Acceleration Exists! Optimization Problems When Oracle Can Only Compare Objective Function Values

1 MIPT 2 Skoltech (Skolkovo Institute of Science and Technology) 3 ISP RAS

Aleksandr Lobanov[1,2,3], Alexander Gasnikov[1,2,3], Andrei Krasnov[1]

ASCOMP 2024

ASCOMP 2024, Innopolis

Setup

Optimization problem, possibly non-convex, possibly stochastic:

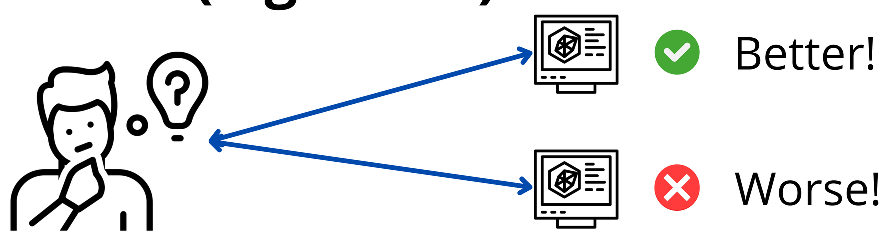
$$\min_{x \in \mathbb{R}^d} \left\{ f(x) := \mathbb{E}_{\xi \sim \mathcal{D}} f_\xi(x) \right\}$$

Feedback, Deterministic Order Oracle:

$$\phi(x, y) = \text{sign}\left[f(x) - f(y) + \delta(x, y) \right]$$

Motivation

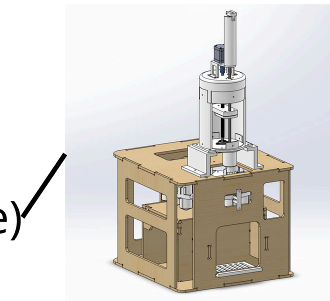
Human Feedback (e.g. RLHF):



Our motivation: AI chocolate, AI wine, etc.

(Valio's chocolate)

(Smart coffee machine)



Friendly method to our oracle

Algorithm Golden Ratio Method (GRM)

Input: Interval $[a, b]$
Initialization: Choose constants $\epsilon > 0$
$y \leftarrow a + (1 - \rho)(b - a)$
$z \leftarrow a + \rho(b - a)$ — $\rho = \frac{1}{\Phi} = \frac{\sqrt{5}-1}{2}$
while $b - a > \epsilon$ **do**
  if $\phi(y, z) = -1$ **then**
    $b \leftarrow z$
    $z \leftarrow y$
    $y \leftarrow a + (1 - \rho)(b - a)$
  else
    $a \leftarrow y$
    $y \leftarrow z$
    $z \leftarrow a + \rho(b - a)$
  end if
end while
Return: $\frac{a+b}{2}$

$$\frac{z - a}{b - z} = \frac{y - a}{z - y}$$

Trouble: GRM solve a one-dimensional optimization problem

Assumptions

***L*-coordinate-Lipschitz smoothness:**

$$|\nabla_i f(x + h\mathbf{e}_i) - \nabla_i f(x)| \leq L_i |h|; \quad i \in [d], x \in \mathbb{R}^d$$

(Strong) convexity w.r.t. the norm $\|\cdot\|_{[1-\alpha]}$**:**

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu_{1-\alpha}}{2} \|y - x\|^2_{[1-\alpha]}$$

Define norms (*Y. Nesterov, 2012*)**:**

$$\|\cdot\|_{[\alpha]} := \sqrt{\sum_{i=1}^d L_i^\alpha x_i^2} \text{ and } \|\cdot\|^*_{[\alpha]} := \sqrt{\sum_{i=1}^d \frac{1}{L_i^\alpha} x_i^2}$$

Non-Accelerated Methods

Algorithm Random Coordinate Descent with Order Oracle (OrderRCD)

Input: $x_0 \in \mathbb{R}^d$, random generator $\mathcal{R}_\alpha(L)$
for $k = 0$ **to** $N - 1$ **do**
1. choose active coordinate $i_k = \mathcal{R}_\alpha(L)$ — $p_\alpha(i) = L_i^\alpha / S_\alpha$ — $S_\alpha := \sum_i^d L_i^\alpha$
2. compute $\eta_k = \text{argmin}_\eta \{ f(x_k + \eta \mathbf{e}_{i_k}) \}$ via (GRM)
3. $x_{k+1} \leftarrow x_k + \eta_k \mathbf{e}_{i_k}$

end for
Return: x_N

Non-convex setting:

$$\frac{1}{N} \sum_{k=0}^{N-1} \left(\|\nabla f(x_k)\|^*_{[1-\alpha]} \right)^2 \leq \mathcal{O}\left(\frac{S_\alpha F_0}{N} + S_\alpha \epsilon + S_\alpha \Phi \Delta \right)$$

Convex setting:

$$\mathbb{E}\left[f(x_N) \right] - f(x^*) \leq \mathcal{O}\left(\frac{S_\alpha R^2_{[1-\alpha]}}{N} + \frac{2 S_\alpha R^2_{[1-\alpha]} (\epsilon + \Phi \Delta)}{F_{N-1}} \right)$$

Strongly convex setting:

$$\mathbb{E}\left[f(x_N) \right] - f(x^*) \leq \left(1 - \frac{\mu_{1-\alpha}}{S_\alpha} \right)^N F_0 + \frac{2 S_\alpha \epsilon}{\mu_{1-\alpha}} + \frac{2 c S_\alpha \Phi \Delta}{\mu_{1-\alpha}}$$

Prior works

Reference	Nesterov (2012)	Gorbunov et al. (2019)	Saha et al. (2021)	Tang et al. (2023)	This paper	
Non-convex	✗	$\mathcal{O}\left(\frac{dS_\alpha F_0}{\varepsilon^2}\right)$	✗	$\mathcal{O}\left(\frac{dLF_0}{\varepsilon^2}\right)$	$\tilde{\mathcal{O}}\left(\frac{S_\alpha F_0}{\varepsilon^2}\right)$	
Convex	$\mathcal{O}\left(\frac{S_\alpha R^2_{[1-\alpha]}}{\varepsilon}\right)$	$\mathcal{O}\left(\frac{dS_\alpha R^2_{[1-\alpha]}}{\varepsilon}\log\frac{1}{\varepsilon}\right)$	$\mathcal{O}\left(\frac{dLR^2}{\varepsilon}\right)$	✗	$\tilde{\mathcal{O}}\left(\frac{S_\alpha R^2_{[1-\alpha]}}{\varepsilon}\right)$	
Strongly convex	$\mathcal{O}\left(\frac{S_\alpha}{\mu_{1-\alpha}}\log\frac{1}{\varepsilon}\right)$	$\mathcal{O}\left(\frac{S_\alpha}{\mu_{1-\alpha}}\log\frac{1}{\varepsilon}\right)$	$\mathcal{O}\left(d\frac{L}{\mu}\log\frac{1}{\varepsilon}\right)$	✗	$\tilde{\mathcal{O}}\left(\frac{S_\alpha}{\mu_{1-\alpha}}\log\frac{1}{\varepsilon}\right)$	$\tilde{\mathcal{O}}\left(\frac{S_{\alpha/2}}{\sqrt{\mu_{1-\alpha}}}\log\frac{1}{\varepsilon}\right)$
Order Oracle?	✗	✓	✓	✓	✓	✓
Acceleration?	✗	✗	✗	✗	✗	✓

Accelerated Method

Algorithm Accelerated Coordinate Descent Method with Order Oracle (OrderACDM)

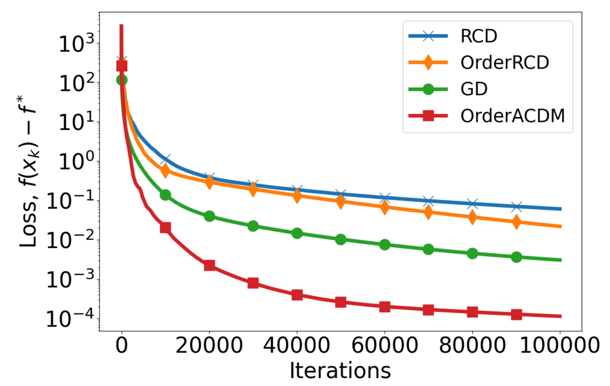
Input: $x_0 = z_0 \in \mathbb{R}^d, \mathcal{R}_\alpha(L), A_0 = 0, B_0 = 1, \beta = \frac{\alpha}{2}$
for $k = 0$ **to** $N - 1$ **do**
1. choose active coordinate $i_k = \mathcal{R}_\beta(L)$ — $p_\beta(i) = \frac{L_i^{\alpha/2}}{S_{\alpha/2}}$
2. find parameter a_{k+1} from $a_{k+1}^2 S_\beta^2 = A_{k+1} B_{k+1}$, where $A_{k+1} = A_k + a_{k+1}$ and $B_{k+1} = B_k + \mu_{1-\alpha} a_{k+1}$
3. $\alpha_k \leftarrow \frac{a_{k+1}}{A_{k+1}}$
4. $\beta_k \leftarrow \frac{\mu_{1-\alpha} a_{k+1}}{B_{k+1}}$
5. $y_k \leftarrow \frac{(1-\alpha_k) x_k + \alpha_k (1-\beta_k) z_k}{1 - \alpha_k \beta_k}$
6. compute $\eta_k = \text{argmin}_\eta \{ f(y_k + \eta \mathbf{e}_{i_k}) \}$ via (GRM)
7. $x_{k+1} \leftarrow y_k + \eta_k \mathbf{e}_{i_k}$
8. $w_k \leftarrow (1 - \beta_k) z_k + \beta_k y_k + \frac{a_{k+1} L_{i_k}^\alpha}{B_{k+1} p_\beta(i)} \eta_k \mathbf{e}_{i_k}$
9. compute $\zeta_k = \text{argmin}_\zeta \{ f(w_k + \zeta \mathbf{e}_{i_k}) \}$ via (GRM)
10. $z_{k+1} \leftarrow w_k + \zeta_k \mathbf{e}_{i_k}$

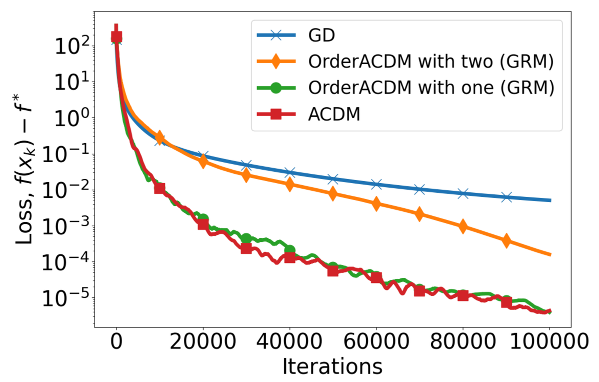
end for
Return: x_N

Convergence in strongly convex setting:

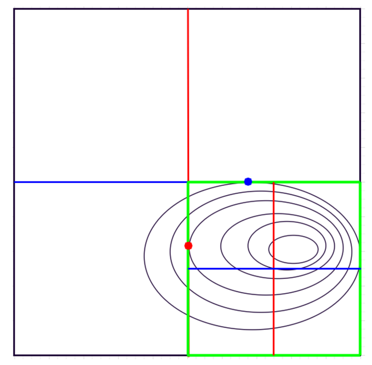
$$\mathbb{E}\left[f(x_N) \right] - f(x^*) \leq \left(1 - \frac{\sqrt{\mu_{1-\alpha}}}{S_{\alpha/2}} \right)^N F_0$$

Experiments





Low-dimensional case



Golden ration method + Nesterov's 2D generalization

Total number of Order Oracle calls:

$$\sim \log^2\left(\frac{LR}{\varepsilon}\right)$$

Stochastic Order Oracle

$$\phi(x, y, \xi) = \text{sign}\left[f(x, \xi) - f(y, \xi) \right]$$ → New feedback

Algorithm: $x_{k+1} = x_k - \eta_k \phi(x_k + \gamma \mathbf{e}_k, x_k - \gamma \mathbf{e}_k, \xi_k) \mathbf{e}_k$

(γ: smoothing parameter; $\mathbf{e}_k$: uniformly distributed on the Euclidean sphere)

Asymptotic convergence: $\sqrt{N}(x_k - x^*) \sim \mathcal{N}(0, V)$, where the matrix V is:

$$V = \frac{\eta^2}{d} \left(2\eta(1 - 1/d) \frac{c}{\sqrt{d}} \alpha \nabla^2 f(x^*) - I \right)^{-1}$$

$2\eta(1 - 1/d)\frac{c}{\sqrt{d}}\alpha \nabla^2 f(x^*) > I$; $\alpha = \int \|z\|^{-1} dP(z) < \infty$