Name: ______________ Student ID: ______________ Scores: ______________

AIE6003 Homework 4
Instructor: Xiaopeng Li **Deadline**: May 6

Notice: If you are trying to invoke a result outside of this course, then you should put a reference to it. The questions with 1 star are easy, with 2 stars are medium, and with 3 stars are hard. Easy and medium level of questions are likely to occur in the exams, while hard questions are not likely to occur in the exams.

1*. (15 points) Denote $[m] := \{1, \dots, m\}$ for any positive integer m. Let

$$B_*(\tau) := \{X \in \mathbb{R}^{n\times n} : \|X\|_* \le \tau\}, \qquad \tau > 0.$$

Throughout this problem, you may use without proof the result from our lecture slides.

(a) (4 points) Let $A = \sigma uv^\top$, where $\sigma > 0$ and $\|u\|_2 = \|v\|_2 = 1$. Consider the fully observed problem

$$\min_{X\in B_*(\tau)} \frac{1}{2}\|X - A\|_F^2.$$

Show that the unique optimizer is $X^\star = \min\{\sigma, \tau\}uv^\top$.

(b) (4 points) Starting from $X_0 = 0$, consider the following update. Set $S_0 := \tau uv^\top$. Then choose the scalar $\eta^\star \in [0, 1]$ that gives the smallest objective value along the segment from 0 to S_0:

$$\eta^\star \in \operatorname*{argmin}_{0\le\eta\le1} \frac{1}{2}\|\eta S_0 - A\|_F^2, \qquad X_1 := \eta^\star S_0.$$

Show that $\eta^\star = \min\left\{\frac{\sigma}{\tau}, 1\right\}$ and hence $X_1 = X^\star$. Conclude that, for this special rank-one fully observed problem, this rank-one update solves the constrained problem in one iteration.

(c) (3 points) Now return to matrix completion. Define $M_{\text{coh}} = e_1e_1^\top$ and $M_{\text{inc}} = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$. Suppose $\Omega \subset [n] \times [n]$ is sampled uniformly without replacement with $|\Omega| = m$. Show that if $(1, 1) \notin \Omega$, then $P_\Omega(M_{\text{coh}}) = 0$. Conclude that on this event, no algorithm can distinguish M_{coh} from the zero matrix using only the observations. Compute $\mathbb{P}((1, 1) \notin \Omega)$.

(d) (4 points) Let $n = 40$, $\tau = 1$, and

$$f(X) = \frac{1}{2}\|P_\Omega(X - Y)\|_F^2, \qquad Y = P_\Omega(M).$$

For each trial, sample the same index set Ω for both matrices, with $m \in \{2n, 4n, 6n, 8n\}$. Run Frank-Wolfe from $X_0 = 0$ for 200 iterations with $\eta_t = \frac{2}{t+2}$. If the gradient is zero, you may leave the iterate unchanged. Report the average relative recovery error

$$\frac{\|X_{200} - M\|_F}{\|M\|_F}$$

over 20 trials, and plot the average error versus m for both M_{inc} and M_{coh}. Briefly explain why the coherent case is much harder even though both matrices have rank 1 and nuclear norm 1.

2**. (15 points) Consider the LASSO problem

$$\min_{\beta\in\mathbb{R}^d} F(\beta) := \frac{1}{2}\|A\beta - b\|_2^2 + \lambda\|\beta\|_1,$$

where $A = [a_1, \dots, a_d] \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$, and $\lambda > 0$.

(a) (3 points) Show that $\beta = 0$ is optimal if and only if $\|A^\top b\|_\infty \le \lambda$.

(b) (4 points) Let $S \subset [d]$, and let $\hat\theta \in \mathbb{R}^{|S|}$ be an optimizer of the restricted Lasso problem

$$\min_{\theta \in \mathbb{R}^{|S|}} \frac{1}{2}\|A_S\theta - b\|_2^2 + \lambda\|\theta\|_1.$$

Pad $\hat\theta$ with zeros outside S to obtain $\tilde\beta \in \mathbb{R}^d$. Show that if $|a_i^\top(A\tilde\beta - b)| \le \lambda$ for all $i \notin S$, then $\tilde\beta$ is a global minimizer of the full Lasso problem.

(c) (4 points) Now consider the group Lasso problem

$$\min_{\beta \in \mathbb{R}^d} \frac{1}{2}\|A\beta - b\|_2^2 + \lambda \sum_{g=1}^m \|\beta_{B_g}\|_2,$$

where $B_1, \ldots, B_m$ partition $[d]$. Let $\mathcal{G} \subset [m]$ be a set of active groups (nonzero variables), and let $\tilde\beta$ be an optimizer of the restricted problem on the coordinates in $\bigcup_{g\in\mathcal{G}} B_g$, padded with zeros on the inactive groups. Prove that $\tilde\beta$ is globally optimal for the full group Lasso whenever $\|A_{B_g}^\top(A\tilde\beta - b)\|_2 \le \lambda$ for $g \notin \mathcal{G}$.

(d) (4 points) Suppose the certificate in part (b) fails. Namely, for some $i \notin S$, $|a_i^\top(A\tilde\beta - b)| > \lambda$. Let $r := A\tilde\beta - b$ and consider changing only coordinate i: $\beta(\delta) = \tilde\beta + \delta e_i$. Derive the exact minimizer of $F(\beta(\delta))$ over $\delta \in \mathbb{R}$, and show that the objective decreases strictly. Quantify the decrease in terms of $a_i^\top r$, λ, and $\|a_i\|_2$.

3***. (20 points) For $\mu \ge 0$, consider

$$f_\mu(x,y) = \frac{\mu}{2}\|x\|_2^2 + x^\top A y - \frac{\mu}{2}\|y\|_2^2, \qquad A \in \mathbb{R}^{m\times n}.$$

Let $L := \|A\|_{\mathrm{op}}$. Define

$$z = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad K = \begin{pmatrix} 0 & A \\ -A^\top & 0 \end{pmatrix}, \qquad F(z) = \begin{pmatrix} \nabla_x f_\mu(x,y) \\ -\nabla_y f_\mu(x,y) \end{pmatrix},$$

(a) (5 points) Show that $F(z) = (\mu I + K)z$. For $\mu > 0$, prove that F is μ-strongly monotone and ℓ-Lipschitz with $\ell = \sqrt{\mu^2 + L^2}$. Conclude that the saddle point is unique and equals $z^\star = 0$.

(b) (6 points) Simultaneous GDA with step size $\eta > 0$ is

$$z_{t+1} = z_t - \eta F(z_t) = \big((1-\eta\mu)I - \eta K\big)z_t.$$

Use $K^\top = -K$ to prove the exact identity

$$\|z_{t+1}\|_2^2 = (1-\eta\mu)^2\|z_t\|_2^2 + \eta^2\|Kz_t\|_2^2.$$

Deduce that GDA is a global contraction for every initialization if and only if

$$0 < \eta < \frac{2\mu}{\mu^2 + L^2}.$$

Find the best fixed step size and the corresponding worst-case contraction factor.

(c) (5 points) Now specialize to the scalar bilinear case $\mu = 0$, $m = n = 1$, $A = [\sigma]$, $\sigma > 0$. Let $w_t := x_t + iy_t$. Show that simultaneous GDA satisfies

$$w_{t+1} = (1 + i\eta\sigma)w_t,$$

so the last iterate expands for every $\eta > 0$ unless $w_t = 0$. For extragradient (EG),

$$z_{t+1/2} = z_t - \eta F(z_t), \qquad z_{t+1} = z_t - \eta F(z_{t+1/2}),$$

show that

$$w_{t+1} = \big(1 + i\eta\sigma - \eta^2\sigma^2\big) w_t.$$

Determine exactly for which $\eta > 0$ EG converges linearly, and find the step size in this interval that gives the smallest contraction factor.

(d) (4 points) For the same scalar bilinear problem, consider the implicit proximal-point step

$$z_{t+1} = z_t - \eta F(z_{t+1}).$$

Writing again $w_t = x_t + iy_t$, solve this two-dimensional linear system explicitly and show that

$$w_{t+1} = \frac{1 + i\eta\sigma}{1 + \eta^2\sigma^2}\, w_t.$$

What is the contraction factor?