Optimization for Data Science
June 4, 2018

1. (7 POINTS) Describe the gradient method and prove that it converges at a linear rate in the strongly convex case.

2. (6 POINTS) Describe in depth the stochastic gradient approaches that exploit variance reduction and explain the reasons why those methods work better than the classic stochastic gradient.

3. (7 POINTS) Consider the problem
$$\min_{x\in\Delta} f(x),$$
with f continuously differentiable and convex function and
$$\Delta = \{x \in \mathbb{R}^n : e^\top x = 1,\ x \geq 0\}.$$
Calculate the maximum stepsize that can be taken at a point x_k along the away-step direction (i.e., d_k^{AS}) and the pairwise Frank Wolfe direction (i.e., $d_k^{PWFW} = d_k^{FW} + d_k^{AS}$).

4. (8 POINTS) Consider the mean-risk problem:
$$\min_{x\in\Delta} \quad \gamma\sqrt{x^\top M x} - c^\top x$$
where $M \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, $c \in \mathbb{R}^n$, $\gamma > 0$ is the risk-aversion parameter and
$$\Delta = \{x \in \mathbb{R}^n : e^\top x = 1,\ x \geq 0\}.$$
Analyze in depth its properties. Then describe a method for finding an optimal solution and properly justify the choice.

5. (8 POINTS) Consider the problem
$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}x^\top Q x + c^\top x,$$
with $Q \in \mathbb{R}^{n\times n}$ symmetric positive definite matrix and $c \in \mathbb{R}^n$. Let $\gamma_1, \dots, \gamma_n$ be the eigenvalues of the matrix Q. Prove that the gradient method defined as follows:
$$x_{k+1} = x_k - \frac{1}{\gamma_k}\nabla f(x_k),$$
with $k \geq 1$, converges in at most n iterations to the solution of the problem.