ПРО ШВИДКІСТЬ ЗБІЖНОСТІ СТОХАСТИЧНОГО КІНЦЕВО-РІЗНИЦЕВОГО МЕТОДУ ОПТИМІЗАЦІЇ

В.І. НОРКІН, А.Ю. КОЗИРЕВ
Національний технічний університет України
«КПІ ім. І. Сикорського»
vladimir.norkin@gmail.com

Анотація. *Досліджена швидкість збіжності стохастичного кінцево-різницевого методу оптимізації опуклих ліпшицевих функцій на опуклій обмеженій множині. В методі використовуються стохастичні центральні кінцеві-різніцеві оцінки градієнтів згладжених функцій. Показано, що оцінки швидкості збіжності всередньому обратно пропорційні квадртному корню із числа ітерацій.*

Ключові слова: *опукла оптимізація, метод згладжування, стохастичний кінцево-різницевий метод, швидкість збіжності.*

Для розвязання негладкої оптимізаційної задачі $\left[\min_{x\in X} F(x)\right]$ розглянемо так звані згладжені (або усереднені) функції наступного вигляду [1 - 3]:

$$F_h(x) = \frac{1}{v_n h^n} \int_{B_h(x)} F(y)dy\,, \quad B_h(x) = \left\{ y \in \mathrm{R}^n \,\middle|\, \|y - x\|_2 \le h \right\},$$

де $\|\cdot\|_2$ – евклідова норма в просторі R^n; V_n та S_n - одинична куля та її поверхня у R^n; s_n та $v_n = ns_n$ – площа поверхні та об'єм одиничної кулі у R^n; $N(y)$ – зовнішня одинична нормаль до одиничної кулі у R^n в точці y; $h > 0$ - параметр згладжування.

Якщо функція $F(\cdot)$ ліпшицева з субдифференціалом $\partial F(\cdot)$, то її градієнт можна представити у вигляді

$$\nabla F_h(x) = \frac{1}{v_n h^n} \int_{B_h(x)} \partial F(y)dy\,.$$

Градієнт $F_h(x)$ обчислюється поверхневими інтегралами

$$\nabla F_h(x) = \frac{n}{hs_n}\int_{S_n}\big(F(x+y)-F(x)\big)N(y)dS =$$

$$= \frac{n}{2hs_n}\int_{S_n}\big(F(x+hy)-F(x-hy)\big)N(y)dS.$$

Якщо позначити $\tilde{y}$ випадковий вектор, рівномірно розподілений на одиничній сфері $S_1(0)=\left\{y \in \mathrm{R}^n : \|y\|=1\right\}$, то градієнт $\nabla F_h(x)$ можна представити у вигляді математичних очікувань [1 - 3]:

$$\nabla F_h(x) = n \cdot \mathrm{E}_{\tilde{y}}\frac{1}{h}\big(F(x+h\cdot\tilde{y})-F(x)\big)\cdot\tilde{y} =$$

$$= n \cdot \mathrm{E}_{\tilde{y}}\frac{1}{2h}\big(F(x+h\cdot\tilde{y})-F(x-h\cdot\tilde{y})\big)\cdot\tilde{y}.$$

В роботі [1] на основі подання градієнтів згладжених функцій у вигляді математичних очікувань центральних і симетричних різниць значень функції за випадковими напрямками побудовані і обґрунтовані стохастичні методи оптимізації негладких ліпшицевих функцій.

Теорема 1 [1]. Нехай $F(x)$ - ліпшицева функція в околі кулі $\left\{x \in \mathrm{R}^n : \|x\| \le r\right\}$ з субградієнтними множинами Кларка $\partial F(x)$; множина $F^* = \left\{x : \|x\| \le r, 0 \in \partial F(x)\right\}$ ніде не щільна, наприклад, скінчена; відома точка $\bar{x}_0$, $\|\bar{x}_0\| < r$, така, що $F(\bar{x}_0) < \min_{\|x\|=r} F(x$

Розглянемо ітераційний процес:

$$x_{k+1} = \begin{cases} x_k - \rho_k\eta_k, & \|x_k\| \le r, \\ x_0, & \|x_k\| > r; \end{cases} \quad x_0 = \bar{x}_0, \qquad k = 0,1,\ldots,$$

де $\eta_k = \dfrac{1}{2h_k}\big(F(x_k + h_k\tilde{y}_k) - F(x_k + h_k\tilde{y}_k)\big)$; випадкові вектори $\left\{\tilde{y}_k\right\}$ незалежні і рівномірно розподілені на поверхні одиничної сфери з центром у початку координат.

Нехай виконані умови: $\rho_k > 0$, $h_k \ge h_{k+1} > 0$;

$$\lim_k h_k = \lim_k \rho_k/h_k = \lim_k\left(h_k - h_{k+1}\right)/\rho_k = 0;$$

$$\sum_{k=0}^{\infty} \rho_k^2 < \infty , \qquad \sum_{k=0}^{\infty} \rho_k = +\infty .$$

Тоді всі граничні точки випадкової послідовності $\{x_k\}$ майже напевно належать множині $X^* = \{x : \|x\| \le r, 0 \in \partial F(x)\}$ і числова послідовність $\{F(x_k)\}$ майже напевно має границю.

Має місце наступний результат про швидкість збіжності стохастичного кінцево-різницевого методу з усередненням траєкторії на опуклих функціях.

Теорема 2. Припустимо, що функція $F(\cdot)$ опукла і ліпшицева з деякою константою L (в евклідовій нормі) на опуклій обмеженій множині $X \subseteq \mathrm{R}^n$ і справедлива оцінка $\|X\| = \sup_{x\in X} \|x\| \le D$;

послідовність $\{x_t\}$ будується рекурентно за наступним правилом:

$$x_{t+1} = \pi_X \left(x_t - \rho_t \cdot \eta_t \right), \qquad x_0 \in X , \qquad t = 1,...,T ,$$

$$\eta_t = \frac{1}{2h_t}\left(F(x_t + h_t \tilde{y}_t) - F(x_t - h_t \tilde{y}_t) \right) \tilde{y}_t ,$$

де ρ_t – детерміновані крокові множники; h_t – параметри згладжування; випадкові вектори $\{\tilde{y}_t\}$ незалежні і рівномірно розподілені на поверхні одиничної сфери; π_X – оператор проектування на опуклу множину X .

Тоді при постійному кроці $\rho_t = \rho = D/(L\sqrt{T})$ і фіксованому параметрі згладжування h точка $\bar{x}_T = (1/T)\sum_{t=1}^{T} x_t$ задовольняє співвідношенню:

$$\mathrm{E}\left[F_h(\bar{x}_T) - \min_{x\in X} F_h(x) \right] \le nLD/\sqrt{T} ,$$

і якщо $h \le nD/\sqrt{T}$, то

$$\mathrm{E}\left[F(\bar{x}_T) - \min_{x\in X} F(x) \right] \le 3nLD/\sqrt{T} .$$

При змінному кроці $\rho_t = D/(L\sqrt{t})$ і постійному h для будь-якого $t \le T$ і $\bar{x}_t = (1/t)\sum_{k=1}^{t} x_k$ виконано:

$$\mathrm{E}F_h(\bar{x}_t) - F_h(x^*) \le \frac{nLD}{2} \times \frac{1+\ln t}{\sqrt{t}-1}.$$

Зауваження. Більш детальний аналіз [3] для алгоритму:

$$x_{t+1} = \pi_X\left(x_t - \rho \cdot \eta_t\right), \qquad x_0 \in X, \qquad t = 1,...,T,$$

$$\eta_t = \frac{1}{K}\sum_{k=1}^{K} \frac{1}{2h}\left(F(x_t + h\cdot \tilde{y}_t^k) - F(x_t - h\cdot \tilde{y}_t^k)\right)\cdot \tilde{y}_t^k,$$

$$\rho = D/(L\sqrt{nT}), \;\; h = D\sqrt{n/T},$$

де $\left\{\tilde{y}_t^k\right\}$ - незалежні і рівномірно розподілені на поверхні одиничної сфери дає для $\bar{x}_T = (1/T)\sum_{t=1}^{T} x_t$ наступний результат:

$$\mathrm{E}\left[F(\bar{x}_T) - \min_{x\in X} F(x)\right] \le Const\frac{LD}{\sqrt{T}}\left(1+\sqrt{\frac{n}{K}}\right)$$

з деякою абсолютною числовою константою $Const$.

Подяка. Роботу виконано у рамках проекту Національного фонду досліджень України № 2020.02/0121 “Analytical methods and machine learning in control theory and decision-making in conditions of conflict and uncertainty”.

Література.

1. Норкин В.И. Два алгоритма случайного поиска для минимизации недифференцируемых функций. Математические методы исследования операций и теории надежности. Ред. Ю.М. Ермольев, И.Н. Коваленко. – Киев: Институт кибернетики, 1978. – С. 36-40.
2. Duchi J., Jordan M., Wainwright M., Wibisono A. Optimal rates for zero-order optimization: the power of two function evaluations // IEEE Transactions on Information Theory. – 2015. – Vol. 61(5). – P. 2788-2806.
3. Shamir O. An Optimal Algorithm for Bandit and Zero-Order Convex Optimization with Two-Point Feedback // J. of Machine Learning Research, 18 (2017), 1-11.