Conditional Gradient Method

Grinev Timofey

December 23th, 2022

Problem statement

Consider a problem:

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

where $\mathcal{X} \subset \mathbb{R}^n$.

Usually: $\mathcal{X}$ - compact, convex and bounded set, f - convex and smooth on $\mathcal{X}$ with L-Lipschitz gradient.

Solution

Usual scheme is as follows:

- Compute auxiliary point
  $\hat{x}^k = Argmin_{x \in \mathcal{X}} \left\{ f(x^k) + < \nabla f(x^k), x - x^k > \right\}$
- Choose step-size: $\gamma_k = \frac{2}{k+2}$ (function agnostic) or
  $\gamma_k \in Argmin_{0 \leqslant \gamma \leqslant 1} \left\{ f(x^k + \gamma(\hat{x}^k - x^k)) \right\}$ (line search)
- $x^{k+1} = x^k + \gamma_k(\hat{x}^k - x^k)$

This algorithm is known as *Frank-Wolfe algorithm* (vanilla version).

Convergence speed

Under the assumptions above, we have:

$$f(x^k) - f(x^*) = \mathcal{O}\left(\frac{1}{k}\right)$$

If f is μ-strongly convex, we also have

$$\|x^k - x^*\|_2 \leqslant \sqrt{\frac{f(x^k) - f(x^*)}{\mu}} = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$$

LASSO

Consider the following problem with $A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m$:

$$\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n, \lambda \geqslant 0}$$

This problem also has an alternate form:

$$\|Ax - b\|_2^2 \to \min_{\|x\|_1 \leqslant M}$$

In the second form, LASSO can be solved by the Frank-Wolfe algorithm.

Form equivalence

Generally speaking, both M and λ are free hyperparameters that are not strictly connected. Except for the KKT condition of complementary slackness:

$$\lambda(\|x\|_1 - M) = 0.$$

However, some works connect them (thus, creating equivalent problems) by constructing the so-called *the bounded norm cone*. For more info, see *Q. Denoyelle, V. Duval, G. Peyre, and E. Soubies, "The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy," Inverse Problems, vol. 36, p. 014001, Jan. 2020.*

Accelerated Nesterov scheme

Nesterov accelerated schemes are some sort of momentum based methods, where we increase the impulse of our descent step in each iteration of the minimization process. Also, it can be easily applied to composite minimiztion problem:

$$F(x) = f(x) + g(x) \to \min_{x \in \mathbb{R}^n},$$

where f is still "very good", while g is convex but not smooth.

It's clear that it corresponds to the LASSO problem with $f(x) = \|Ax - b\|_2^2$ and $g(x) = \lambda \|x\|_1$.

Algorithm

Algorithm 5: FISTA algorithm with constant step

Input : f convex with ∇f Lipschitz constant L; g convex
Output : $x_k \simeq \arg\min_{x\in\mathbb{R}^n} \{f(\mathbf{x}) + g(\mathbf{x})\}$
Initialization: $x_0 \in \mathbb{R}^n$

$\mathbf{y}_1 = \mathbf{x}_0$;

$t_1 = 1$;

for $k = 1, \dots$ **do**

$$x_k = \mathrm{prox}_{\frac{1}{L}g}\left(\mathbf{y}_k - \frac{1}{L}\nabla f(\mathbf{y}_k)\right)$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$\mathbf{y}_k = \mathbf{x}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_k - \mathbf{x}_{k-1})$$

end

where $prox_h(z) \in Argmin_{y\in\mathbb{R}^n}\left(h(y) + \frac{1}{2}\|z - y\|_2^2\right)$

Computation skips

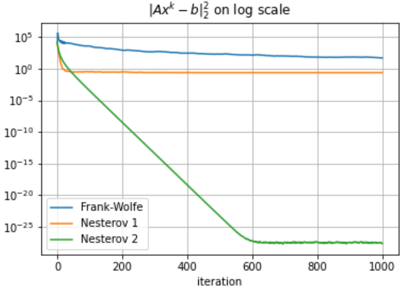
For the LASSO problem some computations can be obtained straightforward.

For Frank-Wolfe we need to minimize a linear function to obtain the auxiliary point. Thus, it is sufficient to check only corner points of the set $\{x : \|x\|_1 \leqslant M\}$: $\pm M\mathbf{e}_j,\ j \in \{1, .., n\}$.

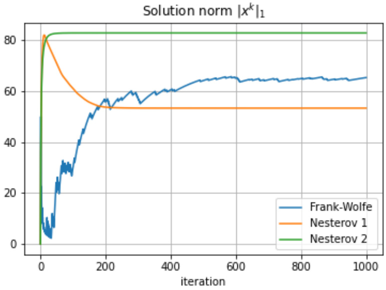
For the Nesterov accelerated method by substituting $h \to \frac{1}{L}\|\cdot\|_1$, we get $prox_{\frac{1}{L}g}(z) = sign(z) \cdot (|z| - \frac{\lambda}{L})_+$.

Experiment

$$A \in \mathbb{R}^{200\times500},\ A_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$
$$9x_k^* \overset{i.i.d.}{\sim} Unif(\{-1,1\}) + \mathcal{N}(0,0.1),\ \ 1 \leqslant k \leqslant 50$$



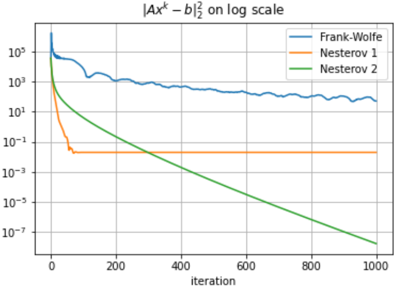
(a) Target function



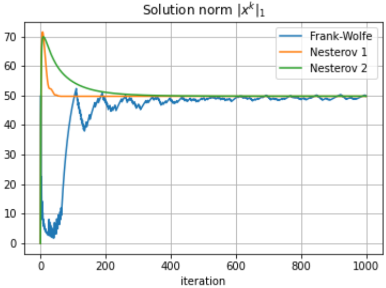
(b) Solution norm

Experiment

$$A \in \mathbb{R}^{700\times 500}, A_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$
$$9x_k^* \overset{i.i.d.}{\sim} Unif(\{-1,1\}) + \mathcal{N}(0,0.1),\ 1 \leqslant k \leqslant 50$$



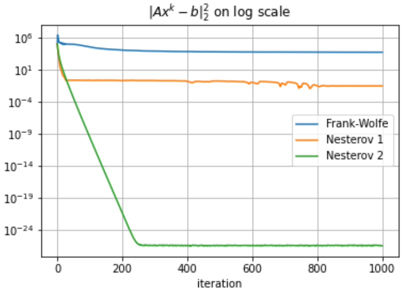
(a) Target function



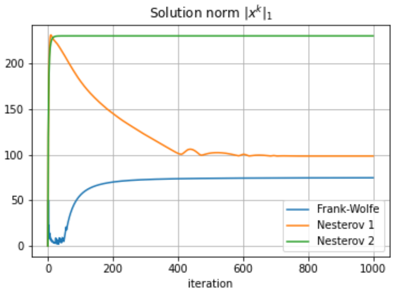
(b) Solution norm

Experiment

$$A \in \mathbb{R}^{1000\times4000},\, A_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$
$$9x_k^* \overset{i.i.d.}{\sim} Unif(\{-1,1\}) + \mathcal{N}(0,0.1),\ 1 \leqslant k \leqslant 50$$



(a) Target function

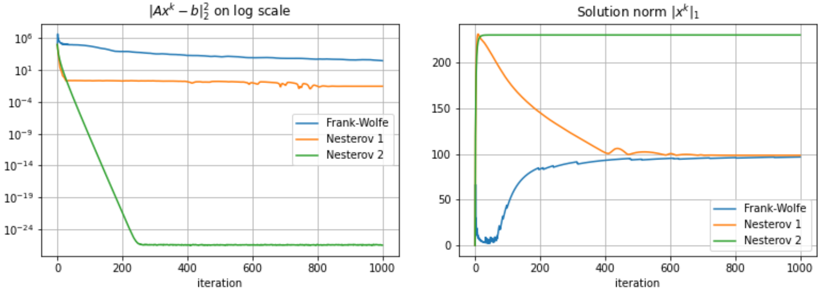


(b) Solution norm

Experiment

$$A \in \mathbb{R}^{1000\times 4000},\ A_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

$$9x_k^* \overset{i.i.d.}{\sim} Unif(\{-1,1\}) + \mathcal{N}(0,0.1),\ \ 1 \leqslant k \leqslant 50$$

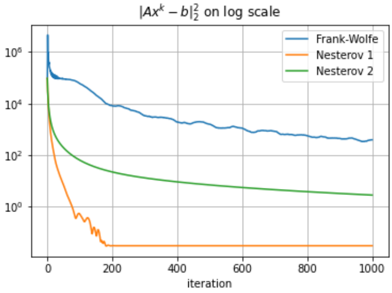


(a) Target function (b) Solution norm

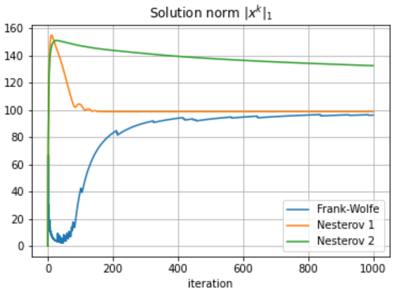
Figure: M is increased

Experiment

$$A \in \mathbb{R}^{1000\times 4000},\, A_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$
$$9x_k^* \overset{i.i.d.}{\sim} Unif(\{-1,1\}) + \mathcal{N}(0, 0.1),\ 1 \leqslant k \leqslant 50$$



(a) Target function



(b) Solution norm

Thanks!