Homework Assignment 1:

Implementation of Basic Large Scale Optimization Algorithms

ELEC-E5431 – Large Scale Data Analysis (LSDA)

- Return your report preferably in pdf-format. Recommended deadline to return your report is by **February 15, 2024** through MyCourse.
- Enclose you codes as well!!!
- In subject line: write ELEC-E5431, your name and student ID.
- Show all the steps of your work. Your reasoning is the most important component for grading in addition to the code!
- It is preferable if you code in Python or MATLAB, but it is not a limitation.

Problem 1 *[70/100 points for this assignment]*

The optimization problem to be addressed is a simple quadratic function minimization, that is,

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x}$$

where the matrix $\mathbf{A}$ and vector $\mathbf{b}$ are appropriately generated. Use the same $\mathbf{A}$ and $\mathbf{b}$ while comparing different methods.

Hints:

- $\mathbf{A}$ should be such that $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is non-negative, that is, $\mathbf{A}$ is positive semi-definite, and $\mathbf{b}$ should be in the range of $\mathbf{A}$.

- In fact, by solving the above unconstrained minimization problem, you solve a system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$. Indeed, the gradient of the objective function is $\mathbf{A}\mathbf{x} - \mathbf{b}$, and it should be equal to 0 at optimality. Thus, you can find optimal $\mathbf{x}^*$ using back-slash (or matrix inversion followed by computing the product $\mathbf{A}^{-1}\mathbf{b}$) operators in MATLAB. The optimal objective value can be then obtained by simply substituting such $\mathbf{x}$ into the objective of the above optimization problem. It is suitable for small and mid size problems, but the matrix inversion is prohibitively too expensive to be able to solve a system of linear equations for large scale problems. Thus, the only option for large scale problems is the use of algorithms that you implement in this assignment!

To be able to produce convergence figures for the algorithms that you test, let the dimension of $\mathbf{x}$ be 100 variables or few 100's (but after producing the figures also play with higher dimensions to see when the matrix inversion fails, but the large scale optimization methods still work fine and some also quite fast).

Set the tolerance parameter for the stopping criterion for checking the convergence to 10^{-5}. For example, check if $\|\nabla f(\mathbf{x})\| \leq 10^{-5}$, and limit the total number of iterations by 5000 if the predefined tolerance is still not achieved.

Task 1: *Gradient Descent Algorithm.* Implement Gradient Descent Algorithm for solving the above optimization problem. Use correctly selected fixed step size. Draw the experimental convergence rate, i.e., draw the convergence plot of $\log |f(\mathbf{x}^t) - f^*|$ versus iteration count t, for the algorithm, and compare it to the theoretically predicted one.

Because, we know that

$$f(\mathbf{x}^t) - f^* \leq \frac{1}{2\mu} \|\nabla f(\mathbf{x}^t)\|^2$$

you can equivalently draw

$$\log \|\nabla f(\mathbf{x}^t)\| = \log \|\mathbf{A}\mathbf{x}^t - \mathbf{b}\|$$

versus iteration count t.

Task 2: *Conjugate Gradient Algorithm.* Implement Conjugate Gradient Algorithm (page 11 in Summary Notes) for solving the above optimization problem. Draw the experimental convergence rate for the algorithm, and compare it to the theoretically predicted one.

Task 3: *FISTA.* Implement FISTA (page 12 in Summary Notes) for solving the above optimization problem. Draw the experimental convergence rate for the algorithm, and compare it to the theoretically predicted one.

Task 4: *Coordinate Descent.* Implement the Coordinate Descent method (page 20 in Summary Notes) for solving the above optimization problem. Draw the experimental convergence rate for the algorithm, and compare it to the theoretically predicted one.

Task 5: *Comparisons and Conclusion.* Compare the results (in terms of the iterations required and the overall computation time) for different methods (including, for example, the standard MATLAB back-slash operator) and draw your overall conclusions. Observe up to which dimension Python or MATLAB still can invert a matrix, that is, define the dimension after which the problem turns to be large scale in the context of your implementation.

Problem 2 *[15/100 points for this assignment]*

Projection onto intersection of convex sets. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be closed convex sets such that $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \emptyset$.

Give a complete proof that for the following problem

$$\text{find } \mathbf{x} \in \mathcal{C}_1 \cap \mathcal{C}_2$$

the subgradient method with Polyak's stepsize rule is equivalent to the following alternating projections:

$$\mathbf{x}^{t+1} = \mathcal{P}_{\mathcal{C}_1}(\mathbf{x}^t), \quad \mathbf{x}^{t+2} = \mathcal{P}_{\mathcal{C}_2}(\mathbf{x}^{t+1}).$$

Almost complete proof can be found in the lecture notes on Subgradient Methods, but explain it in more detail (how you understand it) and prove also (by straightforward calculation) the equality

$$\nabla \left(\frac{1}{2} \text{dist}^2_{\mathcal{C}_i}(\mathbf{x}^t) \right) = \mathbf{x}^t - \mathcal{P}_{\mathcal{C}_i}(\mathbf{x}^t)$$

that is not proven there. Here $\text{dist}_{\mathcal{C}}(\mathbf{x}) = \min_{\mathbf{z} \in \mathcal{C}} \|\mathbf{x} - \mathbf{z}\|_2$.

Problem 3 *[15/100 points for this assignment]*

Generalized Pythagorean Theorem states that

$$D_\varphi(\mathbf{z}, \mathbf{x}) \geq D_\varphi(\mathbf{z}, \mathbf{x}_{\mathcal{C},\varphi}) + D_\varphi(\mathbf{x}_{\mathcal{C},\varphi}, \mathbf{x})$$

where $D_\varphi(\mathbf{z}, \mathbf{x})$ is Bregman divergence between points $\mathbf{z}$ and $\mathbf{x}$ for generating function φ and $\mathbf{x}_{\mathcal{C},\varphi} = \mathcal{P}_{\mathcal{C},\varphi}(\mathbf{x}) = \arg\min_{\mathbf{z} \in \mathcal{C}} D_\varphi(\mathbf{z}, \mathbf{x})$ is Bregman projection of point $\mathbf{x}$ to set $\mathcal{C}$.

Prove that if $\mathcal{C}$ is affine plane, then Generalized Pythagorean Theorem holds as equality. Prove also that Generalized Pythagorean Theorem is equivalent to Pythagorean Theorem if in addition $D_\varphi(\mathbf{z}, \mathbf{x}) = \text{dist}^2_{\mathcal{C}}(\mathbf{x})$.