Gradient Methods for Inexact Strongly Convex Model

Artem Agafonov
Optimization Class Project. MIPT

Introduction

On this poster, some analogs of the (δ, L, μ)-oracle by Devolder-Glineur-Nesterov [1], [3] are introduced. At the same time, various types of conditions of relative smoothness and relative strong convexity of the objective function are highlighted. The gradient method and restarts of fast gradient method that allow for inexact model are studied.

Many optimization methods are based on idea of approximating function with a simpler function-model. [2] After this replacement, it is proposed to solve the optimization problem not for the original function, but for its model. This can be said, for example, about such popular methods as the gradient method (GM) and the fast gradient method (FGM). Both methods are based on the idea of approximating a function at the initial point (current method position) by its majorising paraboloid of revolution and choosing the minimum point of the paraboloid of revolution as the new position of the method.

(δ, L, μ)-model

Let function f be μ-strongly convex on convex set Q. We say that it is equipped with (δ, L, μ) - model (left (δ, L, μ) - model) if for any $y \in Q$ if we can compute a pair (f_δ, ψ_δ) such that

$$\mu V(x, y) \le f(y) - (f_\delta(x) + \psi_\delta(y, x)) \le LV(x, y) + \delta,$$

where $\psi_\delta(y, x)$ is convex in y, $\psi_\delta(y, x) = 0$, $\delta > 0$.

Inaccurate Problem

Let us consider the concept ot inaccurate problem solving [2].

Consider the next problem: $\psi(x) \to \min\limits_{x \in Q}$, where $\psi(x)$ is convex.

Then $\operatorname{Arg\,min}_{\widetilde{\delta}} \psi(x)$ is the set of $\widetilde{x}$ such that $\exists h \in \partial\psi(\widetilde{x}),\ \langle h, x - \widetilde{x}\rangle \ge$ $\ge -\widetilde{\delta}\ \ \forall x \in Q$. Any element from $\operatorname{Arg\,min}_{\widetilde{\delta}} \psi(x)$ we call $\operatorname{argmin}_{\widetilde{\delta}} \psi(x)$.

Gradient Method for (δ, L, μ)-model

1: **Initialization:** choose $x^0 \in Q$.
2: **for** $k = 0, \ldots$ **do**
3: Compute x^{k+1}:

$$x^{k+1} = \operatorname*{argmin}_{\widetilde{\delta}} \left\{ \underbrace{\psi_\delta(x, x^k) + LV(x, x^k)}_{\Psi(x, x^k)} \right\},$$

4: **end for**

Convergence Rate of Gradient Method

Assume that function f is equipped with a (δ, L, μ)-model w.r.t. $V(x, y)$. Then, after of k iterations version of Gradient Method, f satisfies

$$f(y_{k+1}) - f(x_*) \le LV(x_*, x^0)\exp\left((-k+1)\frac{\mu}{L}\right) + \delta + \widetilde{\delta},$$

where $y_k = \operatorname{argmin}_{i=1,\ldots,k}(f(x_i))$.

FGM for (δ, L)-model and Its Restarts

1: **Initialization:** Choose $x^0 \in Q$, N – number of steps, $\{\delta_k\}_{k=0}^{N-1}$, $\{\widetilde{\delta}_k\}_{k=0}^{N-1}$ – sequences and $L_0 > 0$.
2: 0-**step:**

$$y_0 := x_0,\ u_0 := x_0,\ L_1 := \frac{L_0}{2},\ \alpha_0 := 0,\ A_0 := \alpha_0$$

3: **for** $k = 1, \ldots$ **do**
4: Find α_{k+1} :

$$A_k + \alpha_{k+1} = L_{k+1}\alpha_{k+1}^2$$

5:

$$A_{k+1} := A_k + \alpha_{k+1}, \quad y_{k+1} := \frac{\alpha_{k+1}u_k + A_k x_k}{A_{k+1}}$$

6:

$$\phi_{k+1}(x) = V(x, u_k) + \alpha_{k+1}\psi_{\delta_k}(x, y_{k+1})$$
$$u_{k+1} := \operatorname*{argmin}_{x \in Q}{}^{\widetilde{\delta}_k} \phi_{k+1}(x)$$

7:

$$x_{k+1} := \frac{\alpha_{k+1}u_{k+1} + A_k x_k}{A_{k+1}}$$

8: **if** $F_{\delta_k}(x_{k+1}) \le F_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x_{k+1}, y_{k+1}) + \frac{L_{k+1}}{2}\|x_{k+1} - y_{k+1}\|^2 + \delta_k$ **then**
9: $L_{k+2} := \frac{L_{k+1}}{2}$ and move to next step
10: **else**
11: $L_{k+1} := 2L_{k+1}$ and repeat this step
12: **end if**
13: **end for**

We will restart this method to obtain method for strongly convex function f which is equipped with (δ, L, μ)-model.

Right (δ, L, μ)-model

Let function f be μ-strongly convex on convex set Q. We say that it is equipped with right (δ, L, μ) - model if for any $y \in Q$ if we can compute a pair (f_δ, ψ_δ) such that $\mu V(y, x) \le f(y) - (f_\delta(x) + \psi_\delta(y, x)) \le LV(x, y) + \delta$, where $\psi_\delta(y, x)$ is convex in y, $\psi_\delta(y, x) = 0$, $\delta > 0$.

Note, that definitions of right and left (δ, L, μ)-models are equal in the case, when the following inequality on prox-function d(x) is satisfied: $d(x - y) \le C_n\|x - y\|^2, C_n = O(\log n)$.

Convergence Rate of Restarts of FGM

Assume that function f is equipped with a right (δ, L, μ)-model w.r.t. $V(x, y)$. And the following inequality holds: $\varepsilon \ge \frac{12\mu}{L}\left(9\delta\left\lceil\sqrt{\frac{L}{\mu}}\right\rceil^3 + L\widetilde{\delta}\left\lceil\sqrt{\frac{L}{\mu}}\right\rceil\right)$. Then restart of Fast Gradient Method achieve accuracy ε of function in

$$M = \left\lceil \log_4 \frac{\mu R^2}{\varepsilon} \right\rceil \cdot \left\lceil 6\sqrt{\frac{L}{\mu}} \right\rceil$$

iterations.

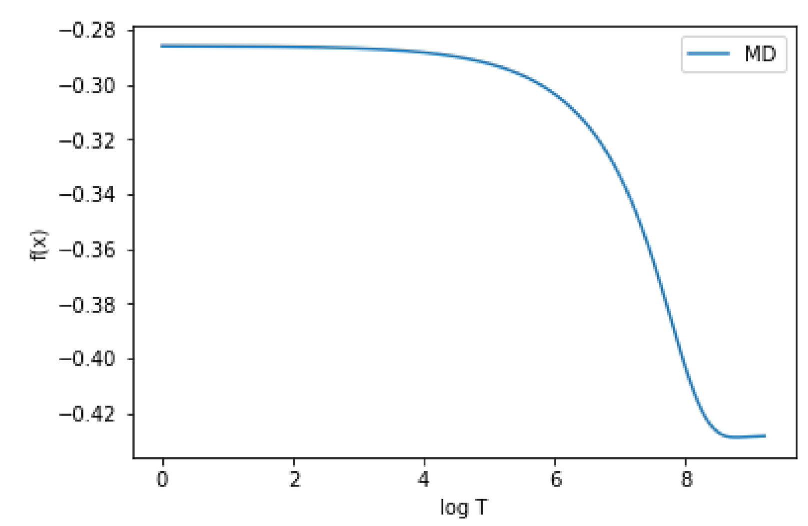
Numerical examples

- $F(x) = \frac{1}{2}\|Ax - b\|_2^2 + \mu\sum\limits_{k=1}^{n} x_k \ln x_k \to \min\limits_{\sum_{k=1}^{n} x_k = 1, x \ge 0}$, $n = 6$, x_0 random from simplex, 10 launches;
- $f(x) = x_1^2 + 2x_2^2 + \ldots + Nx_N^2 \to \min\limits_{x \in B_1(0)}$, $N = 10000$;
- $f(x) = \|Ax - b\|_2^2 \to \min\limits_{\sum_{k=1}^{n} x_k = 1, x \ge 0}$, $n = 50$, x_0 random from simplex, 10 launches.

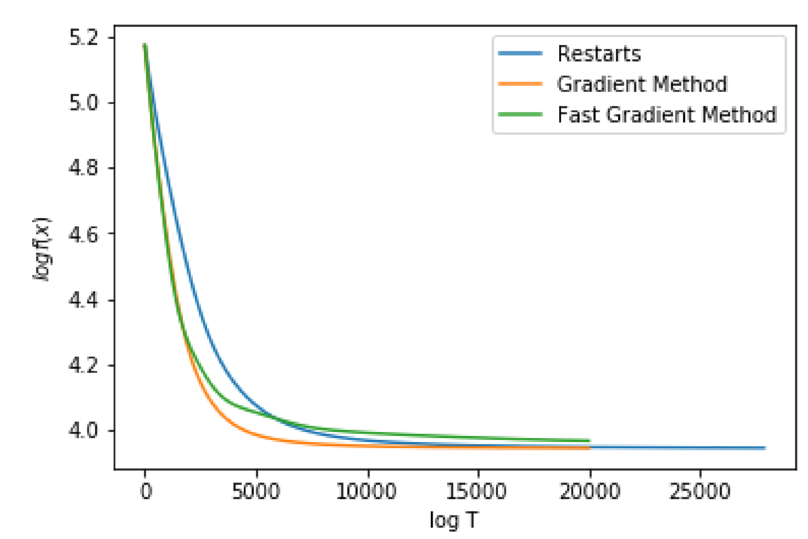
Consider a numerical example with the following function

Results

- In the first experiment the Gradient method converges quickly.



- In the second experiment GM and FGM converges in a small number of iterations, so the time was measured. GM: 0.00612 sec. FGM: 0.0104 sec.
- In the third experiment all three methods were tested.



Conclusion

The linear rate of convergence of gradient method is justified. Some numerical experiments are presented. An approach to the problem of the accumulation of errors for the fast gradient method is proposed using a special technique of its restarts.

[1] *Devolder O., Glineur F., Nesterov Yu.* First-order methods of smooth convex optimization with inexact oracle.
[2] *Tyurin A. I., Gasnikov A. V.* Fast gradient descent method for convex optimization problems with an oracle that generates a (δ, L)-model of a function in a requested point.
[3] *Devolder O., Glineur F., Nesterov Yu.* First-order methods with inexact oracle: the strongly convex case.