Сравнение различных методов скорейшего спуска на примере квадратичной функции

А.В. Гаглоев

ФУПМ, МФТИ

ТМШ, 2014

Содержание

Введение и Постановка Задачи

Градиентный Метод

Случайный Спуск

Покоординатный Спуск

Примеры Применения

Скорейший Спуск

Задача: Пусть $\mathrm{A} : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{n}}, \mathrm{A} \succeq 0,$ - самосопряжённое линейное отображение с собственными значениями $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$
$\mu \stackrel{\text{def}}{=} \|A\| \|A^{-1}\| = \lambda_n / \lambda_1$

$$f(x) = \frac{1}{2} \langle \mathrm{A} x, x \rangle - \langle b, x \rangle,$$

$x, b \in \mathbb{R}^n$
Требуется найти $x^* = \operatorname*{argmin}_{x \in \mathbb{R}^n} f(x)$
Наискорейший спуск: $\{x^k\}_{k \in \mathbb{Z}_+}$ $(f(x^{k+1}) \leqslant f(x^k))$

$$x^{k+1} = x^k + h^k v^k,$$

$v^k \in \mathbb{R}^n$, $h^k = \operatorname*{argmin}_{h} f(x^k - h f'(x^k))$

Градиентный Наискорейший Спуск

$$v^k = f'(x^k)$$

$$f(x^{k+1}) = f(x^k) - \frac{\|f'(x^k)\|^2}{2f(f'(x^k))} f'(x^k)$$

Канторович Л.В. (1947 г.):

$$\frac{f(x^{k+1})}{f(x^k)} \leqslant 1 - \frac{4}{\left(\sqrt{\mu} + \frac{1}{\sqrt{\mu}}\right)^2}$$

Случайный Скорейший Спуск

$v^k \sim U(S^{n-1}(0,1))$, где $S^{n-1}(a,r) = \{x \in \mathbb{R}^n | \|x-a\| = r\}$

$$q_n(x^k) \overset{\text{def}}{=} \frac{\mathsf{E}[f(x^{k+1})|x^k]}{f(x^k)} = 1 - \frac{1}{f(x^k)} \sum_{i=1}^{n} \lambda_i x_i^2 \mathsf{E} \frac{\lambda_i v_i^2}{\sum\limits_{j=1}^{n} \lambda_j v_j^2}$$

Основная оценка скорости сходимости $(x^k \overset{\text{def}}{=} x, v^k \overset{\text{def}}{=} v)$:

Theorem

$$1 - \mathsf{E}\frac{\lambda_n v_n^2}{\langle \mathrm{A}v, v\rangle} \leqslant q_n(x) \leqslant 1 - \mathsf{E}\frac{\lambda_1 v_1^2}{\langle \mathrm{A}v, v\rangle}$$

Оценки Скорости Сходимости

Corollary

$$1 - \mathsf{E}\frac{\mu v_n^2}{\sum\limits_{i=1}^{n-1} v_i^2 + \mu v_n^2} \leqslant q_n(x) \leqslant 1 - \mathsf{E}\frac{v_1^2}{v_1^2 + \mu \sum\limits_{i=2}^{n} v_i^2}$$

Theorem

При $n \to \infty$ справедливо

$$q_n(x) \leqslant 1 - \frac{1}{n\mu} + O\left(\frac{1}{n^{\frac{3}{2}}}\right)$$

Оценки Скорости Сходимости

Theorem

При $\mu \to \infty$

$$q_2(x) \leqslant 1 - \frac{1}{\sqrt{\mu}} + O\left(\frac{1}{\mu}\right),$$

$$q_3(x) \leqslant 1 - \frac{\ln\sqrt{\mu}}{\mu} + O\left(\frac{1}{\mu^{\frac{3}{2}}}\right),$$

для $4 \leqslant n$

$$q_n(x) \leqslant 1 - \frac{k_n}{\mu} + O\left(\frac{1}{\mu^2}\right), \qquad k_n = \frac{\sqrt{\pi}}{4}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

Точная Оценка

$$q(x^k) = \frac{\mathrm{E}(f(x^{k+1}|x^k)}{f(x^k)} \leq 1 - \mathrm{E}\frac{\lambda_1 v_1^2}{\langle \mathrm{A} v, v\rangle}$$

$$\mathrm{E}\frac{\lambda_1 v_1^2}{\langle \mathrm{A} v, v\rangle} = \frac{\lambda_1}{\lambda_n}\mathrm{E}\frac{v_1^2}{1 - \sum\limits_{i=1}^{n-1}\left(1 - \frac{\lambda_i}{\lambda_n}\right)v_i^2} =$$

$$= \frac{\lambda_1}{\lambda_n}\sum_{k=0}^{\infty}\mathrm{E}v_1^2\left(\sum_{i=1}^{n-1}\left(1 - \frac{\lambda_i}{\lambda_n}\right)v_i^2\right)^k$$

$$\mathrm{E}\prod_{i=1}^{n} v_i^{2\alpha_i} = \frac{\prod\limits_{\substack{\alpha_i \neq 0 \\ 1 \leqslant i \leqslant n}} (2\alpha_i - 1)!!}{\prod\limits_{i=0}^{\alpha_1 + \dots + \alpha_n - 1} (2i + n)}, \qquad \alpha_i \in \mathbb{Z}_+$$

Наглядная Оценка

Неравенство Коши-Буняковского

$$|\langle \xi, \eta \rangle| \leq \|\xi\| \|\eta\|, \qquad \langle \xi, \eta \rangle = \mathsf{E}\xi\eta$$

$$\xi = \frac{v_1}{\langle Av, v \rangle^{1/2}}, \quad \eta = v_1 \langle Av, v \rangle^{1/2}$$

$$\left(\mathsf{E}v_1^2\right)^2 \leq \mathsf{E}\frac{v_1^2}{\langle Av, v \rangle} \mathsf{E}v_1^2 \langle Av, v \rangle,$$

$$q_n(x) \leq 1 - \mathsf{E}\frac{\lambda_1 v_1^2}{\langle Av, v \rangle} \leq 1 - \frac{\lambda_1}{Sp(A)} \frac{1 + \frac{2}{n}}{1 + \frac{2\lambda_1}{Sp(A)}}$$

Покоординатный Скорейший Спуск

$$x^{k+1} = x^k - \frac{\langle a_{i_k}, x^k \rangle}{a_{i_k i_k}} e_{i_k}$$

$$Pr\{i_k = j\} = \frac{a_{jj}}{Sp(A)}$$

где e_i — i-ый координатный вектор

$$\frac{\mathrm{E}[f(x^{k+1})|x^k]}{f(x^k)} \leqslant 1 - \frac{\lambda_1}{Sp(A)}$$

Сложность шага итерации — $O(n)$

Выбор Номеров Координат

Пусть $n = 2^m, m \in \mathbb{N}$. Определим $m + 1$ вектор $P_k \in \mathbb{R}^{2^{m-k}}$, $k = 0, \dots, m$

$$P_0^{(i)} = a_{ii}, \; i = 1, \dots, n.$$

$$P_k^{(i)} = P_{k-1}^{(2i)} + P_{k-1}^{(2i-1)}, i = 1, \dots, 2^{m-k}, \; k = 1, \dots, m$$

Алгоритм:
1. Выбор $i = 1$ в P_m
2. Для всех $k = m, \dots, 1$
Если выбран номер i на шаге k, на шаге $k - 1$ номера $2i$ и $2i - 1$ выбираются с вероятностями $\dfrac{P_{k-1}^{(2i)}}{P_k^{(i)}}$ и $\dfrac{P_{k-1}^{(2i-1)}}{P_k^{(i)}}$
Сложность выбора номера координаты — $O(ln\ n)$

Пример 1

Задача об аппроксимации функции из $L^2[0,1]$ полиномом степени не выше $n-1$

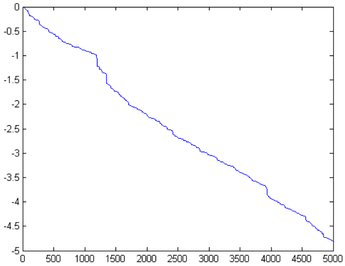
$$f(x) = \frac{1}{2}\langle \mathrm{H_n} x, x\rangle - \langle b, x\rangle$$

$\mathrm{H_n}$, матрица Грама системы функций $\{x^k\}_{k=0}^{n-1}$, - матрица Гильберта размерности $n \times n$ $\left(h_{i,j}^n = \frac{1}{i+j-1}\right)$

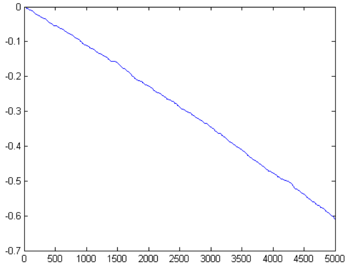
$$Cond(\mathrm{H_n}) = \mathrm{O}\left(\frac{(1+\sqrt{2})^{4\mathrm{n}}}{\sqrt{\mathrm{n}}}\right)$$ при $n \to \infty$.

Пример 1

$$n = 4, \quad Cond(H_n) = 1.5\ 10^4$$



RD



RCDM

Рис.: Зависимость $log(f(x^k))$ от k.

Оценка Эффективности Методов

$1 - q_n(x)$

Теоретические результаты

n	$Cond(H_n)$	RD	GM	RCDM
2	$1.928e+01$	$1.855e-01$	$1.875e-01$	$4.931e-02$
3	$5.240e+02$	$1.347e-02$	$7.604e-03$	$1.752e-03$
4	$1.551e+04$	$4.906e-04$	$2.578e-04$	$5.769e-05$
5	$4.766e+05$	$1.376e-05$	$8.393e-06$	$1.840e-06$

Наблюдаемые результаты

n	$Cond(H_n)$	RD	GM	RCDM
2	$1.928e+01$	$4.746e-01$	$1.875e-01$	$1.340e-01$
3	$5.240e+02$	$3.513e-02$	$7.604e-03$	$3.510e-03$
4	$1.551e+04$	$9.509e-04$	$2.578e-04$	$1.115e-04$
5	$4.766e+05$	$2.642e-05$	$8.393e-06$	$3.710e-06$

Пример 2

$$f(x) = \frac{1}{2}\langle A_n x, x\rangle - \langle b, x\rangle$$

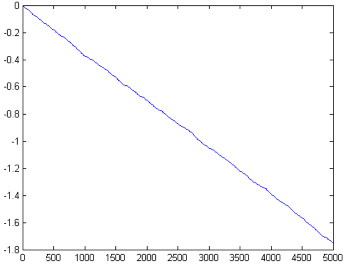
$$A_n = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 2 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 2 & 1 & 0 & \dots & 0 \\ & & & \dots & & & \\ 0 & \dots & 0 & 0 & 1 & 2 & 1 \\ 0 & \dots & 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\lambda_k^n = 2\left(1 - \cos\frac{\pi k}{n+1}\right), \quad k = 1\dots n$$

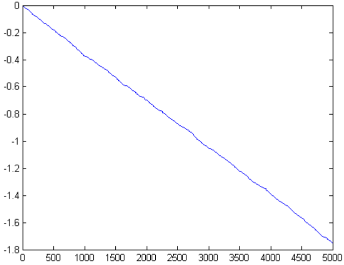
$$Cond(\mathrm{A_n}) = \frac{1 + \cos\frac{\pi}{\mathrm{n+1}}}{1 - \cos\frac{\pi}{\mathrm{n+1}}} = \mathrm{O}(\mathrm{n}^2) \text{ при } n \to \infty.$$

Пример 2

$$n = 40, \quad Cond(A_n) = 3.9\ 10^2$$



RD



RCDM

Рис.: Зависимость $log(f(x^k))$ от k.

Оценка Эффективности Методов

$1 - q_n(x)$

Теоретические результаты

n	$Cond(A_n)$	RD	GM	RCDM
2	3	$3.660e-01$	0.750	$2.500e-01$
10	48.374	$5.430e-03$	$7.937e-02$	$4.051e-03$
20	178.064	$6.476e-04$	$2.221e-02$	$5.585e-04$
30	388.812	$2.008e-04$	$1.023e-02$	$1.710e-04$

Наблюдаемые результаты

n	$Cond(A_n)$	RD	GM	RCDM
2	3	$7.389e-01$	0.750	$4.918e-01$
10	48.374	$1.073e-02$	$7.937e-02$	$8.235e-03$
20	178.064	$1.285e-03$	$2.221e-02$	$1.127e-03$
30	388.812	$3.807e-04$	$1.023e-02$	$3.336e-04$

Выводы

Сравниваемые методы обладают своими областями превосходства

- Случайный наискорейший спуск — большие числа обусловленности, $Cond(A)$, малые размерности, n
- Покоординатный случайный спуск — большие размерности
- Градиентный спуск — промежуточные значения $Cond(A)$ и n

Список литературы I

- Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1983.
- NesterovYu.E. Efficiency of coordinate descent methods on huge-scale optimization problems. — SIAM Journal on Optimization, 2012.— vol. 22.— p. 341-362.
- Levethal D., Lewis A.S. Randomized Methods for Linear Constraints: Convergence Rates and Conditioning.— Math. of Oper. Res., 2010, - vol. 35, No. 3, p. 641-654.
- Николаев Е.Г. Случайный и градиентный выбор направления спуска в задаче поиска минимума функции многих переменных.