Министерство высшего и среднего специального образования РСФСР

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

Препринт № 165

РЕКУРРЕНТНЫЕ НЕРАВЕНСТВА И АЛГОРИТМЫ СТОХАСТИЧЕСКОЙ АППРОКСИМАЦИИ

Я.И.Альбер

С.В.Шильман

Горький 1983

УДК 518.5.,517.9

Исследуются рекуррентные числовые неравенства

$$\lambda_{n+1} \leq (1+\beta_n)\lambda_n - \rho_n \Psi(\lambda_n) + \gamma_n, \quad n=1,2,\ldots,$$

где λ_n, ρ_n, β_n, γ_n - положительные (или неотрицательные) числовые последовательности, $\Psi(\lambda)$ - строго возрастающая функция при $\lambda \geq 0$, $\Psi(0) = 0$. Рассматривается случай с общим (не обязательно степенным) характером поведения $\Psi(\lambda)$ и допускается общий (не обязательно степенной) закон изменения ρ_n и γ_n. Находятся аналитические оценки λ_n, справедливые при всех $n \geq 1$. Полученные результаты применяются для изучения стохастических итеративных алгоритмов. В частности, определяются неасимптотические оценки скорости сходимости алгоритмов Роббинса-Монро, Кифера-Вольфовича и случайного поиска в задачах критериальной оптимизации, находятся асимптотически оптимальные параметры процедур. Исследование осуществляется в рамках достаточно широких классов минимизируемых функций, включающих выпуклые функции и функции со степенным вырождением.

Введение

Итеративные алгоритмы являются широко распространёнными средством численного решения разнообразных задач. Хорошо известны процедуры этого типа, применяющиеся для решения алгебраических и операторных уравнений, вариационных неравенств и задач оптимизации. Итеративные методы стали использоваться и для решения вероятностных задач и, в частности, для минимизации функций при наличии случайных погрешностей. Многие свойства этих вычислительных процедур (сходимость, устойчивость и точность итерационного процесса) исследуются с помощью числовых рекуррентных неравенств следующего вида:

$$\lambda_{n+1} \leq (1+\beta_n)\lambda_n - \rho_n \Psi(\lambda_n) + \gamma_n, \quad n=1,\dots, \tag{I}$$

где λ_n, ρ_n, β_n, γ_n – положительные (или неотрицательные) числовые последовательности, $\Psi(\lambda)$ – возрастающая функция от λ $(\lambda \geq 0)$, $\Psi(0)=0$.

Неравенства (I) послужили самостоятельным предметом исследования в работах авторов [I – 3]. По этим работам можно проследить постепенный прогресс как качества оценок λ_n $(n \geq I)$, так и техники изучения этих неравенств. В частности, в [3] впервые были найдены неасимптотические (справедливые с первого шага итерации) аналитические оценки значений λ_n, неулучшаемые по порядку при $n \to \infty$. При этом законы изменения ρ_n, γ_n задавались в виде степенных последовательностей $\gamma_n = \gamma n^{-s}$, $\rho_n = \rho n^{-t}$, $\rho > 0$, $\gamma > 0$, $0 < t \leq I$, $s > t$, и рассматривались функции $\Psi(\lambda)$ степенного типа: $\Psi(\lambda) = \lambda^p$, $p > 0$.

В настоящей работе приводятся новые результаты, развивающие и обобщающие работу [3]. Именно, изучается случай с общим (не обязательно степенным) характером $\Psi(\lambda)$ и допускается общий (не обязательно степенной) закон изменения ρ_n, γ_n. При этом находятся новые аналитические оценки λ_n, справедливые при всех $n \geq 2$.

Во второй части работы полученные результаты применяются для изучения стохастических итеративных алгоритмов. В частности, определяются неасимптотические оценки скорости сходимости ряда алгоритмов в задачах критериальной оптимизации, находятся асимптотически оптимальные параметры процедур. Ранее в работе авторов [4] были получены неасимптотические оценки скорости сходимости итеративных стохастических алгоритмов в терминах общих функций Ляпунова, градиент которых удовлетворяет условию Липшица-Гёльдера. При этом исследовались невырожденные задачи и задачи со степенным вырождением, а шаговый множитель алгоритмов α_n задавался равным $\alpha(n+n_0)^{-t}$, $0<t\leq 1$, $n_0\geq 0$, $\alpha>0$. В частности, в [4] отмечалось, что если функцию Ляпунова взять в виде $V(x)=f(x)-\min\limits_x f(x)$, то сформулированные оценки описывают скорость сходимости алгоритмов по функционалу (критерию) в среднем. В работе [5] подробно анализировались оценки среднего квадрата ошибки $M[\|x_n-x^*\|^2]$, где x^* - искомое решение, M - знак математического ожидания. Рассматривались алгоритмы Роббинса-Монро, Кифера-Вольфовица и алгоритмы случайного поиска, основанные на статистическом градиенте с парной пробой.

В настоящей работе аналогичные [5] результаты формулируются применительно к задачам критериальной оптимизации.

I. ИССЛЕДОВАНИЕ ЧИСЛОВЫХ РЕКУРРЕНТНЫХ НЕРАВЕНСТВ

Конечной целью данного раздела является получение неасимптотических оценок значений λ_n, удовлетворяющих неравенству (I).

Примем, что ρ_n, γ_n, $\gamma_n\rho_n^{-1}$ - невозрастающие последовательности и существуют при $x\geq I$ неотрицательные функции $\rho(x)$, $\gamma(x)$ - непрерывные невозрастающие, совпадающие при $x=n$ соответственно с ρ_n, γ_n. Обозначим через $r(x)$ какую-либо первообразную от $\rho(x)$, будем считать, что она определена при $x>I$. Кроме того, предположим, что $\sum\limits_{n=1}^{\infty}\rho_n=\infty$.

Введём функцию $\Phi(\lambda)=\int[1/\Psi(\lambda)]d\lambda$, пусть она определена при $\lambda>0$. Наряду с ней рассмотрим обратную функцию $\Phi^{-1}(z)$ на множестве Z значений $\Phi(\lambda)$. Будем считать, что $\Phi^{-1}(z)\geq 0$ на множестве $z\in Z$, при прочих z функцию $\Phi^{-1}(z)$ доопределим нулём.

Далее будут использоваться следующие обозначения:

$$v(x)=\Psi^{-1}\left[\frac{c_0\gamma(x-1)}{\rho(x-1)}+\gamma(x-1)\right], \quad C_1 = V(2), \qquad (I.I)$$

c_0 – свободный параметр, $c_0 > I$, $\Psi^{-1}(\cdot)$ – функция обратная для $\Psi(x)$,

$$w(x,C,\xi) = \Phi^{-1}\left\{\Phi(C)-a\left[r(C)-r(\xi)\right]\right\}, \qquad (I.2)$$

где $x \geqslant 2$, ξ – произвольное число из интервала $[2,\infty]$, $C \geqslant 0$,

$$a = c_0^{-1}(c_0-1),$$

$$u(x,C)=W(x,C,2)=\Phi^{-1}\left\{\Phi(C)-a\left[r(x)-r(2)\right]\right\}; \qquad (I.3)$$

$$A = w(2,\lambda_1,1)=\Phi^{-1}\left\{\Phi(\lambda_1)-a\left[r(2)-r(1)\right]\right\}. \qquad (I.4)$$

Л е м м а. Пусть последовательность неотрицательных чисел λ_n ($n=I, \ldots$) удовлетворяет неравенству

$$\lambda_{n+1} \leqslant (1+\beta_n)\lambda_n-\rho_n\Psi(\lambda_n), \qquad (I.5)$$

где $\rho_n>0$, $\beta_n \geqslant 0$, $\sum\limits_{k=1}^{\infty}\beta_k<\infty$, $\Psi(\lambda)$ – неубывающая функция от $\lambda \geqslant 0$, $\Psi(0)=0$. ~~(θ) = θ.~~ Тогда справедлива оценка

$$\lambda_n \leqslant \bar{C}\,\Phi^{-1}\left[\Phi(\lambda_1)-\bar{C}\sum_{k=1}^{n-1}\rho_k\right], \quad \bar{C} \geqslant \prod_{n=1}^{\infty}(1+\beta_n).$$

Если дополнительно функция $\Psi(\lambda)$ выпуклая при $\lambda \geqslant 0$, то

$$\lambda_n \leqslant \bar{C}\,\Phi^{-1}\left[\Phi(\lambda_1)-\sum_{k=1}^{n-1}(1+\beta_k)^{-1}\rho_k\right],$$

Доказательство. Заменой $\lambda_n = C_{n-1}\mu_n$, $C_n = \prod_1^n (1+\beta_k)$ из (I.5) получаем

$$\mu_{n+1} \le \mu_n - \rho_n C_n^{-1} \Psi\left[C_{n-1}\mu_n\right]. \qquad (1.6)$$

Далее из свойства вогнутости и дифференцируемости $\Phi(\lambda)$ при $\lambda > 0$ следует

$$\Phi(\mu_{n+1}) - \Phi(\mu_n) \le \frac{1}{\Psi(\mu_n)}(\mu_{n+1} - \mu_n) \le -\frac{\rho_n \Psi[C_n \mu_n]}{C_n \Psi(\mu_n)}.$$

Отсюда нетрудно получить все утверждения леммы.

Сформулируем свойство P. На интервале $[2,\infty)$ граф... функций $w(x, v(\xi), \xi)$ и $v(x)$ либо совпадают полнос..ю, либо пересекаются не более чем в двух точках (точки касан..я, за исключением быть может $x = 2$, не включаются в число точек пересе...ия).

Т е о р е м а I. Пусть $\lambda_n \ge 0$ удовлетворяют рекуррентному неравенству (I) при $\beta_n \equiv 0$ и

I) свойство P выполняется для $w(x, v(2), 2) = u(x, C_1)$ и $v(x)$,

2) при $x \to \infty$ $u(x, C_1) \ge v(x)$,

3) C_0 выбрано так, что при $x \to 2$ $u(x, C_1) \ge v(x)$.

Тогда при всех $n \ge 1$

$$\lambda_n \le u(n, C), \quad C = \max[A, C_1]. \qquad (1.7)$$

З а м е ч а н и е I. Вместо $u(x, C)$ и $v(x)$ можно использовать их мажоранты, но тогда к ним и должны относится требования теоремы.

З а м е ч а н и е 2. Вместо условий I), 2) можно потребовать, чтобы графики функций $w(x, v(\xi), \xi)$ пересекали $v(x)$ в точке $x = \xi$ при всех $x \in (2, \infty)$ так, чтобы $w(x, v(\xi), \xi) \le v(x)$ при $x < \xi$ и $w(x, v(\xi), \xi) \ge v(x)$ при $x > \xi$. Для этого достаточно, чтобы производные в точке $x = \xi$ удовлетворяли условию

$$a\Psi[v(\xi)]\rho(\xi) \le \left|\frac{dv}{dx}\right|_{x=\xi}, \qquad \xi \in [2, \infty].$$

Т е о р е м а 2. Пусть $\lambda_n \ge 0$ удовлетворяют рекуррентному неравенству (I) при $\beta_n \equiv 0$ и

а) свойство P выполняется для $v(x)$ и $w(x, v(\xi), \xi)$ при всех $\xi \in [2,\infty)$,

б) $u(x, C_1) \leq v(x)$ при $x \to \infty$.

Тогда имеют место следующие оценки:

I. если $A \leq C_1$, а $C_0 > 1$ выбрано так, что при $x \to 2$ $u(x,C_1) \leq \leq v(x)$, то при $n \geq 2$

$$\lambda_n \leq v(n); \tag{I.8}$$

2. во всех остальных случаях

$$\lambda_n \leq u(n,C) \text{ при } 1 \leq n \leq \bar{x}, \quad C = \max[A, C_1], \tag{I.9}$$

$$\lambda_n \leq v(n) \quad \text{при} \quad n > \bar{x}.$$

Здесь $\bar{x}$ единственный на интервале $(2, \infty)$ корень уравнения $u(x,C) = v(x)$.

Замечание 3. В теореме 2 функцию $v(x)$ можно заменить её мажорантой $\tilde{v}(x)$.

Замечание 4. Требование а) теоремы будет выполняться, если равенство абсолютных значений производных $\frac{d}{dx} w(x, v(\xi), \xi)$ и $\frac{dv}{dx}$ возможно лишь в одной точке ξ из интервала $[2,\infty)$, т.е. уравнение $a\Psi[v(\xi)]\rho(\xi) = \left|\frac{dv}{dx}(\xi)\right|$ имеет не более одного решения на $[2,\infty)$.

Замечание 5. Если неизвестны аналитические выражения для $\rho(x)$ или $r(x)$, то допустимо использование представления $r(n)$ в виде

$$r(n) - r(1) = \sum_{k=1}^{n-1} \rho(k).$$

Замечание 6. Если β_n отличны от нуля и $\sum_1^\infty \beta_n < \infty$, то все утверждения теорем I, 2 сохранятся при условии, что вместо ρ_n будут взяты $\rho_n(1+\beta_n)^{-1}$ (если $\Psi(\lambda)$ выпуклая функция) или $\rho_n \bar{C}^{-1}$ (во всех других случаях), а оценки λ_n будут увеличены в $\bar{C}$ раз, $\bar{C} \geq \geq \prod_1^\infty (1+\beta_n)$.

Указанное утверждение следует из следующих выкладок. Произведём в (I) замену $\lambda_n = \mu_n C_{n-1}$, $C_n = \prod_{k=1}^{n} (1+\beta_k)$. Тогда в силу того, что $C_n \geq 1$, получим соотношение

$$\mu_{n+1}\le \mu_n-\frac{\rho_n}{c_n}\Psi(\mu_n c_{n-1})+\frac{\gamma_n}{c_n}\le \mu_n-\rho_n\vartheta_n\Psi(\mu_n)+\gamma_n ,$$

где $\vartheta_n=(1+\beta_n)^{-1}$ в случае, когда $\Psi(\lambda)$ выпукла, и $\vartheta_n=\bar{C}^{-1}$ во всех других случаях.

Рассмотрим частный класс степенных "подпорок" $\Psi(\lambda)=\lambda^p, p>0$, и

$$\rho_n\ge\frac{b}{(1+\beta_n)^{-1}(n+n_0)^t},\quad \gamma_n=d(n+n_0)^{-s},$$

где $s>t$, $b>0$, $d\ge 0$, $n_0\ge 0$.

Оценки λ_n для этого случая получены в [3] и использованы в [4, 5]. Далее они приводятся в качестве следствий общих теорем I, 2, чем достигается существенное упрощение доказательств. Кроме того, в самих оценках производится некоторое уточнение по отношению к [3–5] мажорирующих констант и условий на C_0.

Следствие I. Пусть $\lambda_n\ge 0$, $n=1,\dots$ удовлетворяют неравенству $\lambda_{n+1}\le(1+\beta_n)\lambda_n-\rho_n\lambda_n+d(n+n_0)^{-s}$, $\rho_n(1+\beta_n)^{-1}\ge b(n+n_0)^{-1}$,

где $n_0\ge 0$, $n_0=\mathrm{const}$, и пусть

$$u(x,C)=C\left(\frac{2+n_0}{x+n_0}\right)^{ab},\quad \tilde{v}(x)=\left(c_0\frac{d}{b}+d\right)\left(\frac{1}{x+n_0-1}\right)^{s-1},$$

$$A=\lambda_1\left(\frac{n_0+1}{n_0+2}\right)^{ab},\quad C_1=\left(c_0\frac{d}{b}+d\right)\left(\frac{1}{n_0+1}\right)^{s-1}.$$

Тогда $\lim\limits_{n\to\infty}\lambda_n=0$ и справедливы следующие оценки.

I. Если $b>s-1$ и неопределённый параметр c_0 ($c_0>1$) выбран так, что $ab>s-1$, $a=(c_0-1)^{-1}c_0$, то

$$\lambda_n\le\bar{C}\,u(n,C),\quad \text{при } 2\le n\le\bar{x},$$

$$\lambda_n\le\bar{C}\,\tilde{v}(n),\quad \text{при } n>\bar{x}, \qquad (I.10)$$

где $\bar{C}=\max[A,C_1]$, $\bar{x}$ – единственный на интервале $(2,\infty)$ корень уравнения $u(x;C)=\tilde{v}(x)$.

Если при этом $\Lambda \le C_1$ и $ab \ge (s-1)(2+n_0)(1+n_0)^{-1}$, то оценка (I.10) выполняется при всех $n \ge 2$.

2. Если $b > s-1$, но $ab \le s-1$ или же $b \le s-1$, то $\lambda_n \le \bar{C} u(n,C)$, $n \ge 1$.

Следствие 2. Пусть $\lambda_n \ge 0$, $n = 1, \ldots$ удовлетворяют неравенству

$$\lambda_{n+1} \le (1+\beta_n)\lambda_n - \rho_n \lambda_n + d(n+n_0)^{-s}, \ \rho_n(1+\beta_n)^{-1} \ge b(n+n_0)^{-t},$$

где $0 < t < 1$, $s > t$, $n_0 = \mathrm{const}$, $n_0 \ge 0$, и пусть

$$\tilde{v}(x) = \left(c_0 \frac{d}{b} + d\right)\left(\frac{1}{x+n_0-1}\right)^{s-t}, \quad C_1 = \left(c_0 \frac{d}{b} + d\right)(n_0+1)^{-(s-t)},$$

$$u(x,C) = C \exp\left\{-\frac{ab}{1-t}\left[(x+n_0)^{1-t} - (2+n_0)^{1-t}\right]\right\},$$

$$\Lambda = \lambda_1 \exp\left\{-\frac{ab}{1-t}\left[(2+n_0)^{1-t} - (1+n_0)^{1-t}\right]\right\}.$$

Тогда $\lim\limits_{n\to\infty} \lambda_n = 0$ и имеют место оценки

$$\lambda_n \le \bar{C} u(n,C), \quad \text{при } 2 \le n \le \bar{x},$$

$$\lambda_n \le \bar{C} \tilde{v}(n), \quad \text{при } n > \bar{x}, \qquad \text{(I.II)}$$

где $C = \max[\Lambda, C_1]$, $\bar{x}$ – единственный на интервале $(2, \infty)$ корень уравнения $u(x,C) = \tilde{v}(x)$.

Если $\Lambda \le C_1$ и $c_0 > 1$ такое, что $ab \ge (s-t)(2+n_0)^t(1+n_0)^{-1}$, то оценка (I.II) справедлива при всех $n \ge 2$.

Следствие 3. Пусть $\lambda_n \ge 0$, $n = 1, \ldots$ удовлетворяют неравенству

$$\lambda_{n+1} \le (1+\beta_n)\lambda_n - \rho_n \lambda_n^p + d(n+n_0)^{-s}, \ \rho_n(1+\beta_n)^{-1} \ge b(n+n_0)^{-1},$$

где $p > 1$, $s > 1$, $n_0 = \mathrm{const}$, и пусть

$$u(x,C)=C\left[1+(p-1)C^{p-1}ab\ln\frac{x+n_0}{2+n_0}\right]^{-1/(p-1)},\quad C_1=\left[\left(\frac{d}{b}c_0\right)^{1/p}+d\right](n_0+1)^{-\frac{s-1}{p}},$$

$$A=\lambda_1\left[1+(p-1)\lambda_1^{p-1}ab\ln\frac{2+n_0}{1+n_0}\right]^{-1/(p-1)}.$$

Тогда $\lim\limits_{n\to\infty}\lambda_n=0$ и, если $c_0>1$ выбирается из условия

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

то при всех $n\geq 1$

$$\lambda_n\leq\bar{C}u(n,C),\quad C=\max\left[A,C_1\right].$$

Следствие 4. Пусть $\lambda_n\geq 0$, $n=1,\ldots$ удовлетворяют неравенству

$$\lambda_{n+1}\leq\lambda_n(1+\beta_n)-p_n\lambda_n^p+d(n+n_0)^{-s},\quad p_n(1+\beta_n)^{-1}\leq b(n+n_0)^{-t},$$

где $p>1$, $s>t$, $0<t<1$, $n_0=\mathrm{const}$, и пусть

$$\tilde{v}(x)=\left[\left(\frac{d}{b}c_0\right)^{1/p}+d\right]\left(\frac{1}{x+n_0-1}\right)^{\frac{s-t}{p}},\quad C_1=\left[\left(\frac{d}{b}c_0\right)^{1/p}+d\right](n_0+1)^{-\frac{s-t}{p}},$$

$$u(x,C)=C\left\{1+C^{p-1}ab\frac{p-1}{1-t}\left[(x+n_0)^{1-t}-(2+n_0)^{1-t}\right]\right\}^{-\frac{1}{p-1}},$$

$$A_1=\lambda_1\left\{1+\lambda_1^{p-1}ab\frac{p-1}{1-t}\left[(2+n_0)^{1-t}-(1+n_0)^{1-t}\right]\right\}^{-\frac{1}{p-1}}$$

Тогда $\lim\limits_{n\to\infty}\lambda_n=0$ и имеют место приводимые далее оценки.

I. Если $\dfrac{s-t}{p}<\dfrac{1-t}{p-1}$, то

$$\lambda_n \leq \bar{C}\, u(n,C) \quad \text{при } 2 \leq n \leq \bar{x},$$

$$\lambda_n \leq \bar{C}\, \tilde{v}(n) \quad \text{при} \quad n > \bar{x}, \tag{I.12}$$

где $C = \max[A, C_1]$, $\bar{x}$ – единственный на интервале $(2, \infty)$ корень уравнения $u(x, C) = \tilde{v}(x)$.

Если при этом $A \leq C_1$ и $c_0 > 1$ выбираются из условия

$$ab\left[\left(\frac{c_0 d}{b}\right)^{1/p} + d\right]^{p-1} \geq \frac{(2+n_0)^t}{(1+n_0)^{\text{æ}}}\frac{s-t}{p}, \quad \text{æ} = 1 - \frac{(s-t)(p-1)}{p},$$

то при всех $n \geq 2$ справедлива оценка (I.12)

2. Если $\dfrac{s-t}{p} \geq \dfrac{1-t}{p-1}$ и $c_0 > I$ выбираются из условия $ab\left[\left(c_0 \dfrac{d}{b}\right)^{1/p} + d\right]^{p-1} \leq \dfrac{s-t}{p}$, то при всех $n \geq I$ $\lambda_n \leq$ $\leq \bar{C}\, u(n, C)$.

Следствие 5. Пусть $\lambda_n \geq 0$, $n = I$ удовлетворяют неравенству

$$\lambda_{n+1} \leq \lambda_n - b(n+n_0)^{-1}\lambda_n^p + d(n+n_0)^{-s},$$

где $p < I$, $s > I$, $n_0 = \text{const}$ и пусть $\bar{s} = \min[s, (s-1)/p]$,

$$\tilde{v}(x) = \left[\left(\frac{d}{b}c_0\right)^{1/p} + d\right]\left(\frac{1}{x+n_0-1}\right)^{\bar{s}}, \quad C_1 = \left[\left(\frac{d}{b}c_0\right)^{1/p} + d\right](n_0+1)^{-\bar{s}},$$

$$u(x;C) = \left[C^{1-p} - (1-p)ab\ln\frac{x+n_0}{2+n_0}\right]^{\frac{1}{1-p}}, \quad A = \left[\lambda_1^{1-p} - (1-p)ab\ln\frac{2+n_0}{1+n_0}\right]^{\frac{1}{1-p}}.$$

Тогда $\lim\limits_{n\to\infty} \lambda_n = 0$ и

$$\lambda_n \leq u(n,C) \qquad 2 \leq n \leq \bar{x},$$

$$\lambda_n \leq \tilde{v}(x) \qquad n > \bar{x}, \tag{I.13}$$

где $C = \max[A, C_1]$, $\bar{x}$ – единственный на интервале $(2, \infty)$ корень уравнения $u(x,C)=\tilde{v}(x)$.

Если при этом $A \le C_1$ и c_0 $(c_0>1)$ выбирается так, что

$$\left[\left(c_0\frac{d}{b}\right)^{1/p}+d\right]^{p-1}ab \ge \frac{2+n_0}{(1+n_0)^{æ}}\bar{s}, \quad æ=1+\bar{s}(1-p),$$

то при всех $n \ge 2$ справедлива оценка (I.I3).

Следствие 6. Пусть $\lambda_n \ge 0$, $n=1,\ldots$ удовлетворяют неравенству

$$\lambda_{n+1} \le \lambda_n - b(n+n_0)^{-t}\lambda_n^p + d(n+n_0)^{-s},$$

где $p<1$, $s>t$, $0<t<1$, $n_0=\mathrm{const}$, и пусть

$$u(x,C)=\left\{C^{p-1}-\frac{1-p}{1-t}ab\left[(x+n_0)^{1-t}-(2+n_0)^{1-t}\right]\right\}^{\frac{1}{1-p}},$$

$$A=\left\{\lambda_1^{1-p}-\frac{1-p}{1-t}ab\left[(2+n_0)^{1-t}-(1+n_0)^{1-t}\right]\right\}^{\frac{1}{1-p}},$$ а $\tilde{v}(x), C_1, \bar{s}, æ$

определяются теми же выражениями, что и в следствии 5. Тогда $\lim_{n\to\infty}\lambda_n=0$ и

$$\lambda_n \le u(n,C) \quad \text{при} \quad 2 \le n \le \bar{x},$$
$$\lambda_n \le \tilde{v}(n) \quad \text{при} \quad n>\bar{x},$$

где $C=\max[A,C_1]$, $\bar{x}$ - единственный на $(2,\infty)$ корень уравнения $u(x,C)=\tilde{v}(x)$.

Если $A \le C_1$ и c_0 выбрано так, что

$$\left[\left(c_0\frac{d}{b}\right)^{1/p}+d\right]^{p-1}ab > \frac{(2+n_0)^t}{(1+n_0)^{æ}}\bar{s},$$

то при всех $n \ge 2$ $\lambda_n \le \tilde{v}(n)$.

Примеры применения приведённых утверждений см. в [4 - 6].

2. О НЕАСИМПТОТИЧЕСКИХ ОЦЕНКАХ СКОРОСТИ СХОДИМОСТИ ИТЕРАТИВНЫХ СТОХАСТИЧЕСКИХ АЛГОРИТМОВ

Рассмотрим стохастические алгоритмы вида

$$x_{n+1}=\pi_G\left[x_n-\alpha_n S_n(x_n)\right], \quad n = 1,\ldots \qquad (2.1)$$

где x_n - вектор со значениями из R^m, $S_n(x_n)$ - вектор случайных наблюдений на n-й итерации в точке x_n, $\alpha_n \geq 0$ параметр шага. Через $\pi_G(\cdot)$ обозначен оператор проектирования на выпуклое, замкнутое множество $G \subseteq R^m$. Если $G = R^m$, то $\pi_G = E$, где E - единичная матрица.

Формула (2.1) описывает многие известные процедуры решения систем алгебраических уравнений, условной и безусловной оптимизации в условиях неполной вероятностной информации.

Сходимость и скорость сходимости процесса (2.1) можно изучать, "отслеживая" значения $\lambda_n = M\left[V(x_n)\right]$. Если G не совпадает с R^m, то в качестве $V(x)$ нужно брать функцию $\rho(x, X^*)$-квадрат расстояния от точки x до множества X^* всех решений задачи, которое будет считать выпуклым и замкнутым. При $G = R^m$, наряду с $\rho^2(x, X^*)$, можно использовать и другие дифференцируемые функции $V(x)$, лишь бы они удовлетворяли

1) условию $\inf V(x) = 0$,

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

$$\|\nabla V(x) - \nabla V(y)\| \leq L \|x-y\|^{\mu}, \qquad (2.2)$$

где $L = const$, $\mu = const$, $0 < \mu \leq 1$.

Классы задач и способы их решения определяются видом вектора S_n и его свойствами. Будем считать, что выполняются два основных требования.

I. Метод (2.1) является равномерно псевдоградиентным по отношению к выбранной функции Ляпунова $V(x)$ в следующем смысле:

$$(\nabla V(x_n), M\left[S_n | x_n, \ldots, x_1\right]) \geq \theta_n \Psi\left[V(x_n)\right] - h_n - \nu_n V(x_n),$$

где $\theta_n > 0$, $h_n \geq 0$, $\nu_n \geq 0$.

Функция $\Psi(V)$ характеризует в данном случае меру вырожденности задачи. Два последних слагаемых приведенного неравенства возникают при изучении поисковых методов (например, метода Кифера-Вольфовица).

II. $M\left[\| S_n(x_n)\|^{1+\mu} / x_n, \ldots, x_1\right] \leq \sigma_n + \tau_n V(x_n)$, где $\sigma_n \geq 0$, $\tau_n \geq 0$.

Предположения I, II должны быть взаимно согласованы.

Способами, описанными в работе авторов [4], можно вывести (с учётом требований I, II) для λ_n числовое рекуррентное неравенство

$$\lambda_{n+1} \leq (1+\beta_n)\lambda_n - \rho_n \Psi(\lambda_n) + \gamma_n,$$

где β_n, ρ_n, γ_n выражаются через θ_n, h_n, ϑ_n, σ_n, τ_n.

Далее для получения оценок λ_n можно использовать теоремы I, 2 и их следствия. Частный класс стохастических задач будет описан в следующем разделе.

3. ОПТИМАЛЬНЫЕ ПАРАМЕТРЫ И НЕАСИМПТОТИЧЕСКИЕ ОЦЕНКИ СКОРОСТИ СХОДИМОСТИ СТОХАСТИЧЕСКИХ АЛГОРИТМОВ В ЗАДАЧАХ КРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ

Пусть требуется найти точную нижнюю грань f^* дифференцируемой функции $f(x)$

$$\inf_{x \in R^l} f(x) = f^*, \quad f^* > -\infty. \tag{3.I}$$

Для этого применим алгоритм типа (2.I) при $\pi_G = E$.

Будем далее считать, что

I. $f(x) \in C^{1,\mu}$, то есть $f'(x)$ (градиент функции $f(x)$), удовлетворяет условию (2.2);

IIА. $\| f'(x)\|^2 \geq \theta (f(x) - f^*)^m$, $I \leq m \leq 2$, $\theta > 0$ или

IIБ. $M[\| f'(x_n)\|^2] \geq \theta \left(M[f(x_n) - f^*]\right)^m$.

Параметры m и μ должны быть согласованы.

Условие I описывает гладкость функции, условия II – её структуру. В общем случае они не требуют выпуклости функции $f(x)$, и она не обязана иметь единственную точку минимума. Заметим, что если $f(x) \in C^{1,1}$ и является сильно выпуклой, то при $m = I$ и всех $x \in R^l$ справедливо соотношение IIА.

а) Пусть последовательность x_n генерируется алгоритмом Роббинса-Монро (а. Р.М.), то есть в (2.I) примем, что

$$S_n(x_n) = f'(x_n) + \eta_n, \quad \pi_G = E. \tag{3.2}$$

Здесь η_n, $n = 1,\dots$ совокупность независимых случайных векторов с $M[\eta_n] = 0$ и $M[\|\eta_n\|^2] \le \sigma^2 < \infty$. Тогда справедлива лемма.

Лемма. Предположим, что $f(x) \in C^{1,\mu}$ и является выпуклой, параметр α_n в алгоритме (2.1), (3.2) удовлетворяет условию $\sum_{n=1}^{\infty}\alpha_n^2<\infty$.

Тогда для x_n справедливо соотношение

$$M\left[\|f'(x_n)\|^2\right] \ge \theta\left(M\left[f(x_n)-f^*\right]\right)^2.$$

С учётом условий I, IIБ при $\alpha_n=\alpha(n+n_0)^{-t}$, $0<t\le 1$, $\alpha>0$, $n_0\ge 0$ для алгоритма (2.1), (3.2) можно получить неравенство

$$\lambda_n \le \lambda_n - b(n+n_0)^{-t}\lambda_n + d(n+n_0)^{-s}, \tag{3.3}$$

где $s=(1+\mu)t$.

Применяя следствия 1-4 теорем 1, 2, получаем следующие результаты.

1. Пусть $m = 1$, $\mu = 1$, при этом $s = 2t$. Этот случай охватывает класс сильно выпуклых функций. Асимптотически максимальная скорость сходимости может быть получена при $b > s-1 = 1$ и $t = t^* = 1$. Тогда $\alpha_n=\alpha(n+n_0)^{-1}$ и $M[f(x_n)-f^*]\le O(n^{-1})$.

Как известно, аналогичные оценки имеют место для а Р.М. при тех же условиях выпуклости и для среднего квадрата ошибки $M\left[\|x_n-x^*\|^2\right]$ (см. [5]).

На начальном этапе поиска и при достаточно больших n оценки могут иметь разное аналитическое описание. Вследствие этого выводы, касающиеся начала поиска, существенно другие. В частности, при достаточно больших значениях $M\left[f(x_1)-f^*\right]$ получаемая скорость сходимости тем выше, чем значения t ближе к нулю. Поскольку $t = 0$ соответствует постоянному параметру шага, то, как и в [4, 5], отсюда вытекает целесообразность использования неубывающего параметра α_n на начальном этапе поиска. Ещё раз подчеркнём, что этот вывод справедлив в тех случаях, когда неопределённость информации о значении f^* превосходит неопределённость помех.

2. Пусть $m = 2$, $0<\mu\le 1$. Этот случай включает класс выпуклых функций. Из следствий 3, 4 следует, что при $t<1$ можно получить более высокую асимптотическую скорость сходимости, чем при $t = 1$. Наибольшее её значение достигается при $t = t^* = 2/(2+\mu)$, при этом

$$M\left[f(x_n)-f^*\right]\le O\left[\left(\frac{1}{n}\right)^{\mu/(2+\mu)}\right],$$

отсюда, в частности, следует, что порядок оценки скорости сходимости не превосходит I/3. Последний получается при $\mu = 1$, $t = t^* = 2/3$. С уменьшением μ асимптотически оптимальная скорость сходимости падает, а $t^* \to 1$.

Выводы о начальном участке поиска в случае 2 и далее в случае 3 в рамках допустимых значений t сохраняются теми же, что и в случае I.

3. Теперь пусть $1 < m < 2$, $0 < \mu \le 1$. Рассмотрим асимптотические оценки. Наибольшая скорость сходимости получается при $t = t^* = m/[m(1+\mu) - \mu]$, при этом

$$M[f(x_n) - f^*] \le O\left\{ n^{-\mu/[m(1+\mu)-\mu]} \right\}.$$

При $\mu = 1$ отсюда следует $t^* = m/(2m-1)$, то есть $\alpha_n = \alpha(n+n_0)^{-\frac{m}{2m-1}}$ и

$$M[f(x_n) - f^*] \le O\left\{ n^{-1/(2m-1)} \right\}.$$

Полученные оценки выявляют характер зависимости скорости сходимости от показателя гладкости μ и параметра m, описывающего структуру функции. При этом оценки случаев I, 2 монотонно переходят друг в друга.

б) Алгоритм Кифера-Вольфовица (а. К.В.).

Для решения задачи оптимизации (3.I) применим алгоритм (2.I) при $\pi_G = E$ и векторе $S_n(x_n)$, имеющем в каждой точке компоненты

$$\frac{1}{2c_n}\left[f(x_n + c_n e_i) - f(x_n - c_n e_i) + \eta_{ni} \right], \quad i = \overline{1, l}.$$

Здесь $\{e_i\}$ – ортонормированный базис в R^l, $\eta_n = \{\eta_{n,i}\}$ – вектор, характеризующий случайную погрешность вычисления разностей функции $f(x)$. Будем считать, что совокупность $\{\eta_n\}$ состоит из независимых при разных n случайных векторов, $M[\eta_n] = 0$, $M[\|\eta_n\|^2] \le \sigma^2$, последовательность $c_n \ge 0$ задана в виде $c_n = c(n+n_0)^{-r}$, $c > 0$, $n_0 \ge 0$, $0 < r \le \frac{\mu}{1+\mu} t$. Предполагается также, что выполняются условия I, II разд. 3. Тогда для $\lambda_n = M[f(x_n) - f^*]$ можно получить неравенство вида (3.3), в котором $S = \min[t + 2\mu r, (1+\mu)(t-r)]$. Отсюда заключаем, что все утверждения следствий I–4 применимы и для алгоритма К.В. с соответетствующими изменениями значений b, d и S.

Рассмотрим вопрос о выборе параметров t и r с точки зрения получения максимальной асимптотической скорости сходимости. Анализ показывает, что последняя достигается при $r = r^* = \mu t/(1+3\mu)$, $t = t^* = m(1+3\mu)(m+3\mu m+2m\mu^2-2\mu^2)^{-1}$. В этом случае имеем

$$M\left[f(x_n)-f^*\right]=0\left[\left(\frac{1}{n}\right)^{\frac{2\mu^2}{m+3\mu m+2m\mu^2-2\mu^2}}\right].$$

В частности, при $\mu = 1$ получаем следующие далее оценки.

Пусть $1 \leq m \leq 2$, тогда $t^* = 2m/(3m-1)$, $r^* = m/2(3m-1)$ и

$$M\left[f(x_n)-f^*\right]\leq 0\left[\left(\frac{1}{n}\right)^{1/(3m-1)}\right].$$

Отсюда следует, что если $\mu = 1$ и $m = 2$, то $t^* = 4/5$, $r^* = 1/5$ и

$$M\left[f(x_n)-f^*\right]\leq 0\left(n^{-1/5}\right).$$

Если же $\mu = 1$, а $m = 1$, то $t^* = 1$, $r^* = 1/4$ и

$$M\left[f(x_n)-f^*\right]\leq 0\left(n^{-1/2}\right).$$

Последнее соотношение совпадает с известным результатом для сильно выпуклых функций, справедливым в отношении среднего квадрата ошибки $M\left[\|x_n-x^*\|^2\right]$.

Аналогичные выводы могут быть сделаны для алгоритма случайного поиска, основанного на статистическом градиенте с парными пробами.

Литература

1. Альбер Я.И. Рекуррентные числовые и дифференциальные неравенства. I. – В сб.: Численные методы анализа (прикл. математика) СО АН СССР, Сибирский энерг. ин-т, 1980, с. 30 – 50.
2. Альбер Я.И. Рекуррентные числовые и дифференциальные неравенства. II. – Препринт НИРФИ, № 130, 1979.
3. Альбер Я.И., Шильман С.В. Рекуррентные числовые и дифференциаль – ные неравенства III. – Препринт НИРФИ, № 134, 1980.
4. Альбер Я.И., Шильман С.В. Неасимптотические оценки скорости сходимости стохастических итеративных алгоритмов. – Автоматика и телемеханика, 1981, № 1, с. 41 – 54.
5. Альбер Я.И., Шильман С.В. О методах стохастической аппроксимации. – ДАН СССР, 1981, т. 255, № 2, с. 265 –269.
6. Альбер Я.И., Шильман С.В. Метод обобщенного градиента: сходимость, устойчивость и оценки погрешности. – Ж.вычисл. матем. и матем.физ. 1982, т. 22, № 4, с. 814 – 823.

Дата поступления статьи
2 марта 1983г.

Яков Иосифович Альбер

Семен Вольфович Шильман

РЕКУРРЕНТНЫЕ НЕРАВЕНСТВА И АЛГОРИТМЫ СТОХАСТИЧЕСКОЙ АППРОКСИМАЦИИ

Подписано в печать 25.05.83. МЦ 00366. Формат 60 х 84 1/16. Бумага множительная. Печать офсетная. Объем 1,06 усл. печ. л. Тираж 120.

Заказ 2954. Бесплатно.

Отпечатано на ротапринте Горьковского научно-исследовательского радиофизического института, 603600, Горький ГСП-51, ул. Лядова 25/14, т. 36-90-91.