Optimization for Data Science
July 4, 2019

1. (6 POINTS) Describe in depth the stochastic gradient approach and explain why it is widely used in data science.
2. (7 POINTS) Describe in depth block coordinate gradient descent with Gauss-Southwell rule and randomized block coordinate gradient descent method, highlighting the differences between the two methods.
3. (7 POINTS) Consider the ridge regression problem:

$$\min_{w\in\mathbb{R}^d} f(w) = \frac{1}{2n}\|X^\top w - y\|_2^2 + \frac{\lambda}{2}\|w\|_2^2,$$

   Describe the gradient of the function and prove that there exists a matrix C s.t. $\nabla f(w) = C(w - w^*)$, with w^* solution of the problem.
4. (8 POINTS)Consider the two sets:
   - $D_1 = \{x \in \mathbb{R}^n : \ a \leq x \leq b, \}$, with $a, b \in \mathbb{R}^n$ and $a < b$;
   - $D_2 = \{x \in \mathbb{R}^n : \|x\|_1 \leq \tau\}$, with $\tau > 0$.

   Describe how to calculate the Frank-Wolfe direction for the problem

$$\min_{x\in D_1} f(x).$$

   Furthermore give the computational cost. Finally, consider problem

$$\min_{x\in D_2} f(x),$$

   and the origin point $x_k = 0$. Calculate the Away-Step direction in x_k choosing the minimal representation w.r.t. S_k (i.e., the one with the smallest number of nonzeroes) and the related maximum stepsize.
5. (4 POINTS) Calculate the condition number[1] of the matrix C obtained at exercise 3 and explain what happens when $\lambda \to 0$ and $\lambda \to \infty$.

[1]Keep in mind that the condition number is the ratio $\kappa(C) = \frac{\sigma_{max}(C)}{\sigma_{min}(C)}$ and that

$$\sigma_{max}(C) = \|C\|_2 = \max_{x\neq 0} \frac{\|Cx\|_2}{\|x\|_2} \text{ and } \ \sigma_{min}(C) = \min_{x\neq 0} \frac{\|Cx\|_2}{\|x\|_2}.$$