Optimization for Data Science
September 21, 2018

1. (6 POINTS) Describe in depth the PageRank problem and the related mathematical model. Furthermore, propose a possible algorithm for its solution.

   Solution 1. See Notes Section 4.7. A possible formulation, often used in practice, is the following:

   $$\min_{x \in \mathbb{R}^n} \|\hat{A}x - x\|^2 + \gamma[e^\top x - 1]^2$$

   with $\gamma > 0$ penalty parameter, and $\hat{A}$ the PageRank matrix. Taking into account that we deal with huge-scale instances, we might use the randomized block coordinate gradient descent method.

2. (7 POINTS) Describe in depth gradient method and accelerated gradient method, and explain the differences between the two.

   Solution 2. See Notes Section 5.6.5.

3. (7 POINTS) Given the set of samples $V = \{v_1, \ldots, v_n\}$, with $v_i \in \mathbb{R}^m, v_i \geq 0$, $i = 1, \ldots, n$ and n very large, consider the following problem:

   $$\min_{x \in conv\{v_1,\ldots,v_n\}} f(x) = \|Ax - b\|_2^2 + \lambda\|x\|_1,$$

   with $A \in \mathbb{R}^{l\times m},\ b \in \mathbb{R}^l$ and $\lambda > 0$. Describe the properties of the problem and propose an efficient algorithm for its solution.

   Solution 3. If we build the matrix

   $$M = [v_1 \ldots v_n] \in \mathbb{R}^{m\times n},$$

   we can rewrite the problem as follows:

   $$\begin{array}{ll} \min & f(My) = \|AMy - b\|_2^2 + \lambda e^\top (My) \\ s.t. & e^\top y = 1 \\ & y \geq 0. \end{array}$$

   It is easy to see that the problem has a convex and quadratic objective function. We can hence use either a Frank-Wolfe approach or the projected gradient in order to get a solution.

4. (8 POINTS) Consider the two sets:

   - $D_1 = \{x \in \mathbb{R}^n :\ a^\top x = b,\ 0 \leq x \leq 1\}$, with $a \in \mathbb{R}^n,\ b \in \mathbb{R},\ a, b > 0$;
   - $D_2 = \{x \in \mathbb{R}^n : \|Qx\|_2 \leq 1\}$, with $Q \in \mathbb{R}^{n\times n}$ symmetric positive definite matrix.

   Describe how to calculate the Frank-Wolfe direction for the problems

   $$\min_{x\in D_1} f(x), \ \min_{x\in D_2} f(x).$$

   Furthermore give the computational costs for the two different tasks.

Solution 4. At each iteration, the Frank-Wolfe direction is

$$d_k = \hat{x}_k - x_k.$$

Hence, in order to get $\hat{x}_k$, we need to respectively solve problems:

$$\min_{x \in D_1} \nabla f(x_k)^\top x, \ \min_{x \in D_2} \nabla f(x_k)^\top x.$$

- $D_1 = \{x \in \mathbb{R}^n : \ a^\top x = b, \ x \geq 0\}$, with $a \in \mathbb{R}^n, \ b \in \mathbb{R}, \ a, b > 0$. In this case we have a polytope of vertices:

$$v_i = \frac{b}{a_i} e_i, \ i = 1, \dots, n.$$

  Taking into account the fundamental theorem of linear programming, we can focus on the vertices of the polytope D_1 and write

$$\min_i \nabla f(x_k)^\top v_i = \min_i \nabla_i f(x_k) \frac{b}{a_i} = \min_i \frac{\nabla_i f(x_k)}{a_i}.$$

  So, the solution of the Frank-Wolfe problem has a cost $\mathcal{O}(n)$.

- $D_2 = \{x \in \mathbb{R}^n : \|Qx\|_2 \leq 1\}$, with $Q \in \mathbb{R}^{n \times n}$ symmetric positive definite matrix. In this case, we consider the new variables $y = Qx$. Thus writing the problem as follows:

$$\begin{array}{ll} \min & \nabla f(x_k)^\top (Q^{-1} y) \\ s.t. & \|y\|_2 \leq 1. \end{array}$$

  It is easy to see that the above problem has a solution:

$$\hat{y}_k = -\frac{Q^{-1} \nabla f(x_k)}{\|Q^{-1} \nabla f(x_k)\|_2}.$$

  Hence, we get $\hat{x}_k = Q^{-1} \hat{y}_k$, and the cost for calculating the direction is $\mathcal{O}(n^2)$.

5. (8 POINTS) Consider the problem of solving the linear system:

$$Ax = b, \tag{1}$$

with $A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^m$. Let $m \geq n$ and assume A has full rank, so that there is a unique solution $x^* \in \mathbb{R}^n$. The problem can be rewritten as

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \|Ax - b\|_2^2. \tag{2}$$

Show[1] that the gradient method $x_{k+1} = x_k - \alpha \nabla f(x_k)$, with stepsize

$$\alpha = \frac{1}{\sigma_{max}(A)^2},$$

[1] Keep in mind that

- $\sigma_{max}(A) = \|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$ and $\sigma_{min}(A) = \min_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$;
- For a given x, we have $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$;
- $\sigma_{max}(I - \alpha A^\top A)^2 = 1 - \alpha \sigma_{min}(A)^2$.

applied to Problem (2) has the following rate:

$$\|x_{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\sigma_{min}(A)^2}{\sigma_{max}(A)^2}\right) \|x_k - x^*\|_2^2.$$

Solution 5. We first consider the fact that $Ax^* = b$. then we can write the gradient:

$$\nabla f(x) = A^\top (Ax - b) = A^\top (Ax - Ax^*) = A^\top A(x - x^*).$$

Then we replace the expression in the gradient step

$$x_{k+1} = x_k - \alpha A^\top A(x_k - x^*),$$

thus subtracting x^* on both sides we get:

$$x_{k+1} - x^* = x_k - x^* - \alpha A^\top A(x_k - x^*) = (I - \alpha A^\top A)(x_k - x^*).$$

Taking the squared norm on both sides, and keeping in mind what is written in the footnotes, we get

$$\|x_{k+1} - x^*\|_2^2 \leq \|I - \alpha A^\top A\|_2^2 \|x_k - x^*\|_2^2 \leq \sigma_{max}(I - \alpha A^\top A)^2 \|x_k - x^*\|_2^2 = (1 - \alpha \sigma_{min}(A)^2) \|x_k - x^*\|_2^2.$$

Now we just need to replace the stepsize α to get our result:

$$\|x_{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\sigma_{min}(A)^2}{\sigma_{max}(A)^2}\right) \|x_k - x^*\|_2^2.$$