Homework 7 for Math 173B - Winter 2025

Questions:

(0) Write the names of all people you discussed this assignment with.

(0') Submit all your computer codes as part of this assignment. In particular, for each question, your code must be presented as part of your answer.

(1) A trick called splitting is often used to write certain optimization problems in a form amenable to ADMM implementation. Let A be an $m \times n$ matrix, let $\lambda \in \mathbb{R}$, and consider the (unconstrained) optimization problem

$$(P_1) \qquad \min_{x \in \mathbb{R}^n} f(Ax) + \lambda g(x).$$

The splitting trick consists of rewriting the above problem as

$$(P_2) \qquad \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} f(z) + \lambda g(x) \quad \text{subject to} \quad Ax - z = 0.$$

a) Why are (P_1) and (P_2) equivalent?

b) Derive an ADMM algorithm for solving (P_1). You should use the equivalent problem (P_2).

(2) Consider, now, a data classification/regression setting where one has data $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, ..., N$, and wishes to solve the optimization problem, over the variables $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$, given by

$$\min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \sum_{i=1}^{N} \frac{1}{2}(w^T x_i + b - y_i)^2 + \frac{\lambda}{2}\|w\|_2^2.$$

(a) Write down an equivalent form of the above problem, using the splitting – in the spirit of problem (1) – given by

$$z_i = w^T x_i + b.$$

(b) Write down the augmented Lagrangian associated with your equivalent problem from part (a).

(c) Derive an ADMM based algorithm for solving the problem.

(3) Consider, again, a data classification/regression setting where one has data $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, ..., N$, and wishes to solve the optimization problem, over the variables $w \in \mathbb{R}^d$ given by

$$\min_{w} \frac{1}{N} \sum_{i=1}^{N} \log\left(1 + e^{-w^T x_i y_i}\right) + \frac{\lambda}{2}\|w\|^2. \tag{1}$$

(a) This time, using the equivalent formulation

$$\min_{w,z} \frac{1}{N} \sum_{i=1}^{N} \log\left(1 + e^{-w^T x_i y_i}\right) + \frac{\lambda}{2}\|z\|^2 \quad \text{s.t.} \quad w - z = 0 \tag{2}$$

write an ADMM based algorithm for solving the optimization problem. You may not be able to find a closed form expression for some of the variable updates, so you should replace these with optimization algorithms of your choosing.

(b) Implement and apply your algorithm, as in Assignment 3, for classifying MNIST digits (it's enough to work with two digits). You may select your own stopping criterion.

(i) Provide a plot showing $F(w, z) := \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + e^{-w^T x_i y_i}\right) + \frac{\lambda}{2}\|z\|^2$ at each iteration, and another plot showing $\|w - z\|$ at each iteration. You may want to use log-scale for the y-axis.

(ii) Comment on the performance of your algorithm. Your comments should address convergence as a function of number of iterations, as well as the computational complexity of each iteration.

(iii) Now, use the w you found from part (i) to classify the first 500 *test* data points associated to each of the two handwritten digits. Recall that you need to use the function $y = \text{sign}(w^T x)$ to classify. What was the classification error rate associated with the two digits on the test data (this should be a number between 0 and 1)? What was it on the training data?

(iv) What are the potential advantages and disadvantages, compared to other methods from this course, of using ADMM (or one of its variants) for solving this problem?