Optimization for Data Science

September 4, 2019

1. (7 POINTS) Describe the BCGD method with Gauss-Southwell rule and prove that it converges at a sublinear rate in the 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
$$\Delta = \{x \in \mathbb{R}^n : a^\top x = b,\ x \geq 0\},$$
with $a \in \mathbb{R}^n, a \geq 0$ and $b > 0$. 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
$$\Delta = \{x \in \mathbb{R}^n : a^\top x = b,\ x \geq 0\},$$
with $a \in \mathbb{R}^n, a \geq 0$ and $b > 0$. Analyze in depth its properties. Then describe a method for finding an optimal solution and properly justify the choice.

5. (4 POINTS) Consider the gradient method with stepsize $\alpha_k = 1/L$. Prove that
$$g_N^* \leq \frac{1}{\sqrt{N+1}} \left[2L(f(x_0) - f^*)\right]^{1/2},$$
with
$$g_N^* = \min_{1\leq k\leq N} \|\nabla f(x_k)\|,$$
and f^* optimal value of the function.