Улучшение критерия сходимости в безградиентных методах оптимизации в негладких случаях

Тафинцев Артём
Научный руководитель - Гасников А. В.

Краткое напоминание

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \; f(x, y)$$

$$f(x, y) \triangleq \mathbb{E}_\xi \left[f(x, y, \xi) \right]$$

$$\varphi(z, \xi) \triangleq f(z, \xi) + \delta(z)$$

Input: iteration number N

$z^1 \leftarrow \arg\min_{z \in \mathcal{Z}} d(z)$ **for** $k = 1, \ldots, N$

do

Sample $\mathbf{e}^k, \xi^k$ independently

Initialize γ_k

Calculate $g(z^k, \xi^k, \mathbf{e}^k)$ via (3)

$z^{k+1} \leftarrow \mathrm{Prox}_{z^k} \left(\gamma_k g(z^k, \xi^k, \mathbf{e}^k) \right)$

end

Output:

$\hat{z}^N \leftarrow \left(\sum_{k=1}^N \gamma_k \right)^{-1} \sum_{k=1}^N \gamma_k z^k$

$$E[f(x^N_{res}, y_*) - f(x_*, y^N_{res})] \leq M_{case1} D \sqrt{2/N} + \sqrt{d} \triangle D \tau^{-1} + 2\tau M_2$$

$$\epsilon_{sad}(z^N) = max_{y' \in Y} f(x^N, y') - min_{x' \in X} f(x', y^N)$$

Основная теорема

$$E[\epsilon_{sad}(z^N)] \leq M_{case1} D \sqrt{2/N} + \sqrt{d} \triangle D \tau^{-1} + 2\tau M_2$$

Леммы

1. Лемма 1. Пусть вектор e случайно выбран из Евклидовой сферы. Тогда для любого $r \in \mathbb{R}^d$(док-во через неравенство о средних)

$$E_e[|\langle e, r\rangle|] \leq \|r\|_2\sqrt{d}$$

2. Лемма 2. Пусть функция $f(z)$ M_2-Липшицева. Тогда

$$\sup_{z \in Z} |f^\tau(z) - f(z)| \leq \tau M_2$$

4. Лемма 4. Выполнено неравенство

$$E_{\xi,e}[\langle g(z,\xi,e), r\rangle] \geq \langle \nabla f^\tau(z), r\rangle - d\Delta\tau^{-1}E_e[|\langle e, r\rangle|]$$

6. Лемма 6.

$$E_{\xi,e}[\|g(z,,e)\|_q^2] \leq ca_q^2 dM_2^2 + d^2 a_q^2 \Delta^2/\tau^2$$

Краткое изложение доказательства

По определению z^{k+1} мы получаем(Ben-Tal and Nemirovski 2013 - ref. 2) для всех $u \in Z$

$$\gamma_k \langle g(z^k, e^k, \xi^k), z^k - u \rangle \leq V_{z^k}(u) - V_{z^{k+1}}(u) + \gamma_k^2 \|g(z^k, e^k, \xi^k)\|_q^2 / 2$$

Берём матожидание и суммируем по всем $k = 1, ..., N$(нер-во 1):

$$\sum_{k=1}^{N} \gamma_k E_{e^k, \xi^k} [\langle g(z^k, e^k, \xi^k), z^k - u \rangle] \leq V_{z^1}(u) + \sum_{k=1}^{N} \frac{\gamma_k^2}{2} E_{e^k, \xi^k} [\|g(z^k, e^k, \xi^k)\|_q^2]$$

$$\sum_{k=1}^{N} \gamma_k E_{e^k, \xi^k} [\langle g(z^k, e^k, \xi^k), z^k - u \rangle] \geq \sum_{k=1}^{N} \gamma_k \langle \nabla f^\tau(z^k), z^k - u \rangle$$

$$+ \sum_{k=1}^{N} \gamma_k E_{e^k} [|d \triangle \tau^{-1} e^k, z^k - u|]$$

Основные оценки

$$E_{e^k}[|\langle e^k, z^k - u\rangle|] \leq \|z^k - u\|_2/\sqrt{d}$$

$$\sum_{k=1}^{N} \frac{\gamma_k^2}{2} E_{e^k,\xi^k}[\|g(z^k, e^k, \xi^k)\|_q^2] \leq c a_q^2 d M_2^2 + d^2 a_q^2 \triangle^2 \tau^{-2}$$

$$\epsilon_{sad}(z^N) \leq max_{y' \in Y} f^\tau(x^N, y') - min_{x' \in X} f^\tau(x', y^N) + 2\tau M_2$$

$$\leq \frac{1}{\sum\limits_{k=1}^{N} \gamma_k}(max_{y' \in Y} \sum_{k=1}^{N} \gamma_k f^\tau(x^k, y') - min_{x' \in X} \sum_{k=1}^{N} \gamma_k f^\tau(x', y^k)) + 2\tau M_2$$

Основные оценки

$$\sum_{k=1}^{N} \gamma_k \langle \nabla f^\tau(z^k), z^k - u \rangle = \sum_{k=1}^{N} \gamma_k \langle (\nabla_x f^\tau(x^k, y^k), -\nabla_y f^\tau(x^k, y^k))^T, (u^T, -z^T) \rangle =$$

$$\sum_{k=1}^{N} \gamma_k (\langle \nabla_x f^\tau(x^k, y^k), x^k - x \rangle - \langle \nabla_y f^\tau(x^k, y^k), y^k - y \rangle) \geq$$

$$\sum_{k=1}^{N} \gamma_k ((f^\tau(x^k, y^k) - f^\tau(x, y^k)) - (f^\tau(x^k, y^k) - (f^\tau(x^k, y))) = \sum_{k=1}^{N} \gamma_k (f^\tau(x^k, y) - f^\tau(x, y^k))$$

Смысл результата и его сравнение с предыдущими

$$E[\epsilon_{sad}(z^N)] \leq \frac{3M_{case1}D}{N} + \sqrt{d}\triangle D\tau^{-1} + 2\tau M_2$$

$$E[\epsilon_{sad}(z^N)] \leq M_{case1}D\sqrt{2/N} + \sqrt{d}\triangle D\tau^{-1} + 2\tau M_2$$

$$E[f(x_{res}^N, y_*) - f(x_*, y_{res}^N)] \leq M_{case1}D\sqrt{2/N} + \sqrt{d}\triangle D\tau^{-1} + 2\tau M_2$$

Что дальше?

Что будет, если шум не ограничен константой?

$$M_{\text{case1}}^2 \triangleq \mathcal{O}\left(da_q^2 M_2^2 + d^2 a_q^2 \Delta^2 \tau^{-2}\right)$$

$$\mathbb{E}\left[\hat{\epsilon}_{\text{sad}}\right] \leq \sqrt{\frac{2}{N}} M_{\text{case1}} \sqrt{\mathbb{E}[V_{z^1}(z^\star)]} + \Delta \mathcal{D} \sqrt{d} \tau^{-1} + 2\tau M_2.$$

Ссылки

1. Shamir, O. (2017). An optimal algorithm for bandit and zero-order convex optimization with two-point feedback.
2. Ben-Tal, A. and Nemirovski, A. (2013). Lectures on modern convex optimization: analysis, algorithms, and engineering applications.
3. https://arxiv.org/pdf/2005.05913.pdf
4. https://arxiv.org/pdf/2202.06114.pdf
5. https://drive.google.com/file/d/1VgutizAY$_{C}ieL0dZapHp9yuTv0QMaB_{W}/view$