

Улучшение критерия сходимости в безградиентных методах оптимизации в негладких случаях

Тафинцев Артём

Постановка задачи

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

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

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

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

Input: iteration number N

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

do

 Sample \mathbf{e}^k, ξ^k independently

 Initialize γ_k

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

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

end

Output:

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

Мотивация

$$E[\epsilon_{sad}(z^N)] \leq \frac{3M_a D}{\sqrt{N}} + \frac{\Delta D da_q}{\tau} + 2\tau M_2$$

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

Проблема такой оценки: требуется дополнительное условие на функцию(сильная выпуклость)

$$f(x, y) = (x - 1)(y + 2)$$

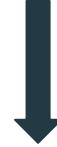
$$f(x_{res}^N, y) - f(x, y_{res}^N) = 0$$

Ход решения

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

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

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



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

Ход решения

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

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



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

Ход решения

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

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

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

Результаты

Простой алгоритм $E[\epsilon_{sad}(z^N)] \leq \frac{3M_a D}{\sqrt{N}} + \frac{\Delta D d a_q}{\tau} + 2\tau M_2$

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

Основной алгоритм
(результат)

$$E[\epsilon_{sad}(z^N)] \leq \frac{\sqrt{2} M_a D}{\sqrt{N}} + \frac{\Delta D d}{\tau} + 2\tau M_2$$

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/1VgutizAYCieL0dZapHp9yuTv0QMaBW/view>