DSA5103 Assignment 2

Instructions While all languages are acceptable, it is recommended that you code using Python or MATLAB. You must write your own code. Due on March 12, 11:59pm. Please submit a pdf file including answers for Q1, Q2, Q3(c,d,e,f). Insert the figures for Q3(c,d) into the pdf file, report the results for Q3(e) in the form of Table 1 in the pdf file, and summarize your comparisons for Q3(f) in the pdf file. In addition, you should submit the codes for Q3 (.m file or .ipynb file) that I can run for generating the numerical results reported in the pdf file. No programming is required for Q1 and Q2.

1. **KKT**

   Consider the problem ($x \geq 0$ means $x_i \geq 0$, $i \in [n]$)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & -\sum_{i=1}^{n} \log(1 + x_i) \\ \text{s.t.} \quad & x \geq 0 \\ & x_1 + \cdots + x_n = 1. \end{aligned}$$

   (a) Check Slater's condition.

   (b) Write KKT conditions.

2. **Coordinate descent**

   Apply coordinate descent method for

$$\min_{x=(x_1;x_2) \in \mathbb{R}^2} \quad f(x_1, x_2) = 2x_1^2 - 6x_1x_2 + 5x_2^2 - 4x_1 - 3x_2$$

   with initial point $x^{(0)} = (0; 0)$. Find $x^{(1)}$ and $x^{(2)}$.

3. **Lasso**

   For solving the Lasso problem

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|X\beta - Y\|^2 + \lambda\|\beta\|_1,$$

we write our own codes to implement the methods: coordinate descent, proximal gradient, accelerated proximal gradient, and accelerated proximal gradient with restart techniques.

(a) Set $n = 1000$, $p = 5000$. Generate an $n \times p$ random matrix X where each entry X_{ij} is a random variable drawn from the standard normal distribution $N(0,1)$, and then standardize each column of X — subtracting its mean and dividing by its standard deviation. Generate a random sparse vector $\beta^* \in \mathbb{R}^p$ with approximately $0.05p$ nonzero entries (by "scipy.sparse.random" in Python or "sprandn" in MATLAB). Let $Y = X\beta^* + 0.01\epsilon$ where $\epsilon_i \sim N(0,1)$, $i \in [n]$ is the Gaussian noise.

(b) Set the penalty parameter be $\lambda = 0.1\|X^T Y\|_\infty$. We always use the initial point $\beta^{(0)} = 0$. Let $L = \lambda_{\max}(X^T X)$ ($\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix), step size $\alpha = 1/L$ for PG and APG. Implement the methods for solving the Lasso problem: coordinate descent (CD), proximal gradient (PG), accelerated proximal gradient (APG), and accelerated proximal gradient with restart techniques (APG-restart) (one may choose to restart every 100 iterations). The codes should be terminated when it achieves the accuracy of tolerance $= 10^{-3}$ in the relative residual error for an approximate solution $\beta^{(k)}$, namely,

$$r(\beta^{(k)}) := \left\|\beta^{(k)} - S_\lambda\left(\beta^{(k)} - X^T(X\beta^{(k)} - Y)\right)\right\|_2 < \text{tolerance}.$$

(c) For CD, PG, APG, APG-restart, plot a figure of the relative residual error in base-10 logarithmic scale, i.e., $\log_{10}(r(\beta^{(k)}))$, against iterations k.

(d) For CD, PG, APG, APG-restart, plot a figure of the relative residual error in base-10 logarithmic scale, i.e., $\log_{10}(r(\beta^{(k)}))$, against running time t_k, where t_k denotes the time taken for the first k iterations of a particular method.

(e) Test your codes for APG and APG-restart to see whether it can achieve the accuracy of tolerance $= 10^{-10}$ in the relative residual error for an approximate solution $\beta^{(k)}$, namely,

$$r(\beta^{(k)}) := \left\|\beta^{(k)} - S_\lambda\left(\beta^{(k)} - X^T(X\beta^{(k)} - Y)\right)\right\|_2 < 10^{-10}$$

within 3000 iterations. Report the relative residual error, iterations, and time for APG and APG-restart in Table 1.

	relative residual error $r(\beta^{(k)})$	iterations	time (sec)
APG			
APG-restart			

Table 1

(f) Set tolerance $= 10^{-6}$. If the step size $\alpha = 1/L = 1/\lambda_{\max}(X^T X)$ is increased slightly to $\alpha = 1.5/L = 1.5/\lambda_{\max}(X^T X)$. Do your codes for APG and APG-restart run faster, or do they diverge? Report the comparisons of numerical results with different step sizes ($\alpha = 1/L$ vs. $\alpha = 1.5/L$).