\Psi(\lambda)=\lambda^{p}$, $p < 1$ не является выпуклой, поэтому формулируемые ниже теоремы удается применить лишь к детерминированным итеративным процессам.

Теорема 5. Пусть $\beta_n \equiv 0$, $\alpha_n \vartheta_n \geq \frac{b}{n+n_0}$, $b>0$,

$$\alpha_n x_n \leq \frac{d}{(n+n_0)^s}, \quad d>0, \quad s>1.$$

Тогда при любых c_0, $c_0 > 1$,

$$\lambda_n \leq \left[C^{1-p}-(1-p)\frac{c_0-1}{c_0} b \ln \frac{n+n_0-1}{n_0+1}\right]^{\frac{1}{1-p}}, \quad 2\leq n\leq \bar{x}, \tag{21}$$

$$\lambda_n \leq \left[\gamma_1(1,p)\left(\frac{d}{b} c_0\right)^{\frac{1}{p}}+\gamma_2(1,p)d\right]\left(\frac{1}{n+n_0-1}\right)^{\bar{s}}, \quad n>\bar{x}, \tag{22}$$

где $\bar{s}=\min\left[\frac{s-1}{p}, s\right]$, $\bar{x}$ – есть единственный на интервале $(2, \infty)$ корень уравнения

$$\left[\gamma_1(1,p)\left(\frac{d}{b} c_0\right)^{\frac{1}{p}}+\gamma_2(1,p)d\right]\left(\frac{1}{x+n_0-1}\right)^{\bar{s}} =$$

$$= \left[C^{1-p}-(1-p)\frac{c_0-1}{c_0} b \ln \frac{x+n_0-1}{n_0+1}\right]^{\frac{1}{1-p}}.$$

При этом, если

$$\frac{c_0-1}{c_0} b\ \gamma_3^{-1}(1,p) \geq \left[\gamma_1(1,p)\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(1,p)d\right]^{\frac{1-p}{\bar{s}}}$$

и $\lambda_0 \leq c\ (s,1,p)$, то оценка (22) выполняется при всех $n \geq 2$.

Теорема 6. Пусть $\beta_n \equiv 0,\ \alpha_n \vartheta_n \geq \frac{b}{(n+n_0)^t},\ 0<t<1,$

$$b>0, \qquad \alpha_n \varkappa_n \leq \frac{d}{(n+n_0)^s},\ d>0,\ s>t.$$

Тогда при любых $c_0 > 1$.

$$\lambda_n \leq \left\{c^{1-p} - \frac{(1-p)}{1-t}\,\frac{c_0-1}{c_0} b\left[(n+n_0-1)^{1-t}-(n_0+1)^{1-t}\right]\right\}^{\frac{1}{1-p}} \quad (23)$$

$$2 \leq n \leq \bar{x},$$

$$\lambda_n \leq \left[\gamma_1(t,p)\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)d\right]\left(\frac{1}{n+n_0-1}\right)^{\bar{s}},\ n>\bar{x}, \quad (24)$$

где $\bar{s}=\min\left[\frac{s-t}{p},\ s\right]$, $\bar{x}$ – единственный на интервале $(2,\ \infty)$ корень уравнения

$$\left[\gamma_1(t,p)\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)d\right]\frac{1}{(x+n_0-1)^{\bar{s}}} =$$

$$=\left\{c^{1-p}-\frac{(1-p)}{1-t}\cdot\frac{c_0-1}{c_0}b\left[(x+n_0-1)^{1-t}-(n_0+1)^{1-t}\right]\right\}^{\frac{1}{1-p}}$$

При этом, если $\lambda_1 \leqslant c\,(s,t,p)$ и

$$\frac{c_0-1}{c_0}b\,\gamma_3^{-1}(t,p)\geqslant\left[\gamma_1(t,p)\left(\frac{d}{b}c_0\right)^{\frac{1}{p}}+\gamma_2(t,p)d\right]^{1-p}\bar{s},$$

то оценка (24) выполняется при всех $n \geqslant 2$.

Все приведенные теоремы вытекают, как уже отмечалось из исследования рекуррентных числовых неравенств типа (9), (13). Это исследование было выполнено авторами, результаты его достаточно подпобно описаны в работе [10]. По этой причине доказательство теорем 1 ÷ 6 здесь не приводится.

5. Анализ некоторых конкретных алгоритмов

1. Алгоритм Роббинса-Монро (А.Р.М.)

Пусть требуется найти (единственное) решение x^* системы уравнений:

$$\varphi(x)=0, \quad \varphi\cdot R^m \longrightarrow R^m.$$

В условиях, когда предоставляется возможность вычислять значения вектора $\varphi(x)$ лишь в фиксированных точках и с некоторой аддитивной погрешностью, для поиска x^* можно использовать алгоритм вида (1). При этом

$$S_n(x_n) = \varphi(x_n) + \eta_n,$$

где η_n - последовательность независимых случайных векторов,

$$M[\eta_n] = 0, \; M[\|\eta_n\|^2] \leq \sigma_1^2 .$$

Возьмем

$$\alpha_n = \alpha (n + n_0)^{-t}, \; 0 < t \leq 1, \; n_0 \geq 0$$

и предположим, что для $x \in Q$ выполняются требования

А. $(\varphi(x), x - x^*) \geq \theta \|x - x^*\|^{2p}, \; p \geq 1, \theta > 0,$

Б. $\|\varphi(x)\| \leq \sigma_2^2 + \tau \|x - x^*\|^{2q}, q > 0, \tau > 0, \sigma_2 = \text{const}$

Из А и Б следуют соответственно условия 1 Б при $\theta_n = \theta$, $h_n \equiv 0$, $\nu_n \equiv 0$ и 2 Б при $\sigma_n = \sigma_2^2 + \sigma_1^2 = \sigma^2$, $\tau_n = \tau$.

В принятых условиях для $\lambda_n = M[\|x_n - x^*\|^2]$ справедливо неравенство (13), в котором

1) при $q = p$ и $n_0 > \left(\frac{\alpha\tau}{2\theta}\right)^{\frac{1}{t}}$; $\vartheta_n = 2\theta - \tau\alpha_n > 0$, $\beta_n = 0$,
$\varkappa_n = \alpha_n \sigma^2$,

или

2) при $q = p_2 = 1$, $\beta_n = \tau\alpha_n^2$, $\vartheta_n = 2\theta$,
$\varkappa_n = \alpha_n \sigma^2$.

Подставляя $\alpha_n = \alpha(n + n_0)^{-t}$ из (13) можно получить

$$\lambda_{n+1} \le (1+\beta_n)\lambda_n - \frac{b_1}{(n+n_0)^t}\lambda_n^p - \frac{d}{(n+n_0)^{2t}}, \qquad (25)$$

где $d = \alpha^2\sigma^2$, $b_1 > 0$, $b_1 = \alpha\left(2\theta - \frac{\alpha\tau}{(n_0+1)^t}\right)$ при $q = p$ или $b_1 = 2\alpha\theta$ при $q = p = 1$.

Полученное неравенство (25) позволяет применить теоремы 1 ÷ 4 для оценивания скорости сходимости А.Р.М., если положить $s = 2t$, $b = b_1$ при $\beta_n = 0$, $b = b_1(1+\beta_1)^{-1}$ при $\beta_n \neq 0$.

Сравним вытекающие из теоремы 1–4 результаты с известными асимптотическими оценками, полученным при $p = 1$. Заметим, что выбором c_0 мажорирующую константу в (15), (17) можно сделать сколь угодно близкой к константам, приведенным в леммах Чжуна [5]. Отсюда следует, что при $n \to \infty$, $t = 1$, $b > s-1$ и $0 < t < 1$ $b > 0$ оценки теорем 1, 2 совпадают с известными асимптотическими оценками, установленными для А.Р.М. при условии сильной псевдоградиентности. В случае $t = 1$, $0 < b < s-1$ порядок оценки теоремы 1 совпадает с оценками Вентера с точностью до произвольного множителя $a = \frac{c_0 - 1}{c_0} < 1$. При этом в теореме 1 явно указана мажорирующая константа, которая отсутствует в оценках Вентера. Эта константа, в отличие от результатов многих авторов, остается ограниченной при $b \to s-1$.

Что касается $p > 1$, то все оценки, вытекающие из теорем 3 и 4 являются новыми.

Теперь рассмотрим вопрос о выборе оптимальных (с точки зрения полученных асимптотических оценок) параметров А.Р.М. Хорошо известно, что максимальная асимптотическая скорость достигается при $t = 1$ и при этом $\lambda_n = 0\ (n^{-1})$. Однако, как показывают полученные в работе оценки, этот результат справедлив только при условии сильной псевдоградиентности ($p = 1$). Кроме того, он не может быть распространен на весь итеративный процесс, поскольку найденные оценки при достаточно больших λ_1 описываются, как правило, различными аналитическими выражениями при $n \leq \bar{x}$ и при $n > \bar{x}$. Следует особо подчеркнуть, что для фиксированных значений параметров p, b, d и n_0 и достаточно больших λ_1 оценки скорости сходимости при $n \leq \bar{x}$ достигают максимального значения при $t \to 0$. Это подтверждает целесообразность выбора неубывающего шагового множителя d_n на начальном этапе поиска решения, когда начальное отклонение достаточно велико.

Зависимость оценок от параметра p позволяет проанализировать влияние структуры оператора $\varphi(x)$ на скорость сходимости. В частности, при фиксированном t

из оценок следует уменьшение с ростом p скорости сходимости А.Р.М. В асимптотике максимальная скорость достигается при $t = \frac{p}{2p-1}$, при этом

$$\lambda_n = O\left[\left(\frac{1}{n}\right)^{\frac{1}{2p-1}}\right].$$

2. Алгоритм Кифера–Вольфовица (А.К.В)

Для решения задачи

$$\min f(\overset{o}{x}), \ f: R^m \longrightarrow R^1,$$

применим процедуру (1), где вектор $S_n(x_n)$ имеет компоненты

$$\left\{(2c_n)^{-1}\left[f(x_n + c_n e_i) - f(x_n - c_n e_i) + \eta_{ni}\right]\right\} i = 1, \ldots, m, \{e_i\}$$

– ортонормированный базис в R^m, $\eta_n [\eta_{n1}, \ldots, \eta_{nm}]$ – последовательность независимых случайных векторов, c_n – заданная последовательность положительных чисел.

Предположим, что $f(x)$ дифференцируема и существует такое $\varepsilon > 0$, что ее градиент $\varphi(x)$ при

$$\| x - y \| < \varepsilon, \ x, y \in R^m,$$

удовлетворяет условию Липшица–Гельдера

$$\| \varphi(x) - \varphi(y) \| \leq L \| x - y \|^{\gamma}, \ L > 0, \ \gamma > 0.$$

Потребуем также, чтобы для $\varphi(x)$ выполнялись те же условия А и В, которые были введены для А.Р.М.

Используя формулу Лагранжа, можно показать, что

$$S_n(x_n) = \varphi(x_n) + \Delta S_n(x_n) + \frac{1}{2}\eta_n c_n^{-1},$$

$$\|\Delta S_n(x_n)\| \leq \sqrt{m}\,\varkappa(1+\mu)^{-1} c_n^{\gamma}.$$

В силу условий А, В справедливо неравенство (13), в котором $\lambda_n = M\left[\|x_n - x^*\|^2\right]$ и

1) $\beta_n = 0,\ \vartheta_n = 2\theta - \delta - 2\tau\alpha_n > 0 \left(p = q, n_0 > \left(\frac{2\tau\alpha}{2\theta\delta}\right)^{\frac{1}{t}}\right)$, или

2) $\beta_n = 2\tau\alpha_n^2,\ \vartheta_n = 2\theta - \delta > 0\ (p = q = 1)$; δ -

произвольное положительное число $(\delta < 2\theta)$,

$$\varkappa_n = 2\alpha_n\left[(1+\gamma)^{-2}\varkappa^2 m\, c_n^{2\gamma} + \sigma_2^2 + 2^{-3} c_n^{-2}\sigma_1^2 + \alpha_n^{-1} Q c_n^{2p\gamma(2p-1)}\right],$$

$$Q = \frac{\varkappa\sqrt{m}\,(2p-1)}{2p(1+\gamma)}\left(\frac{\varkappa\sqrt{m}}{p\delta(1+\gamma)}\right)^{\frac{1}{2p-1}}.$$

При выводе (13) используется получаемое из неравенства Юнга соотношение

$$(1+\mu)^{-1} L c_n^{\gamma}\sqrt{m}\,\lambda_n^{\frac{1}{2}} \leq Q c_n^{\frac{2p\gamma}{2p-1}} + 2^{-1}\delta\lambda_n^{p}.$$

Пусть $\alpha_n = \alpha (n+n_0)^{-t}$, $c_n = c\,(n+n_0)^{-\tau}$, $\alpha > 0$, $c>0$, $0<\tau<\frac{t}{2}$.

Тогда для А.К.В. имеет место неравенство (25), в котором b_1 следует заменить на $b_1 - \alpha\delta$, а $s = \min\left[2(t-\tau),\ t + 2(2p-1)^{-1} p\mu\tau\right]$, $d = 2\alpha^2 \max\left[(1+\mu)^{-2} \mathcal{L}^2 m c^{2\gamma},\ \sigma_2^2,\ 2^{-3} c^{-2} \sigma_1^2,\ Q\alpha^{-1} c^{\frac{2p\mu}{2p-1}}\right]$.

Таким образом, все утверждения и оценки теорем 1–4 справедливы и для А.К.В. с указанными значениями параметров. Это позволяет все выводы относительно оценок скорости сходимости А.Р.М. на начальном этапе поиска перенести и на метод А.К.В.

Рассмотрим отдельно вопрос о выборе оптимальных (с точки зрения асимптотических оценок) параметров t и τ для А.К.В. Из выражения для s вытекает, что $\max\limits_{\tau}(s-t) = p\gamma \,/\, (2p+p\gamma-1)$ и достигается при $\tau^* = (2p-1)t \,/\, 2(2p+p\gamma-1)$. Из теорем 1–4 следует, что наибольшая скорость сходимости по найденным оценкам достигается при $(1-t)(p-1)^{-1} = p\max\limits_{\tau}(s-t)$.

Если p фиксировано, а $\tau = \tau^*$, то это равенство выполняется при $t^* = \dfrac{2p+p\gamma-1}{(1+\gamma)(2p-1)}$. Для значений $t = t^*$, $\tau = \tau^*$ асимптотическая оценка имеет вид

$$\lambda_n \leq O\left[\left(\frac{1}{n}\right)^{\frac{\gamma}{(1+\gamma)(2p-1)}}\right].$$

В частности, при $p = 1$ получаем $t^* = 1$, $\tau^* = \left[2(1+\gamma)\right]^{-1}$, $\lambda_n = O\left[n^{-\frac{\gamma}{1+\gamma}}\right]$. Если $p = \gamma = 1$, то $t^* = 1$ $\tau^* = 1/4$, $\lambda_n = O\left[\left(\frac{1}{n}\right)^{\frac{1}{2}}\right]$. Это полностью совпадает с известными результатами Дӱча [5]. При всех других значениях γ и p приведенные оценки и соответствующие им оптимальные параметры t^* и τ^* являются новыми. Они показывают, что для заданного класса минимизируемых функций асимптотическая скорость сходимости А.К.В. существенно зависит от его гладкости (показатель γ) и от его структуры (показатель p).

3. Алгоритм случайного поиска (А.С.П.)

Для решения задачи поиска экстремума функции $f(x)$ можно применить А.С.П., основанный на статистическом градиенте с парной пробой. Эта процедура описывается соотношением (1), в котором $S_n(x_n) =$

$$= (2c_n)^{-1}\left[f(x_n + c_n U_n) - f(x_n - c_n U_n) + \eta_n\right] U_n .$$

Здесь U_n, $n = 1,2,\ldots$ последовательность независимых случайных векторов, равномерно распределенных на единичной сфере с центром в нуле пространства R^m, совокупности случайных величин $\{\eta_n\}$ и случайных векторов $\{U_n\}$ взаимно независимы, η_n и c_n имеют тот же смысл, что и в А.К.В. Для этого алгоритма

$$S_n(x_n) = \frac{1}{m}\,\varphi(x_n) + \Delta S_n(x_n) + \frac{1}{2}\eta_n c_n^{-1},$$

$$\|\Delta S_n(x_n)\| \le \mathcal{L}(1+\gamma)^{-1} c_n^{\gamma}$$

и поэтому справедливы все (с точностью до мажорирующих констант) оценки, что и для метода К.В.

При одних и тех же исходных данных А.С.П. может обеспечивать такую же асимптотическую скорость, что и А.К.В., но при этом требует меньшего числа вычислений функции $f(x)$. Однако это утверждение не всегда справедливо на начальном этапе поиска экстремума.

4. Метод обобщенного градиента (М.О.Г.)

Пусть требуется найти точку минимума x^* функции $f(x)$, $x \in G$, где $f(x)$ выпуклая, непрерывная, в общем случае недифференцируемая функция, $f(x) \ge f^* > -\infty$. Обозначим через $Q(x)$ множество ее обобщенных градиентов (субдифференциал). Тогда для решения поставленной задачи широко используется алгоритм вида (1), где $S_n(x_n) = q(x_n)$, $q(x)$ – обобщенный градиент (субградиент) в точке x, $q(x) \in Q(x)$.

Потребуем, чтобы для $x \in G$ и всех $q(x) \in Q(x)$ выполнялись условия

А. $(q(x), x - x^*) \geq \theta \|x - x^*\|^{2p}$, $p > 0, \theta > 0$,

Б. $\|q(x)\| \leq \bar{\varphi}(\|x - x^*\|)$,

где $\bar{\varphi}(y)$, $y \geq 0$ является неубывающей функцией, $\bar{\varphi}(0) < \infty$. Тем самым будут выполняться условия 1 А (при $h_n \equiv 0$, $\nu_n \equiv 0$) и 2 Б, если в качестве $V(x)$ взять $V(x) = \|x - x^*\|^2$ $(L = 2, \mu = 1)$.

Пусть $\alpha_n = \alpha(n + n_0)^{-t}$, $0 < t \leq 1$, где n_0 выбирается согласно лемме 1. Тогда, если $\|x_1 - x^*\|^2 < R_0$, то будет справедливо неравенство (9), которое позволяет применить теоремы 1 ÷ 6 для оценки скорости сходимости метода обобщенного градиента. При этом следует положить $b = \alpha\theta$, $d = g(R_1)$, $s = 2t$.

Обсудим некоторые вытекающие отсюда результаты. Во-первых, асимптотические оценки, получающиеся из теорем 1 – 6, совпадают с асимптотическими результатами, установленными ранее только для $p = 1$, $t = 1$ [11]. При других значениях параметров они являются новыми. На основе найденных оценок можно осуществить выбор оптимальных (с точки зрения асимптотических оценок) параметров М.О.Г. Приведем результаты такого выбора.

1. Если $\rho = 1$, то $t^* = 1$ (t^* – оптимальное значение t), при этом $\lambda_n = 0(n^{-1})$, $(\lambda_n = \|x_n - x^*\|^2)$,

2. Если $\rho > 1$, то $t^* = (2\rho - 1)^{-1}$ и $\lambda_n = 0(n^{-\frac{\rho}{2\rho-1}})$,

3. Если $\rho < 1$, то $t^* = 1$ и тогда $\lambda_n = 0(n^{-\bar{S}})$,

$$\bar{S} = \min\left[\frac{1}{\rho}, 2\right].$$

Из последней оценки следует, что при $\rho_2 \leq 1/2$ $\lambda_n = 0(n^{-2})$, а при $\rho > 1/2$, $\|x_n - x^*\|^2 = 0(n^{-\frac{1}{\rho}})$, то есть скорость сходимости при $n \to \infty$ выше $0(n^{-2})$ не достигается. Эти результаты, также как и в случае стохастических алгоритмов, нельзя распространять на весь итерационный процесс, поскольку при достаточно больших λ_1 оценки имеют различные аналитические выражения при $n \leq \bar{x}$ и при $n > \bar{x}$. Причем на начальном этапе поиска $(n \leq \bar{x})$ скорость сходимости возрастает с уменьшением t, что подтверждает целесообразность использования неубывающего α_n в начале поиска и для метода обобщенного градиента.

Теоремы 1 ÷ 6, показывают также, что на скорость сходимости существенное влияние оказывают не только параметры алгоритма, но и структура (показатель ρ) класса минимизируемых функций. Причем с ростом ρ

(при неизменных других параметрах алгоритма и задачи) величина оценок скорости сходимости уменьшается.

В заключение особо подчеркнем, что случай $\rho < 1$ включает в себя класс функционалов, недифференцируемых в точке экстремума x^*, в частности, кусочно-линейного типа:

$$(g(x), x - x^*) \geq \theta \| x - x^* \| \quad (\rho = \tfrac{1}{2}).$$

ЛИТЕРАТУРА

1. Цыпкин Я.З. Адаптация и обучение в автоматических системах. – М.: Наука, 1968.
2. Катковник В.Я. Линейные оценки и стохастические задачи оптимизации. – М.: Наука, 1976.
3. Ермольев Ю.Н. Методы стохастического программирования. – М.: Наука, 1976.
4. Юдин Д.Б. Задачи и методы стохастического программирования. – М.: Советское радио, 1979.
5. Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. – Киев: Наукова думка, 1979.
6. Вазан М. Стохастическая аппроксимация. – М.: Мир, 1972.
7. Невельсон М.Б., Хасьминский Р.З. Стохастическая аппроксимация и рекуррентное оценивание. – М.: Наука, 1972.

8. Поляк Б.Т., Цыпкин Я.З. Псевдоградиентные алгоритмы адаптации и обучения. – Автоматика и телемеханика, 1973, № 3, с. 45–68.

9. Поляк Б.Т. Сходимость и скорость сходимости итеративных стохастических алгоритмов. 1. Общий случай.– Автоматика и телемеханика, 1976, № 12, с. 83–94.

10. Альбер Я.И., Шильман С.В. Рекуррентные числовые и дифференциальные неравенства. Ш. Препринт № 134, Горький: НИРФИ, 1980.

11. Bruck R.E. The iterative solution of the equation $y \in x+Tx$ for a monotone operator in Hilbert space.–Bull.Am.Math. Soc.1973,v.79,N6,p.1258–1261.

НЕАСИМПТОТИЧЕСКИЕ ОЦЕНКИ СКОРОСТИ СХОДИМОСТИ ПСЕВДОГРАДИЕНТНЫХ ИТЕРАТИВНЫХ АЛГОРИТМОВ

Яков Иосифович Альбер
Семен Вольфович Шильман

Подписано в печать 19.09.80 г. МЦ 00775. Формат 60 x 84 1/16
Бумага множительная марки А-1. Печать офсетная. Объем 2,1 п.л.
Тираж 1[illegible]0 экз. Заказ 2422. Бесплатно.

Горьковский ордена Трудового Красного Знамени научно-исследовательский радиофизический институт (НИРФИ), г.Горький, 603600, ГСП-51, ул. Лядова 25/14, т. 38-90-91, д. 5-09.

Отпечатано на ротапринте НИРФИ