535520 Fall 2024: Optimization Algorithms (**Due: 2024/12/1, 21:00**)

Homework 2: Projected GD, Frank-Wolfe, and Mirror Descent

Submission Guidelines: Your deliverables shall consist of 2 separate files – (i) A PDF file: Please compile all your write-ups and your report into one .pdf file (photos/scanned copies are acceptable; please make sure that the electronic files are of good quality and reader-friendly); (ii) A zip file: Please compress all your source code into one .zip file. Please submit your deliverables via E3.

Problem 1 (Convergence of PGD for Convex and Smooth Problems) (5+5+5+5+5=25 points)

As discussed in Lecture 8, we mentioned that for a convex and L-smooth function, the convergence rate of PGD is

$$f(x_t) - f(x^*) \leq \frac{3L\|x_0 - x^*\|^2 + \big(f(x_0) - f(x^*)\big)}{t+1}. \tag{1}$$

In this problem, let us prove this result formally in a step-by-step manner.

(a) To begin with, let us show the following lemma: For any $x, z \in C$, let $\bar{x} = \prod_C(x - \frac{1}{L}\nabla f(x))$ and $g_C(x) = L(x - \bar{x})$ (note that here we basically reuse the same notation as in our lecture slides). Then, we have

$$f(z) \geq f(\bar{x}) + g_C(x)^\top (z - x) + \frac{1}{2L}\|g_C(x)\|^2. \tag{2}$$

(Hint: Consider $f(z) - f(\bar{x}) = (f(z) - f(x)) - (f(\bar{x}) - f(x))$ and then utilize the convexity and smoothness conditions)

(b) By using the result in (a), show that the one-step improvement can be written as

$$f(x_{t+1}) - f(x_t) \leq -\frac{1}{2L}\|g_C(x_t)\|^2. \tag{3}$$

(c) Next, let us connect $\|g_C(x_t)\|$ to the objective function $f(x_t)$: Show that

$$\|g_C(x_t)\| \geq \frac{f(x_{t+1}) - f(x_t)}{\|x_t - x^*\|}. \tag{4}$$

(Hint: Find a proper way to apply the result in (a) and use Cauchy-Schwarz inequality)

(d) Define the sub-optimality gap at the t-th iteration as $\Delta_t := f(x_t) - f(x^*)$. By using the results in (a)-(c), show that $\Delta_{t+1} - \Delta_t \leq \frac{-\Delta_{t+1}^2}{2L\|x_0 - x^*\|^2}$.

(e) Finally, by using the result in (d), use an induction argument to show the convergence rate of PGD in (1).

Problem 2 (Projection Theorem) (10+10=20 points)

In this problem, let's prove the fundamental Projection Theorem, which typically involves two very useful properties as follows:

(a) Let C be a convex set. Given some vector $x \in \mathbb{R}^d$, a vector $x_C \in C$ is equal to the projection $\Pi_C(x)$ if and only if

$$(x - x_C)^\top (z - x_C) \leq 0, \quad \forall z \in C. \tag{5}$$

(Hint: You just need to prove this in both directions by leveraging the FONC-C and the basic geometric properties of projection)

(b) Given a convex set C, the projection operator $\Pi_C : \mathbb{R}^d \to C$ is non-expansive, i.e.,

$$\|x_C - z_C\| \leq \|x - z\|, \quad \forall x, z \in \mathbb{R}. \tag{6}$$

Problem 3 (Bregman Divergence) (5+10+10=25 points)

Recall from Lecture 10 that we learned the Bregman divergence, which is a key component in Mirror Descent. In this problem, you will have the oppotunity to verify a few useful properties:

(a) $\nabla_y D_\phi(y\|x) = \nabla\phi(y) - \nabla\phi(x)$.

(b) For any strictly convex functions $\phi_1 : X \to \mathbb{R}, \phi_2 : X \to \mathbb{R}$ and $\lambda \geq 0$, we have

$$D_{\phi_1+\lambda\phi_2}(y\|x) = D_{\phi_1}(y\|x) + \lambda D_{\phi_2}(y\|x) \tag{7}$$

(c) As mentioned in Lecture 10, please show that the Bregman divergence satisfies the Generalized Pythagorean Theorem, i.e.,

$$D_\phi(z\|x) \geq D_\phi(z\|\bar{x}) + D_\phi(\bar{x}\|x), \tag{8}$$

where $\bar{x}$ is the Bregman projection of x onto the feasible set C.

Problem 4 (Gurobi Optimization Solver for Frank-Wolfe) (30 points)

In this homework, you will leverage the Gurobi Optimization Solver again to solve constrained optimization problems. Specifically, let us implement Frank-Wolfe method with the help of Gurobi optimization solver, which can be used through its Python API, and reproduce the performance of Frank-Wolfe in the top-left subfigure of Figure 2 of the paper (`https://arxiv.org/abs/2002.07003`) on a Portfolio Management dataset (provided to you on E3). Portfolio Management can be formulated as a constrained problem can be described as

$$\min_x \quad f(x) := -\sum_{i=1}^{n} \log(a_i^\top x) \tag{9}$$

$$\text{subject to} \quad \sum_{j=1}^{p} x_j = 1, x \geq 0, \tag{10}$$

where each a_i is a p-dimensional vector. To facilitate gradient computation, you may write your code in either PyTorch or TensorFlow. If you are a beginner in learning the deep learning framework, please refer to the following tutorials:

- PyTorch: `https://pytorch.org/tutorials/`
- Tensorflow: `https://www.tensorflow.org/tutorials`

For the introduction to the Python API for Gurobi Optimization Solver, please see:

- To use Gurobi, you need to install Gurobipy (`https://pypi.org/project/gurobipy/`)
- Examples can be found at `https://www.gurobi.com/documentation/`

For the deliverables, please submit the following:

- Technical report: Please summarize all your experimental results in 1 single report (and please be brief)
- All your source code