Optimization for Data Science
September 20, 2019

1. (6 POINTS) Describe in depth the stochastic variance reduced gradient methods.
2. (7 POINTS) Describe in depth the Frank-Wolfe method and its main variants.
3. (8 POINTS) Given the following problem:

$$\min_{x \in \Delta} \frac{1}{2} x^\top Q x + c^\top x,$$

with $\Delta = \{x \in \mathbb{R}^n : e^\top x = 1,\ x \geq 0\}$ and $Q \in \mathbb{R}^{n \times n}$ positive definite matrix. Calculate the computational cost of performing the exact line search at a point $x_k \in \Delta$ when using the Pairwise Frank-Wolfe direction (assume that the gradient is given).

4. (7 POINTS) Consider the Boosting problem and explain PROs and CONs of using the Gradient method for solving it and calculate the gradient at a point w_k. Suggest an alternative solution method (please motivate the answer).
5. (4 POINTS) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on $\mathbb{R}^n$ and bounded from below. Let $\{x_k\}$ be an infinite sequence such that for all k we have

   (a)
   $$\nabla f(x_k)^\top d_k < 0;$$

   (b)
   $$f(x_{k+1}) = f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \alpha_k \nabla f(x_k)^\top d_k$$
   with $\gamma \in (0, 1)$;

   (c) there exists a value $\mu > 0$ such that
   $$\alpha_k \geq \mu \frac{|\nabla f(x_k)^\top d_k|}{\|d_k\|^2}.$$

   Prove that
   $$\sum_{k=0}^{\infty} \left(\frac{\nabla f(x_k)^\top d_k}{\|d_k\|} \right)^2 < \infty.$$