ACCELERATED RANDOMIZED COORDINATE METHOD BY COUPLING MIRROR DESCENT AND PRIMAL GRADIENT METHOD

GASNIKOV A.V.

USMANOVA I.N.

Moscow Institute for Physics and Technology

Optimization problem: $f(x) \to \min_{x \in Q}$

Optimization problem: $\min_{x \in Q} f(x)$

$$f(x) \to \min_{x \in Q}$$

First-Order Methods:

Gradient-descent steps

Mirror-descent steps

Optimization problem:

$$f(x) \to \min_{x \in Q}$$

First-Order Methods:

Gradient-descent steps

Mirror-descent steps

Allen-Zhu Z., Orecchia L. ***Linear coupling: An ultimate unification of gradient and mirror descent 2015.***

Optimization problem:

$$f(x) \to \min_{x \in Q}$$

First-Order Methods:

Gradient-descent steps

Mirror-descent steps

Allen-Zhu Z., Orecchia L. ***Linear coupling: An ultimate unification of gradient and mirror descent 2015.***

Coordinate Descent Methods

$$e_i = \begin{pmatrix} 0 \\ ... \\ 0 \\ 1 \\ 0 \\ ... \\ 0 \end{pmatrix} - i$$

$$\nabla_i f(x) = \begin{pmatrix} 0 \\ ... \\ 0 \\ \frac{\partial f(x)}{\partial x_i} \\ 0 \\ ... \\ 0 \end{pmatrix} - i$$

$$e_i = \begin{pmatrix} 0 \\ \dots \\ 0 \\ 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} - i$$

$$\nabla_i f(x) = \begin{pmatrix} 0 \\ \dots \\ 0 \\ \frac{\partial f(x)}{\partial x_i} \\ 0 \\ \dots \\ 0 \end{pmatrix} - i$$

$$\left\{ \begin{array}{l} \|x\|^2 = \sum\limits_{i=1}^{n} L_i x_i^2 \\ \|\nabla f(x)\|_*^2 = \sum\limits_{i=1}^{n} L_i^{-1} (\frac{\partial f(x)}{\partial x_i})^2 \\ d(x) = \frac{1}{2}\|x\|^2 \\ V_x(y) = d(y) - \langle \nabla d(x), y - x \rangle - d(x) \end{array} \right.$$

$$e_i = \begin{pmatrix} 0 \\ \dots \\ 0 \\ 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} - i$$

$$\nabla_i f(x) = \begin{pmatrix} 0 \\ \dots \\ 0 \\ \frac{\partial f(x)}{\partial x_i} \\ 0 \\ \dots \\ 0 \end{pmatrix} - i$$

$$\|x\|^2 = \sum_{i=1}^{n} L_i x_i^2$$

$$\|\nabla f(x)\|_*^2 = \sum_{i=1}^{n} L_i^{-1} \left(\frac{\partial f(x)}{\partial x_i}\right)^2$$

$$d(x) = \frac{1}{2}\|x\|^2$$

$$V_x(y) = d(y) - \langle \nabla d(x), y - x \rangle - d(x)$$

$$Grad_i(x) = x - \frac{1}{L_i}\nabla_i f(x)$$

$$Mirr_x(\xi) = arg\min_{y \in Q}\{\langle \xi, y - x \rangle + V_x(y)\}$$

Accelerated by Coupling Randomized Coordinate Descent (ACRCD)

1. $x_{k+1} = \tau z_k + (1-\tau) y_k, \tau \in [0,1];$
2. $i_{k+1} \in \{1, ..., n\};$
3. $y_{k+1} = \mathrm{Grad}_{i_{k+1}}(x_{k+1});$
4. $z_{k+1} = \mathrm{Mirr}_{z_k}\left(\alpha n \nabla_{i_{k+1}} f(x_{k+1})\right), \alpha > 0.$

For every $u \in Q$

$$\alpha n \langle \nabla_{i_{k+1}} f(x_{k+1}), z_k - u \rangle \leq$$

$$\leq \frac{\alpha^2 n^2}{2} \|\nabla_{i_{k+1}} f(x_{k+1})\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \leq$$

$$\leq \alpha^2 n^2 (f(x_{k+1}) - f(y_{k+1})) + V_{z_k}(u) - V_{z_{k+1}}(u)$$

For every $u \in Q$

$$\alpha n \langle \nabla_{i_{k+1}} f(x_{k+1}), z_k - u \rangle \leq$$
$$\leq \frac{\alpha^2 n^2}{2} \|\nabla_{i_{k+1}} f(x_{k+1})\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \leq$$
$$\leq \alpha^2 n^2 (f(x_{k+1}) - f(y_{k+1})) + V_{z_k}(u) - V_{z_{k+1}}(u)$$

Mirr

For every $u \in Q$

$$\alpha n \langle \nabla_{i_{k+1}} f(x_{k+1}), z_k - u \rangle \leq$$

$$\leq \frac{\alpha^2 n^2}{2} \|\nabla_{i_{k+1}} f(x_{k+1})\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \leq$$

$$\leq \alpha^2 n^2 (f(x_{k+1}) - f(y_{k+1})) + V_{z_k}(u) - V_{z_{k+1}}(u)$$

Mirr

Grad

For every $u \in Q$

$$\alpha n \langle \nabla_{i_{k+1}} f(x_{k+1}), z_k - u \rangle \leq$$

$$\leq \frac{\alpha^2 n^2}{2} \|\nabla_{i_{k+1}} f(x_{k+1})\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \leq$$

$$\leq \alpha^2 n^2 (f(x_{k+1}) - f(y_{k+1})) + V_{z_k}(u) - V_{z_{k+1}}(u)$$

Mirr

Grad

$\mathbb{E}_{i_{k+1}}[\cdot | i_1, ..., i_k]$:

$$\alpha n \langle \nabla f(x_{k+1}), z_{k+1} - u \rangle \leq \alpha^2 n^2 \left(f(x_{k+1}) - \mathbb{E}_{i_{k+1}}[f(y_{k+1}) | i_1, ..., i_k] \right) + V_{z_k}(u) - \mathbb{E}_{i_{k+1}}[V_{z_{k+1}}(u) | i_1, ..., i_k]$$

For every $u \in Q$

$$\alpha n \langle \nabla_{i_{k+1}} f(x_{k+1}), z_k - u \rangle \leq$$

$$\leq \frac{\alpha^2 n^2}{2} \|\nabla_{i_{k+1}} f(x_{k+1})\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \leq$$

$$\leq \alpha^2 n^2 (f(x_{k+1}) - f(y_{k+1})) + V_{z_k}(u) - V_{z_{k+1}}(u)$$

Mirr

Grad

$\mathbb{E}_{i_{k+1}}[\cdot | i_1, ..., i_k]$:

$$\alpha n \langle \nabla f(x_{k+1}), z_{k+1} - u \rangle \leq \alpha^2 n^2 \left(f(x_{k+1}) - \mathbb{E}_{i_{k+1}}[f(y_{k+1}) | i_1, ..., i_k] \right) + V_{z_k}(u) - \mathbb{E}_{i_{k+1}}[V_{z_{k+1}}(u) | i_1, ..., i_k]$$

If

$$\frac{1-\tau}{\tau} = \alpha n^2$$

,

$$\alpha \langle \nabla f(x_{k+1}), x_{k+1} - u \rangle \leq \alpha^2 n^2 \left(f(y_k) - \mathbb{E}[f(y_{k+1}) | i_1, ..., i_k] \right) + V_{z_k}(u) - \mathbb{E}[V_{z_{k+1}}(u) | i_1, ..., i_k]$$

$$\alpha\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle \leq \alpha^2 n^2 \left(f(y_k) - \mathbb{E}[f(y_{k+1})|i_1, ..., i_k]\right) + V_{z_k}(u) - \mathbb{E}[V_{z_{k+1}}(u)|i_1, ..., i_k]$$

$$\alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle \le \alpha^2 n^2 \left(f(y_k) - \mathbb{E}[f(y_{k+1})|i_1, ..., i_k]\right) + V_{z_k}(u) - \mathbb{E}[V_{z_{k+1}}(u)|i_1, ..., i_k]$$

$$\bar{x} = \frac{1}{K}\sum_{k=0}^{K-1} x_k$$

$$u = x_*$$

$$\begin{aligned}\alpha K(\mathbb{E} f(\bar{x}) - f(x_*)) &\le \alpha \sum_{k=0}^{K-1} \mathbb{E}\langle \nabla f(x_k), x_k - x_*\rangle \le \\ &\le \alpha^2 n^2 (f(x_0) - \mathbb{E} f(y_k)) + V_{x_0}(x_*) - V_{z_k}(x_*) \le \\ &\le \alpha^2 n^2 (f(x_0) - f(x_*)) + V_{x_0}(x_*)\end{aligned}$$

$$V_{x_0}(x_*) \le \Theta \qquad\qquad f(x_0) - f(x_*) \le d$$

Setting $\alpha = \frac{1}{n}\sqrt{\frac{\Theta}{d}}$

After $K = 4n\sqrt{\frac{\Theta}{d}}$ time steps

Convergence rate (expectation):

$$\mathbb{E} f(\bar{x}) - f(x_*) \le \frac{2n\sqrt{\Theta d}}{K} \le \frac{d}{2}$$

Markov inequality

$$\mathbb{P}\left(f(\bar{x}) - f(x_*) \geq \frac{3d}{4}\right) \leq \frac{2}{3}$$

Markov inequality	$\mathbb{P}\left(f(\bar{x}) - f(x_*) \geq \frac{3d}{4}\right) \leq \frac{2}{3}$

If we independently (may be parallel) start

$$\log_{3/2}\left(\sigma^{-1}\right)$$

different trajectories ARCDC $(\alpha, \tau; x, d)$

Markov inequality $\mathbb{P}\left(f(\bar{x}) - f(x_*) \geq \frac{3d}{4}\right) \leq \frac{2}{3}$

If we independently (may be parallel) start

$$\log_{3/2}\left(\sigma^{-1}\right)$$

different trajectories ARCDC $(\alpha,\tau;x,d)$

Then with the probability $\mathbb{P} \geq 1 - \sigma$

at least one of trajectories will achieve $f(\bar{x}) - f(x_*) \leq \frac{3d}{4}$

After a few restarts N = $log_{4/3}\left(\frac{d}{\varepsilon}\right)$

$$K \leq 4\left(n\sqrt{\frac{\Theta}{\varepsilon}} + n\sqrt{\frac{3\Theta}{4\varepsilon}} + n\sqrt{\frac{9\Theta}{16\varepsilon}} + ...\right) \leq 16n\sqrt{\frac{\Theta}{\varepsilon}}$$

We will achieve the solution with an accuracy ε

After a few restarts N = $log_{4/3}\left(\frac{d}{\varepsilon}\right)$

$$K \leq 4\left(n\sqrt{\frac{\Theta}{\varepsilon}} + n\sqrt{\frac{3\Theta}{4\varepsilon}} + n\sqrt{\frac{9\Theta}{16\varepsilon}} + ...\right) \leq 16n\sqrt{\frac{\Theta}{\varepsilon}}$$

We will achieve the solution with an accuracy ε

Also, it have to be multiplied on $\log_{3/2}\left(\frac{N}{\sigma}\right)$ times

After a few restarts N = $log_{4/3}\left(\frac{d}{\varepsilon}\right)$

$$K \le 4\left(n\sqrt{\frac{\Theta}{\varepsilon}} + n\sqrt{\frac{3\Theta}{4\varepsilon}} + n\sqrt{\frac{9\Theta}{16\varepsilon}} + ...\right) \le 16n\sqrt{\frac{\Theta}{\varepsilon}}$$

We will achieve the solution with an accuracy ε

$$\alpha = \frac{1}{n}\sqrt{\frac{\Theta}{d}} \qquad \tau = \frac{1}{\alpha n^2 + 1}$$

Accelerated by Coupling Randomized Coordinate Descent (ACRCD)

1. $x_{k+1} = \tau z_k + (1-\tau) y_k, \tau \in [0,1];$
2. $i_{k+1} \in \{1, ..., n\};$
3. $y_{k+1} = \mathrm{Grad}_{i_{k+1}}(x_{k+1});$
4. $z_{k+1} = \mathrm{Mirr}_{z_k}\left(\alpha n \nabla_{i_{k+1}} f(x_{k+1})\right), \alpha > 0.$

$$x_0 = y_0 = z_0$$

$$\alpha = \frac{1}{n}\sqrt{\frac{\Theta}{d}} \qquad \tau = \frac{1}{\alpha n^2 + 1}$$

Theorem

Algorithm ACRCD ensures accuracy $f(\bar{x}) - f(x_*) \leq \varepsilon$

with the probability $\mathbb{P} \geq 1 - \sigma$

after N= $log_{4/3}\left(\frac{d}{\varepsilon}\right)$ restarts of

$\log_{3/2}\left(\frac{N}{\sigma}\right)$ independent sessions of ACRCD

Total number of iterations $O\left(n\sqrt{\frac{\Theta}{\varepsilon}}ln\left(\frac{d}{\varepsilon}\right)ln\left(\frac{ln\left(\frac{d}{\varepsilon}\right)}{\sigma}\right)\right)$

Theorem

Algorithm ACRCD ensures accuracy $f(\bar{x}) - f(x_*) \leq \varepsilon$

with the probability $\mathbb{P} \geq 1 - \sigma$

after N= $log_{4/3}\left(\frac{d}{\varepsilon}\right)$ restarts of

$\log_{3/2}\left(\frac{N}{\sigma}\right)$ independent sessions of ACRCD

Total number of iterations $O\left(n\sqrt{\frac{\Theta}{\varepsilon}}ln\left(\frac{d}{\varepsilon}\right)ln\left(\frac{ln\left(\frac{d}{\varepsilon}\right)}{\sigma}\right)\right)$

Each iteration $O(n)$

EXAMPLE:

$$O\left(n\sqrt{\frac{\Theta}{\varepsilon}}ln\left(\frac{d}{\varepsilon}\right)ln\left(\frac{ln\left(\frac{d}{\varepsilon}\right)}{\sigma}\right)\right)$$

Coordinate: $\Theta = V_{x_o}(x_*) = \frac{1}{2}\|x_*\|^2 = \frac{1}{2}\sum L_i x_i^2$

Gradient: $\frac{1}{2}L\|x_0 - x_*\|_2^2 = \frac{1}{2}L\|x_*\|_2^2 = \frac{1}{2}\sum L x_i^2$

EXAMPLE:

$$O\left(n\sqrt{\frac{\Theta}{\varepsilon}}\,ln\left(\frac{d}{\varepsilon}\right) ln\left(\frac{ln\left(\frac{d}{\varepsilon}\right)}{\sigma}\right)\right)$$

Coordinate: $\Theta = V_{x_o}(x_*) = \frac{1}{2}\|x_*\|^2 = \frac{1}{2}\sum L_i x_i^2$

Gradient: $\frac{1}{2}L\|x_0 - x_*\|_2^2 = \frac{1}{2}L\|x_*\|_2^2 = \frac{1}{2}\sum L x_i^2$

$f(x) = x^T S x$ S = S, T {1,2}

$L \geq n$

$\max L_i \leq 2\sqrt{n}$

Conclusion

The new philosophy shows primal-dual nature of Accelerated Coordinate Method and allows to receive Accelerated Coordinate Method from non-accelerated methods.

Also, it can be simply extended on the case of other non-Euclidean norms.

Acknowledgements

We thank

Y.E. Nesterov,
Alexander Rachlin
Peter Richtarik

for chain of valuable links

Thanks!