Optimization for Data Science
June 26, 2018

1. (6 POINTS) Describe in depth the Gauss-Southwell Block-Coordinate Gradient Descent method.
2. (7 POINTS) Describe in depth Away-Step, Pairwise and Fully Corrective Frank-Wolfe. Furthermore, prove that Fully Corrective Frank-Wolfe converges in a finite number of steps (when minimizing a convex function over a polytope).
3. (7 POINTS) Consider the problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2,$$

with $A \in \mathbb{R}^{m \times n}$ and $m \leq n$. Calculate the gradient related to the objective function, then explain the differences, in terms of computational cost per iteration, between the classic gradient method and a BCGD method (with blocks of dimension 1) when solving the problem.

4. (8 POINTS) Consider the problem:

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2}\|x\|^2 + \max_{i \in I}\{a_i + c_i^\top x\}$$

where $I = \{1, \dots, m\}$, $a_i \in \mathbb{R}$ and $c_i \in \mathbb{R}^n$. Describe the Lagrangian dual of the considered problem and give a possible primal solution.

5. (8 POINTS) Consider the problem

$$\max_{x \in P} f(x),$$

with $P \subseteq \mathbb{R}^n$ non-empty polytope and $f(x)$ continuously differentiable convex function. Consider the Frank-Wolfe variant described in Algorithm 1. Prove that the algorithm converges in a finite number of iterations to a point satisfying optimality conditions[1].

Algorithm 1 `Frank-Wolfe for maximizing a convex function over a polytope`

Choose a point $x_1 \in P$
For $k = 1, \dots$
  Set $\hat{x}_k = \arg\max_{x \in P} \nabla f(x_k)^\top (x - x_k)$
  If $\nabla f(x_k)^\top (\hat{x}_k - x_k) = 0$, then STOP
  Set $x_{k+1} = \hat{x}_k$
End for

[1] Keep in mind that for the considered problem a result similar to the fundamental theorem of linear programming holds. Indeed, at least one of vertex of the polytope is a global maximizer for f over P.