Optimization for Data Science
February 10, 2020

1. (8 POINTS) Describe in depth the randomized block coordinate gradient descent method and the Gauss-Southwell BCGD method highlighting their PROs and CONs. Furthermore, analyze the differences between the two methods.

   Solution 1. See Notes Section 4.5.

2. (7 POINTS) Describe in depth the Frank-Wolfe Algorithm.

   Solution 2. See Notes Section 5.3.

3. (7 POINTS) Describe the Page Rank problem and its mathematical model. Furthermore, propose a method to solve it.

   Solution 3. See Notes Subsection 5.6.5. In case we consider the unconstrained formulation, a randomized block coordinate gradient descent method can be used to get a solution.

4. (6 POINTS) Consider the problem of projecting a point $v \in \mathbb{R}^n$ over the ℓ_1-ball:

$$\begin{array}{ll} \min\limits_{w \in \mathbb{R}^n} & \|w - v\|_2^2 \\ \text{s.t.} & \|w\|_1 \leq r \end{array} \tag{1}$$

   with $r > 0$. Let w be an optimal solution of Problem (1). Prove that for all $i = 1, \dots, n$, we have $w_i v_i \geq 0$.
   Taking into account the theoretical result described above, and assuming that an efficient procedure for projecting over the simplex

$$\Delta = \{w \in \mathbb{R}^n : \ e^\top w = r, \ w \geq 0\}$$

   is available, describe a method for efficiently projecting over the ℓ_1-ball.

   Solution 4. We first prove the theoretical result. Assume by contradiction that the claim does not hold. Thus, we have an index i such that

$$w_i v_i < 0.$$

   Let $\tilde{w}$ be a vector such that $\tilde{w}_i = 0$ and for all $j \neq i$ we have $\tilde{w}_j = w_j$. We hence get that

$$\|\tilde{w}\|_1 = \|w\|_1 - |w_i| \leq r,$$

   and $\tilde{w}$ is feasible for the problem. We further can write

$$\|w - v\|_2^2 - \|\tilde{w} - v\|_2^2 = (w_i - v_i)^2 - (0 - v_i)^2 = w_i^2 - 2w_i v_i > w_i^2 > 0,$$

   thus contradicting optimality of w and proving our result. Defining projection over the ℓ_1-ball is then pretty straightforward. We call $P_\Delta(v)$, $P_{\ell_1}(v)$ the projection of vector v over the unit simplex and ℓ_1-ball, respectively. We have that

$$P_{\ell_1}(v) = \begin{cases} v & \text{if } \|v\|_1 \leq r \\ \text{sgn}(w) \cdot x & \text{otherwise,} \end{cases}$$

   where the product $\cdot$ is intended component-wise and $x = P_\Delta(|w|)$.

5. (8 POINTS) Considering $r_k = -\nabla f(x_k)$, prove that the Away-step Frank-Wolfe direction d_k^{ASFW} and the Pairwise Frank-Wolfe direction d_k^{PW} satisfy a relation of the following form

$$
{r_k}^\top d_k^{ASFW} \geq c \cdot {r_k}^\top d_k^{PW},
$$

with $c > 0$ suitably chosen (please specify c). Furthermore, consider the problem

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^n} f(x) \\
&\text{s.t.} \quad x \in \Delta,
\end{aligned} \tag{2}
$$

with Δ the unit simplex. Prove that at iteration k of the Away-step Frank-Wolfe method, when $(x_k)^i = 0, i \in 1, \ldots, n$ and ${r_k}^\top(e_i - x_k) < 0$, then $(x_{k+1})^i = 0$.

Solution 5. First part is very easy to prove. We just need to use the definition of the two directions to get:

$$
2 \cdot {r_k}^\top d_k^{ASFW} \geq {r_k}^\top (d_k^{AS} + d_k^{FW}) = {r_k}^\top d_k^{PW},
$$

thus getting

$$
{r_k}^\top d_k^{ASFW} \geq \frac{1}{2} {r_k}^\top d_k^{PW}.
$$

Now, we prove the second part. By considering the fact that $(x_k)^i = 0$, we surely cannot choose the vertex e_i to define the away-step direction. Furthermore, since ${r_k}^\top(e_i - x_k) < 0$, $d_k^{FW} = e_i - x_k$ cannot be chosen as the search direction at step k as well. This guarantees that $(x_{k+1})^i = 0$.