Homework 13

Requirements:

1. Digital format (can be typeset or photos, ought to write clearly if written by hand), upload to `https://course.pku.edu.cn/`.
2. Submit by next class
3. A problem is not counted if nobody can work it out
4. Each homework 10 points; 1 point deducted for each week's delay

Problems:

1. Use LADMPSAP to solve a graph construction problem:

$$\min_{\mathbf{Z},\mathbf{E}} ||\mathbf{Z}||_* + \lambda||\mathbf{E}||_{2,1}, \quad \text{s.t.} \quad \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{E}, \mathbf{Z}^T\mathbf{1} = \mathbf{1}, \mathbf{Z} \geq \mathbf{0}, \tag{1}$$

where $\mathbf{1}$ is an all-one vector. Randomly generate $\mathbf{D} \in \mathbb{R}^{200\times 300}$. Hand in your code and report.

2. Prove that the Lipschitz constant for the gradient of the logistic function

$$\frac{1}{s}\sum_{i=1}^{s} \log\left(1 + \exp\left(-y_i(\bar{\mathbf{w}}^T\bar{\mathbf{x}}_i)\right)\right)$$

respect to $\bar{\mathbf{w}}$ is upper bounded by $L_{\bar{w}} \leq \frac{1}{4s}\|\bar{\mathbf{X}}\|_2^2$, where $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_s)$.

Then randomly generate samples and solve the minimization problem of the logistic function by

1) gradient descent;

2) pLADMPSAP by introducing $\bar{\mathbf{w}}_i = \bar{\mathbf{w}}$, $i = 1, \cdots, s$.

Compare their convergence speed numerically. Hand in your code and report.

3. Use block coordinate descent to solve the dictionary learning problem:

$$\min_{\mathbf{D},\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_1, \quad s.t. \quad \|\mathbf{d}_i\|_2 = 1, i = 1, \cdots, K.$$

Randomly generate $\mathbf{Y} \in \mathbb{R}^{200\times 500}$, $\mathbf{D} \in \mathbb{R}^{200\times 400}$ and $\mathbf{X} \in \mathbb{R}^{400\times 500}$. Hand in your code and report showing the difference $\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2$.

4. Use block coordinate descent to solve the low-rank matrix completion problem:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{A}} \frac{1}{2}\|\mathbf{U}\mathbf{V}^T - \mathbf{A}\|_F^2, \quad s.t. \quad \mathcal{P}_\Omega(\mathbf{A}) = \mathcal{P}_\Omega(\mathbf{D}),$$

where $\mathcal{P}_\Omega(\cdot)$ is an operator that extracts entries of a matrix whose indices are in Ω and sets the remaining entries zeros.

Randomly generate $\mathbf{D} = \mathbf{U}_0\mathbf{V}_0^T$ and Ω, where $\mathbf{U}_0 \in \mathbb{R}^{200\times 5}$, $\mathbf{V}_0 \in \mathbb{R}^{300\times 5}$ and $|\Omega| = 0.1 \times 200 \times 300$. Compare with BCD that adds an extra orthogonalization step on $\mathbf{U}$. Hand in your code and report showing your settings and the difference $\|\mathbf{A}^* - \mathbf{D}\|_F$, where $\mathbf{A}^*$ is the optimal solution.

5. [Parallel Projections Algorithm] We are given m closed convex sets $\mathcal{X}_1, \cdots, \mathcal{X}_m$ in $\mathbb{R}^n$, and we want to find a point in their intersection. Consider the equivalent problem

$$\min_{\mathbf{x}, \{\mathbf{y}_i\}} \frac{1}{2} \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{y}_i\|^2,$$
$$s.t.\ \mathbf{x} \in \mathbb{R}^n, \mathbf{y}_i \in \mathcal{X}_i, i = 1, \cdots, m.$$

a) Derive a block coordinate descent algorithm involving projections on each of the sets $\mathcal{X}_i$ that can be carried out independently for each set. State a convergence result for this algorithm.

b) Implement your algorithm, with $m = 5$ and $\mathcal{X}_i$ being 2D disks having nonempty intersection.