

Convex Optimization

Problem set 5

Due Wednesday June 9

1. Consider the projected sub-gradient descent method:

$$x^{(k+1)} \leftarrow \Pi_G(x^{(k)} - t^{(k)}\nabla f(x^{(k)})) \quad (1)$$

where Π_G is a projection into the feasible set. In class we showed that with a specific choice of step sizes $t^{(k)}$, the output:

$$\tilde{x}^{(k)} = \arg \min_{x^{(i)}, i=0, \dots, k} f(x^{(i)})$$

converges to the optimum, with suboptimality after k iterations bounded by $O(\sqrt{\frac{B^2 L^2}{k}})$, where $\|x\| \leq B$ for all $x \in G$ and f is L -Lipschitz.

- (a) Prove that if the domain is bounded and f is Lipschitz, then with any choice of stepsizes such that

$$t^{(k)} \xrightarrow{k \rightarrow \infty} 0 \quad \text{and} \quad \sum_{i=1}^k t^{(i)} \xrightarrow{k \rightarrow \infty} \infty, \quad (2)$$

the output of projected sub-gradient descent will converge to the optimum, i.e.

$$f(\tilde{x}^{(k)}) - \inf_{x \in G} f(x) \xrightarrow{k \rightarrow \infty} 0 \quad (3)$$

- (b) Even if the conditions (2) are satisfied, the rate of convergence might be very bad, and in particular much worse than $O(\frac{B^2 L^2}{k})$. Consider for example step sizes of $t^{(k)} = 1/k$.
- Verify that these stepsizes satisfy (2).
 - What is the bound on the suboptimality after k iterations that you can obtain using these stepsizes? What is the resulting bound on the number of stepsizes to attain suboptimality ϵ ?
 - Show a simple one dimensional function $f : [0, 1] \rightarrow \mathbb{R}$, where using projected sub-gradient descent with these stepsizes does in-fact required an exponential number of steps to reach ϵ -suboptimality.
- (c) **[Optional]** More generally consider stepsizes of the form $t^{(k)} = 1/k^p$.
- For what values of p is Gradient Descent guaranteed to converge?
 - Analyze the bound on the number of iterations required to reach ϵ -suboptimality as a function of p . What is the optimal p ? What happens when p is very small?

2. Recall that by definition, a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is L -Lipschitz with respect to norm $\|\cdot\|$ if for every $x, y \in \mathbb{R}^d$:

$$|f(x) - f(y)| \leq L \|x - y\|.$$

Prove that a function is L -Lipschitz with respect to norm $\|\cdot\|$ if and only if $\|\nabla f(x)\|_* \leq L$ for all subgradients $\nabla f(x)$ at all points x , where $\|\cdot\|_*$ is the dual norm to $\|\cdot\|$.

3. As discussed in class, in order to execute the Mirror Descent algorithm, we have to be able to efficiently solve optimization problems of the form:

$$\min_{x \in G} \langle p, x \rangle + \omega(x) \tag{4}$$

where G is the domain and $\omega(x)$ is the prox function (distance generating function). We already saw that this problem has a simple closed form when G is a Euclidean ball and $\omega(x) = \frac{1}{2} \|x\|_2^2$. In this problem we will consider the prox function:

$$\omega(x) = \sum_{i=1}^n x_i \log x_i$$

- (a) For the domain $G = \{x | x_i \geq 0, \|x\|_1 = 1\}$ and $\omega(x)$ as above, write down the optimization problem (4) as a constrained optimization problem with a single explicit equality constraint (note that the definition of $\omega(x)$ implies implicit inequality constraints).
- (b) Write down the Lagrangian and the KKT conditions for the problem.
- (c) Use the KKT conditions to solve the optimization problem. Write down the expression for the optimal x as a function of p in closed form.
- (d) Use this to write down the step taken at each iteration of Mirror Descent. I.e. write down an explicit and easy to compute expression for $x^{(k+1)}$ as a function of $x^{(k)}$, the subgradient $\nabla f(x^{(k)})$, and the stepsize $t^{(k)}$.
- (e) **[Optional]** Explain how you would solve (4) and implement a Mirror Descent step for $G = \{x | x_i \geq 0, \|x\|_1 \leq 1\}$ with the same prox function.