Comparison of gradient methods for smooth strongly convex optimization

Mikhail Zyatkov

Optimization Class Project. MIPT

Introduction

In connection with recent advances [1], [2] in finding optimal first-order methods of smooth convex optimization, there is a question about the practical application of them. This project proposes to investigate new methods for convergence rate on example functions of mentioned class. Methods ITEM [1], FGD, GD, OGM, TMM will be used as objects of study. Moreover, we propose two adaptive methods, based on ITEM and approach from [3], which adaptive by Lipschitz constant and strong convexity constant. Code available in GitHub [4]

Preliminaries

Consider optimization problem

$$\min_{x \in \mathbb{R}^d} f(x), \tag{1}$$

Let f be a proper, closed and convex function, which is L-smooth and μ -strongly convex, i.e. $\forall x, y \in \mathbb{R}^d$:

$$f(x) \leq f(y)\langle \nabla f(y), x - y\rangle + \frac{L}{2}\|x - y\|^2$$

$$f(x) \geq f(y)\langle \nabla f(y), x - y\rangle + \frac{\mu}{2}\|x - y\|^2$$

Algorithm

We solve optimization problem using ITEM implementation:

Data: $f \in \mathcal{F}_{\mu,L}$ with $0 \leq \mu < L < \infty$, initial guess $x_0 \in \mathbb{R}^d$
Input: $y_{-1} = z_0 = x_0,\ A_0 = 0,\ q = \mu/L$

for $k = 0, 1, \dots$ **do**
  Set $A_{k+1} = \frac{(1+q)A_k + 2(1+\sqrt{(1+A_k)(1+qA_k)})}{(1-q)^2}$
  $\beta_k = \frac{A_k}{(1-q)A_{k+1}}$, and $\delta_k = \frac{1}{2}\frac{(1-q)^2 A_{k+1} - (1+q)A_k}{1+q+qA_k}$
  $y_k = (1-\beta_k)z_k + \beta_k x_k$
  $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$
  $z_{k+1} = (1 - q\delta_k)z_k + q\delta_k y_k - \frac{\delta_k}{L}\nabla f(y_k)$
end

Algorithm 1: Information-Theoretic Exact Method (ITEM)

ensures

$$\|z_N - x_*\|^2 \leq \frac{1}{1+qA_N}\|z_0 - x_*\|^2 \leq \frac{(1-\sqrt{q})^{2N}}{(1-\sqrt{q})^{2N} + q}\|z_0 - x_*\|^2$$

$$f(x_{N+1}) - f_* \leq \psi_N \leq \min\left\{(1-\sqrt{q})^{2(N+1)}, \frac{1}{(N+1)^2}\right\}\frac{L}{1-q}\|z_0 - x_*\|^2$$

Adaptive by strong Convexity Gradient Method

Data: $f \in \mathcal{F}_L$, initial guess $x_0 \in \mathbb{R}^d, L, \mu_0, \beta, \epsilon$
Output: $[x_0, \dots, x_N]$

for $k \geq 0$ **do**
  if $\|\nabla f(x_k)\| \leq \epsilon$ **then**
    break;
  end
  $\mu_k := \beta\mu_{k-1}$;
  $x_k =$ ITEM(f, x_{k-1}, L, μ_k);
  if $\|\nabla f(x_k)\| \leq \frac{1}{2}\|\nabla f(x_{k-1})\|$ **then**
    continue;
  end
  $\mu_k := \frac{\mu_k}{\beta}$;
  if $\|\nabla f(x_k)\| < \|\nabla f(x_{k-1})\|$ **then**
    $x_{k-1} := x_k$
  end
  goto 6;
end

Algorithm 2: Adaptive by strong Convexity ITEM

Adaptive by Lipschitz constant Gradient Method

Data: $f \in \mathcal{F}_L$, initial guess $x_0 \in \mathbb{R}^d, L_0, \mu_0, \beta, \epsilon$
Input: $y_{-1} = z_0 = x_0,\ A_0 = 0,\ q = \mu/L$
Output: $[x_0, \dots, x_N]$

for $k \geq 0$ **do**
  if $\|\nabla f(x_k)\| \leq \epsilon$ **then**
    break;
  end
  $\mu_k := \beta\mu_{k-1}$;
  $x_k, L_k =$ ITEM-GL$(f, x_{k-1}, L_{k-1}, \lceil\sqrt{\frac{8L_{k-1}}{\mu_k}}\rceil, \epsilon)$;
  if $\|\nabla f(x_k)\| \leq \frac{1}{2}\|\nabla f(x_{k-1})\|$ **then**
    continue;
  end
  $\mu_k := \frac{\mu_k}{\beta}$;
  if $\|\nabla f(x_k)\| < \|\nabla f(x_{k-1})\|$ **then**
    $x_{k-1} := x_k$
  end
  goto 6;
end

Algorithm 3: Adaptive by Lipschitz constant ITEM

Numerical example

Consider numerical examples with real data and following analytical smooth convex functions:

$$f_1 = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle; A \in S_{++}$$

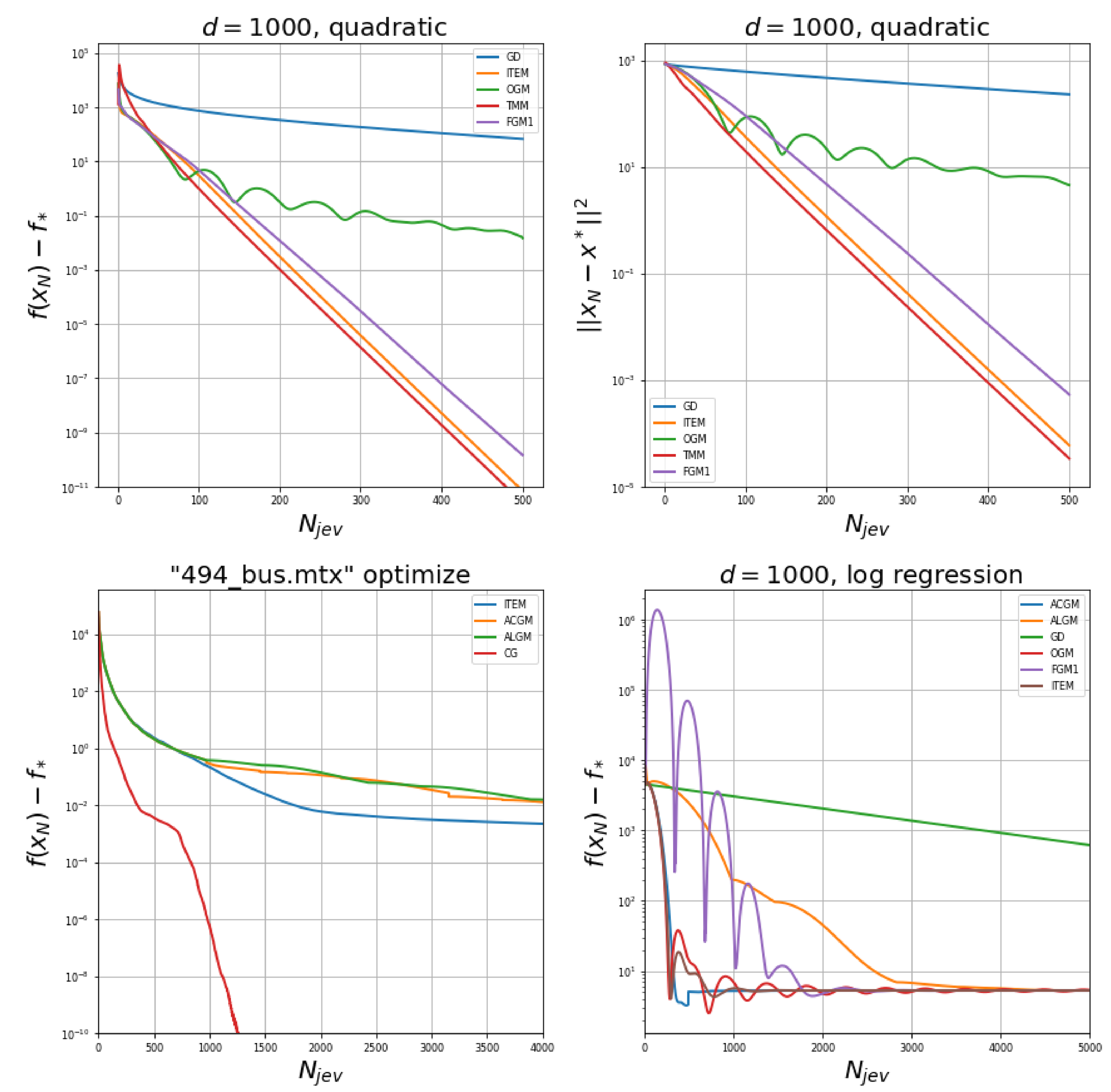
$$f_2 = \sum_{i=1}^{p} g(a_i^T x - b_i) + \frac{\mu}{2}\|x\|^2$$

, where

$$g(x) = \begin{cases} \frac{1}{2}x^2 e^{-r/x}, x > 0 \\ 0, x \leq 0 \end{cases} \tag{2}$$

$$A = [a_1, \dots, a_p], b \in R^p, \|A\| = \sqrt{L - \mu}$$

Experiments and results



Conclusion

As we can see, ITEM method has the best convergence rate from all the methods mentioned above, that was obtained analytically. Moreover, we provided adaptive ITEM methods, which have worthy convergence on many functions.

References

[1] Adrien Taylor and Yoel Drori. An optimal gradient method for smooth strongly convex minimization, 2021.

[2] Yoel Drori and Adrien Taylor. On the oracle complexity of smooth strongly convex minimization, 2021.

[3] Nikita Pletnev. Fast adaptive by constants of strong-convexity and lipschitz for gradient first order methods, 2020.

[4] `https://github.com/mzyatkov/optimization_project`.