LASSO Optimisation Project

Vsevolod Ivanov, Arkadiy Aliev

17.12.2024

Presentation plan

1. Problem statement
2. Nesterov method
3. Frank Wolfe method
4. Experiments
5. Conclusion

Problem statement

LASSO

The LASSO problem in its standard (primal) form is given as:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \|Ax - b\|^2 \quad \text{subject to} \quad \|x\|_1 \leq 1$$

where $A \in \mathbb{R}^{n \times d}$, $b \in \mathbb{R}^d$.
Here we observe, that the constraint set $X = \{x \in \mathbb{R}^d : \|x\|_1 \leq 1\}$ is the unit l_1-ball, the convex hull of the unit basis vectors: $X = conv(\{\pm e_1, \ldots, \pm e_d\})$

Nesterov's method

Nesterov's accelerated gradient method

Let L be parameter of Lipschitzness of the gradient of f and $\mu = \lambda_{min}(A^T A)$. Nesterov method works as follows:

1. Start at random point $x^{(0)}$, initialize $y^{(0)} = x^{(0)}$
2. Set $n = 100$, $\beta^{(k)} = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$ for strongly convex or $\frac{k}{k+3}$ for convex + L-smooth case
3. Repeat steps $4-5$ n times:
4. set $x^{(k+1)} = \Pi_{\|\cdot\|_1 \leq 1}(y^{(k)} - \frac{1}{L}\nabla f(y^{(k)}))$
5. set $y^{(k+1)} = x_k + \beta^{(k)}(x^{(k+1)} - x^{(k)})$
6. After n steps, return $x^{(n)}$

Convergence

$$f(x_k) - f(x_*) \leq \frac{L\|x_0 - x_*\|^2}{(k+1)^2}$$

Frank-Wolfe algorithm

Frank-Wolfe algorithm

1. Start at random point $x^{(0)}$
2. Set $n = 100$, $\gamma_k = \frac{2}{k+1}$
3. Repeat steps $4-5$ n times:
4. set $s^{(k+1)} = argmin_{z \in L_1} \langle \nabla f(x^{(k)}), z \rangle$
5. set $x^{(k+1)} = \gamma_k s + (1-\gamma_k) x^{(k)}$
6. After n steps, return $x^{(n)}$

Convergence

$$f(x_k) - f(x^*) \leq \frac{2L\mathrm{Diam}(B_{L_1})^2}{k+1}$$

Hypotheses

In general, the convergence rate of the Nesterov Algorithm will be faster than that of Frank-wolfe, however, when we take sparse matrices, then in Frank-Wolfe all multiplications are also performed for sparse matrices, that is, each iteration of Frank-Wolfe works faster. In the Nesterov method, at each step we calculate the projection of the vector onto the ball (projection of a sparse vector can be not sparse), which is why the matrices multiplied in the gradient cease to be sparse, namely:

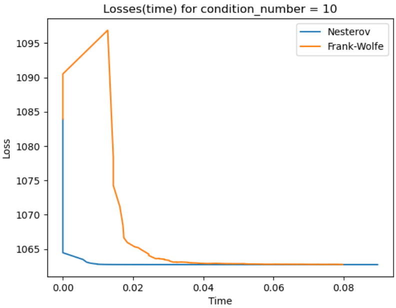
$$\nabla f(x) = A^T(Ax - b)$$

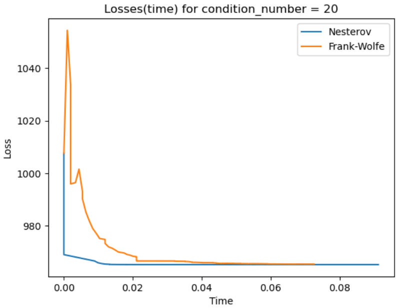
Here, vector x in Nesterov is dense.

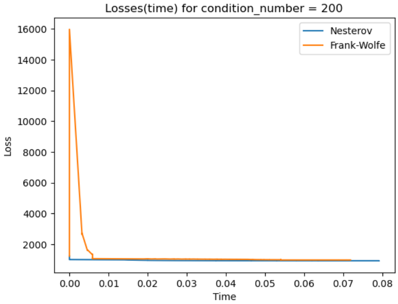
Design of experiments

1. For all experiments, we compare the dependence of the loss on time.
2. We generate random dense matrices and compare the performance of the algorithms for different condition numbers
3. We generate random sparse matrices with given densities and compare the performance of the algorithms on them

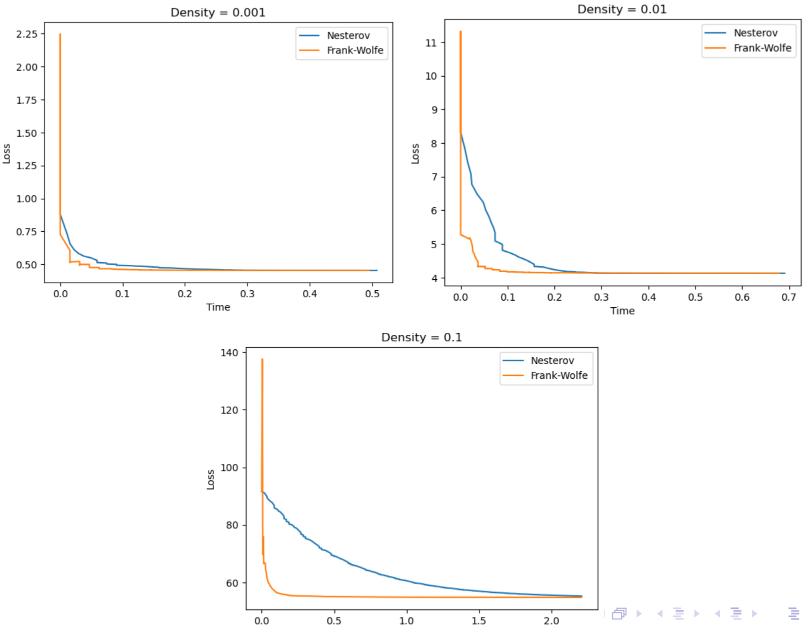
Experiments for dense matrix of size 1000 × 100







Experiments for sparse matrix of size 2000×2000



Conclusion

The Nesterov algorithm will be preferable if you are solving the LASSO problem for dense matrices. However, if you are looking for a solution for sparse matrices, then the Frank-Wolfe algorithm will be more efficient.