Name: _____ Student ID: _____ Scores: _____

AIE6003 Final
Instructor: Xiaopeng Li **Time Limit**: 2 hours

Notice: You can invoke results in the lecture and homework without proving it.

1. (170 points) A linear model has n tunable coefficients $x \in \mathbb{R}^n$, fitted by minimizing a squared-error loss on a training set. After expanding the loss, the training objective is a convex quadratic in the coefficients, $f(x) := x^\top Q x - 2p^\top x$, where $Q = Q^\top \succ 0$ is the feature Gram matrix and $p \in \mathbb{R}^n$ collects the feature–label correlations. On a small or noisy training set, the fitted coefficients can blow up and produce an unstable model; to keep things well-conditioned, we impose a hard cap on their total magnitude by requiring the coefficient vector to lie inside the unit Euclidean ball. Let $\mathbb{B} := \{x \in \mathbb{R}^n : \|x\| \le 1\}$. Consider

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad x^\top x \le 1. \tag{P}$$

   (a) (45 points) Derive the normal cone $N_{\mathbb{B}}(x)$ for every feasible $x \in \mathbb{B}$.

   Solution: For the unit ball $\mathbb{B} = \{x : \|x\| \le 1\}$, the normal cone is

$$N_{\mathbb{B}}(x) = \begin{cases} \{0\}, & \|x\| < 1, \\ \{\mu x : \mu \ge 0\}, & \|x\| = 1. \end{cases}$$

   Indeed, if $\|x\| < 1$, then every sufficiently small direction is feasible, so the only vector normal to all feasible directions is 0. If $\|x\| = 1$ and $\mu \ge 0$, then for any $z \in \mathbb{B}$,

$$\langle \mu x, z - x \rangle = \mu(\langle x, z \rangle - 1) \le \mu(\|z\| - 1) \le 0,$$

   so $\mu x \in N_{\mathbb{B}}(x)$. Conversely, the supporting hyperplane of the Euclidean ball at a boundary point is orthogonal to x, hence every normal vector is a nonnegative multiple of x.

   (b) (70 points) Derive the KKT system and the one-dimensional Lagrange dual problem for (P).

   Solution: The normal-cone optimality condition is

$$-\nabla f(x^*) = 2p - 2Qx^* \in N_{\mathbb{B}}(x^*).$$

   Using part (a), this is equivalent to the existence of $\lambda^* \ge 0$ such that

$$\begin{cases} 2Qx^* - 2p + 2\lambda^* x^* = 0, \\ \|x^*\|^2 \le 1, \\ \lambda^* \ge 0, \\ \lambda^*(\|x^*\|^2 - 1) = 0. \end{cases}$$

   Equivalently,

$$(Q + \lambda^* I)x^* = p, \qquad \lambda^*(\|x^*\|^2 - 1) = 0.$$

   The problem is convex, and Slater's condition holds because $x = 0$ is strictly feasible. Thus the KKT conditions are necessary and sufficient.

For the dual, use the Lagrangian

$$L(x,\lambda) = x^\top Q x - 2p^\top x + \lambda(x^\top x - 1), \qquad \lambda \geq 0.$$

For fixed $\lambda \geq 0$, the minimizer over x satisfies

$$2(Q+\lambda I)x - 2p = 0, \qquad x(\lambda) = (Q+\lambda I)^{-1}p.$$

Therefore

$$g(\lambda) = \inf_x L(x,\lambda) = -p^\top (Q+\lambda I)^{-1} p - \lambda.$$

The one-dimensional dual problem is

$$\max_{\lambda \geq 0} \; -p^\top (Q+\lambda I)^{-1} p - \lambda.$$

By Slater's condition, strong duality holds.

(c) (55 points) Assume $p \neq 0$ and $\lambda_0 \geq 0$. Let $g(\lambda)$ be the dual function from part (b), let $\lambda^* \in \operatorname{argmax}_{\lambda \geq 0} g(\lambda)$, and define projected dual-gradient ascent update

$$\lambda_{t+1} = \left[\lambda_t + L_d^{-1} g'(\lambda_t)\right]_+, \quad \text{where} \quad L_d := \frac{2\|p\|^2}{\lambda_{\min}(Q)^3}.$$

Prove

$$g(\lambda^*) - g(\lambda_T) \leq \frac{L_d |\lambda_0 - \lambda^*|^2}{2T}.$$

Hint: first compute $g'(\lambda)$ and $g''(\lambda)$, then apply projected gradient descent to $h = -g$ on $[0,\infty)$.

Solution: From part (b),

$$g(\lambda) = -p^\top (Q+\lambda I)^{-1} p - \lambda.$$

Differentiating gives

$$g'(\lambda) = p^\top (Q+\lambda I)^{-2} p - 1,$$

and

$$g''(\lambda) = -2p^\top (Q+\lambda I)^{-3} p.$$

Since $Q \succeq \lambda_{\min}(Q) I$ and $\lambda \geq 0$,

$$|g''(\lambda)| \leq \frac{2\|p\|^2}{\lambda_{\min}(Q)^3} = L_d.$$

Hence $h := -g$ is convex and L_d-smooth on $[0,\infty)$.
The update $\lambda_{t+1} = \left[\lambda_t + L_d^{-1} g'(\lambda_t)\right]_+$ defines projected gradient descent on h over $[0,\infty)$. Since $\lambda_0 \geq 0$ and each step projects onto $[0,\infty)$, all iterates satisfy $\lambda_t \geq 0$. For projected gradient descent with stepsize $1/L_d$ on a convex L_d-smooth function,

$$h(\lambda_{t+1}) - h(\lambda^*) \leq \frac{L_d}{2}\left(|\lambda_t - \lambda^*|^2 - |\lambda_{t+1} - \lambda^*|^2\right).$$

Summing from $t = 0$ to $T-1$ gives

$$\sum_{t=0}^{T-1} \left(h(\lambda_{t+1}) - h(\lambda^*)\right) \leq \frac{L_d}{2} |\lambda_0 - \lambda^*|^2.$$

Since projected gradient descent with stepsize $1/L_d$ is descent for h,

$$T\left(h(\lambda_T) - h(\lambda^*)\right) \leq \frac{L_d}{2}|\lambda_0 - \lambda^*|^2.$$

Using $h = -g$ yields

$$g(\lambda^*) - g(\lambda_T) \leq \frac{L_d|\lambda_0 - \lambda^*|^2}{2T}.$$

2. (170 points) After deployment, a large model is fine-tuned by adding a correction matrix $X \in \mathbb{R}^{m\times n}$ to its frozen weights. Storing or transmitting the full correction is expensive, so we want X to be effectively low rank, i.e., expressible as the product of two thin matrices. To keep the correction's overall magnitude under control, we cap its nuclear norm by a budget $\tau > 0$: this is a convex surrogate for a hard rank cap. During training, only a subset of entries $\Omega \subseteq [m] \times [n]$ is monitored, and $Y \in \mathbb{R}^{m\times n}$ records the desired correction values on those entries. The data-fit term and the nuclear-norm ball are

$$F(X) := \frac{1}{2}\|P_\Omega(X - Y)\|_F^2, \qquad \mathcal{N}_\tau := \{X \in \mathbb{R}^{m\times n} : \|X\|_* \leq \tau\}.$$

On top of this, a separate correction layer cleans up the residual: it splits the residual into a smooth low-rank piece and a sparse piece.

(a) (60 points) Given a residual matrix $R \in \mathbb{R}^{m\times n}$, the low-rank + sparse decomposition is captured by the convex objective

$$J(L, S) := \frac{1}{2}\|L + S - R\|_F^2 + \lambda_L\|L\|_* + \alpha\|S\|_1, \qquad \lambda_L, \alpha > 0.$$

where L is the smooth drift component and S is the spike component. Introduce suitable auxiliary matrices and derive the scaled ADMM iteration with penalty parameter $\rho > 0$, using singular-value thresholding and entrywise soft-thresholding.

Solution: Introduce auxiliary matrices Z_L and Z_S to separate the nonsmooth terms:

$$\min_{L,S,Z_L,Z_S} \frac{1}{2}\|L + S - R\|_F^2 + \lambda_L\|Z_L\|_* + \alpha\|Z_S\|_1$$
$$\text{s.t. } L - Z_L = 0, \qquad S - Z_S = 0.$$

With scaled dual variables U_L, U_S, the scaled augmented Lagrangian, up to constants independent of the primal variables, is

$$\frac{1}{2}\|L + S - R\|_F^2 + \lambda_L\|Z_L\|_* + \alpha\|Z_S\|_1 + \frac{\rho}{2}\|L - Z_L + U_L\|_F^2 + \frac{\rho}{2}\|S - Z_S + U_S\|_F^2.$$

Use two ADMM blocks: (L, S) and (Z_L, Z_S). Define $A^k := Z_L^k - U_L^k$ and $B^k := Z_S^k - U_S^k$. The (L, S)-update solves

$$\min_{L,S} \frac{1}{2}\|L + S - R\|_F^2 + \frac{\rho}{2}\|L - A^k\|_F^2 + \frac{\rho}{2}\|S - B^k\|_F^2.$$

The first-order conditions are

$$L + S - R + \rho(L - A^k) = 0, \qquad L + S - R + \rho(S - B^k) = 0.$$

Let

$$T^k := \frac{2R + \rho(A^k + B^k)}{\rho + 2}.$$

Adding and subtracting the two first-order conditions gives

$$L^{k+1} = \frac{1}{2}\left(T^k + A^k - B^k\right), \qquad S^{k+1} = \frac{1}{2}\left(T^k - A^k + B^k\right).$$

The Z_L-update is the proximal operator of the nuclear norm:

$$Z_L^{k+1} = \mathrm{SVT}_{\lambda_L/\rho}\left(L^{k+1} + U_L^k\right).$$

The Z_S-update is the proximal operator of the entrywise ℓ_1 norm:

$$Z_S^{k+1} = \mathcal{S}_{\alpha/\rho}\left(S^{k+1} + U_S^k\right).$$

Finally, the scaled dual updates are

$$U_L^{k+1} = U_L^k + L^{k+1} - Z_L^{k+1}, \qquad U_S^{k+1} = U_S^k + S^{k+1} - Z_S^{k+1}.$$

Here, if $W = P\,\mathrm{diag}(\sigma_i)Q^\top$, then

$$\mathrm{SVT}_c(W) = P\,\mathrm{diag}((\sigma_i - c)_+)Q^\top,$$

and $\mathcal{S}_c$ is entrywise soft-thresholding:

$$[\mathcal{S}_c(W)]_{ij} = \mathrm{sign}(W_{ij})(|W_{ij}| - c)_+.$$

(b) (55 points) To avoid storing the full $m \times n$ matrix X, parametrize $X = UV^\top$ with $U \in \mathbb{R}^{m\times k}$ and $V \in \mathbb{R}^{n\times k}$ ($k \ll \min(m,n)$). The factorization is non-unique, so we add a balancing regularizer that pushes U and V to have the same scale. Derive the gradients of

$$\Psi(U,V) := \frac{1}{2}\|P_\Omega(UV^\top - Y)\|_F^2 + \frac{\gamma}{4}\|U^\top U - V^\top V\|_F^2, \qquad \gamma > 0.$$

Solution: Define

$$R(U,V) := P_\Omega(UV^\top - Y), \qquad W(U,V) := U^\top U - V^\top V.$$

For perturbations $(\Delta U, \Delta V)$, the data-fit differential is

$$\langle R(U,V), \Delta U V^\top + U\Delta V^\top\rangle = \langle R(U,V)V, \Delta U\rangle + \langle R(U,V)^\top U, \Delta V\rangle.$$

Since W is symmetric,

$$d\left(\frac{\gamma}{4}\|W\|_F^2\right) = \gamma\langle UW, \Delta U\rangle - \gamma\langle VW, \Delta V\rangle.$$

Therefore

$$\nabla_U \Psi(U,V) = R(U,V)V + \gamma U\left(U^\top U - V^\top V\right),$$

and

$$\nabla_V \Psi(U,V) = R(U,V)^\top U - \gamma V\left(U^\top U - V^\top V\right).$$

(c) (55 points) An alternative to factorization is to keep X in matrix form but update it by Frank-Wolfe over $\mathcal{N}_\tau$, which never forms a dense matrix: each linear minimization returns a single rank-1 outer

product, so after t iterations the iterate is a convex combination of at most t such outer products. Assume F is L-smooth on $\mathcal{N}_\tau$. Starting with $X_0 = 0$, consider Frank-Wolfe with

$$S_t \in \operatorname*{argmin}_{S \in \mathcal{N}_\tau} \langle \nabla F(X_t), S \rangle, \qquad X_{t+1} = (1-\eta_t)X_t + \eta_t S_t, \qquad \eta_t = \frac{2}{t+2}.$$

Denote $X^* \in \operatorname{argmin}_{X \in \mathcal{N}_\tau} F(X)$. Prove

$$F(X_t) - F(X^*) \le \frac{8L\tau^2}{t+2}, \qquad t \ge 1.$$

Hint: Use smoothness, the Frank-Wolfe oracle, convexity of F, and $\operatorname{diam}(\mathcal{N}_\tau) \le 2\tau$.

Solution: Note $X_0 = 0 \in \mathcal{N}_\tau$. Let $h_t := F(X_t) - F(X^*)$. Since F is L-smooth and $X_{t+1} = (1-\eta_t)X_t + \eta_t S_t$,

$$F(X_{t+1}) \le F(X_t) + \eta_t \langle \nabla F(X_t), S_t - X_t \rangle + \frac{L}{2}\eta_t^2 \|S_t - X_t\|_F^2.$$

By the Frank-Wolfe oracle and convexity,

$$\langle \nabla F(X_t), S_t - X_t \rangle \le \langle \nabla F(X_t), X^* - X_t \rangle \le F(X^*) - F(X_t) = -h_t.$$

Also, for any $X, Z \in \mathcal{N}_\tau$,

$$\|X - Z\|_F \le \|X\|_F + \|Z\|_F \le \|X\|_* + \|Z\|_* \le 2\tau.$$

Thus

$$h_{t+1} \le (1-\eta_t)h_t + 2L\tau^2\eta_t^2.$$

With $\eta_t = 2/(t+2)$,

$$h_{t+1} \le \frac{t}{t+2}h_t + \frac{8L\tau^2}{(t+2)^2}.$$

The smoothness bound for $t = 0$ gives $h_1 \le 2L\tau^2 \le 8L\tau^2/3$. Suppose $h_t \le 8L\tau^2/(t+2)$. Then

$$h_{t+1} \le \frac{t}{t+2} \cdot \frac{8L\tau^2}{t+2} + \frac{8L\tau^2}{(t+2)^2} = \frac{8L\tau^2(t+1)}{(t+2)^2} \le \frac{8L\tau^2}{t+3}.$$

By induction,

$$F(X_t) - F(X^*) \le \frac{8L\tau^2}{t+2}, \qquad t \ge 1.$$

3. (160 points) A power-grid operator must dispatch resources in real time while defending against worst-case demand fluctuations. The operator chooses a load-shifting vector $u \in [-1,1]^m$ ($u_i \in [-1,1]$ is how much of the i-th flexible resource to ramp up or down, in normalized units, capped at full intensity in either direction). An adversary chooses a demand perturbation $v \in [-1,1]^n$ ($v_j \in [-1,1]$ is the percentage shock applied to the j-th load bucket, again capped). The operator's loss combines a coupling term $u^\top C v$ (how the dispatch responds to the demand shock through the network), a fixed dispatch cost $a^\top u$ on the operator side, and a fixed reward $b^\top v$ for the adversary:

$$\psi(u,v) := u^\top C v + a^\top u - b^\top v, \qquad C \in \mathbb{R}^{m \times n},\ a \in \mathbb{R}^m,\ b \in \mathbb{R}^n.$$

The operator minimizes ψ over its action set; the adversary maximizes ψ over its action set.

(a) (40 points) The duality gap measures how far a candidate strategy pair (u, v) is from a saddle-point equilibrium: it adds the operator's regret to the adversary's regret. Derive a closed-form expression for the duality gap $\operatorname{gap}(u, v)$ of ψ on $[-1,1]^m \times [-1,1]^n$.

Solution: For fixed $u \in [-1,1]^m$,

$$\max_{v' \in [-1,1]^n} \psi(u, v') = a^\top u + \max_{v' \in [-1,1]^n} (C^\top u - b)^\top v' = a^\top u + \|C^\top u - b\|_1.$$

For fixed $v \in [-1,1]^n$,

$$\min_{u' \in [-1,1]^m} \psi(u', v) = \min_{u' \in [-1,1]^m} (Cv + a)^\top u' - b^\top v = -\|Cv + a\|_1 - b^\top v.$$

Therefore

$$\mathrm{gap}(u, v) = a^\top u + b^\top v + \|C^\top u - b\|_1 + \|Cv + a\|_1.$$

(b) (70 points) On a tight market day, both sides become risk averse and pay a quadratic penalty for large actions, which is captured by adding $\frac{\mu}{2}\|u\|^2$ to the operator's loss and $\frac{\mu}{2}\|v\|^2$ to the adversary's loss. The action sets are also relaxed: instead of the box, the operator moves freely in $\mathbb{R}^m$ and the adversary moves freely in $\mathbb{R}^n$. The regularized unconstrained payoff becomes

$$\psi_\mu(u, v) := \frac{\mu}{2}\|u\|^2 + u^\top C v - \frac{\mu}{2}\|v\|^2 + a^\top u - b^\top v, \qquad \mu > 0.$$

The regularization makes the game strongly convex in u and strongly concave in v, so simultaneous GDA contracts toward the saddle point as long as the stepsize is in the right range. Prove that simultaneous GDA is a global contraction exactly when

$$0 < \eta < \frac{2\mu}{\mu^2 + \|C\|_{\mathrm{op}}^2}.$$

Solution: For ψ_μ,

$$\nabla_u \psi_\mu(u, v) = \mu u + Cv + a, \qquad -\nabla_v \psi_\mu(u, v) = \mu v - C^\top u + b.$$

Thus the saddle-gradient operator is

$$F(z) = \begin{pmatrix} \mu I_m & C \\ -C^\top & \mu I_n \end{pmatrix} z + \begin{pmatrix} a \\ b \end{pmatrix}, \qquad z = (u, v).$$

Let z^* be the unique saddle point, so $F(z^*) = 0$, and set $e_t = z_t - z^*$. GDA gives

$$e_{t+1} = e_t - \eta(F(z_t) - F(z^*)) = ((1 - \eta\mu)I - \eta K)\, e_t, \qquad K := \begin{pmatrix} 0 & C \\ -C^\top & 0 \end{pmatrix}.$$

Since $K^\top = -K$, we have $\langle e_t, K e_t \rangle = 0$. Hence

$$\begin{aligned} \|e_{t+1}\|^2 &= \|(1 - \eta\mu)e_t - \eta K e_t\|^2 \\ &= (1 - \eta\mu)^2 \|e_t\|^2 + \eta^2 \|K e_t\|^2. \end{aligned}$$

Moreover $\|K\|_{\mathrm{op}} = \|C\|_{\mathrm{op}}$, so the worst-case squared contraction factor is

$$q(\eta)^2 = (1 - \eta\mu)^2 + \eta^2 \|C\|_{\mathrm{op}}^2 = 1 - 2\eta\mu + \eta^2(\mu^2 + \|C\|_{\mathrm{op}}^2).$$

Therefore GDA is a global contraction exactly when $q(\eta)^2 < 1$, i.e.

$$-2\eta\mu + \eta^2(\mu^2 + \|C\|_{\mathrm{op}}^2) < 0.$$

Since $\eta > 0$, this is equivalent to

$$0 < \eta < \frac{2\mu}{\mu^2 + \|C\|_{\mathrm{op}}^2}.$$

(c) (50 points) Drop the regularizers and the action box, leaving only the pure coupling term: the operator and adversary play the centered bilinear game $\min_u \max_v u^\top C v$ over $\mathbb{R}^s \times \mathbb{R}^s$, with $m = n = s$ and C nonsingular. Bilinear games of this type are notoriously bad for explicit GDA, so consider instead the extragradient (EG) update,

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

where $z = (u, v)$ and $F = (\nabla_u(u^\top C v), -\nabla_v(u^\top C v))$, prove that if $0 < \eta < 1/\|C\|_{\mathrm{op}}$, then EG converges linearly to $(0, 0)$, and derive the exact SVD-mode contraction factor.
Hint: apply SVD to C and decouple the original problem into independent 2-d bilinear games.

Solution: For the centered bilinear game,

$$F(z) = Kz, \qquad K := \begin{pmatrix} 0 & C \\ -C^\top & 0 \end{pmatrix}.$$

Let $C = P\Sigma Q^\top$ be an SVD with singular values $\sigma_1 \geq \cdots \geq \sigma_s > 0$. In transformed coordinates $\bar{u} = P^\top u$ and $\bar{v} = Q^\top v$, each singular mode is independent:

$$F_j(\bar{u}_j, \bar{v}_j) = (\sigma_j \bar{v}_j, -\sigma_j \bar{u}_j).$$

The half-step is

$$\bar{u}_{j,t+1/2} = \bar{u}_{j,t} - \eta\sigma_j \bar{v}_{j,t}, \qquad \bar{v}_{j,t+1/2} = \bar{v}_{j,t} + \eta\sigma_j \bar{u}_{j,t}.$$

The EG step gives

$$\begin{aligned} \bar{u}_{j,t+1} &= \bar{u}_{j,t} - \eta\sigma_j \bar{v}_{j,t+1/2} = (1 - \eta^2\sigma_j^2)\bar{u}_{j,t} - \eta\sigma_j \bar{v}_{j,t}, \\ \bar{v}_{j,t+1} &= \bar{v}_{j,t} + \eta\sigma_j \bar{u}_{j,t+1/2} = \eta\sigma_j \bar{u}_{j,t} + (1 - \eta^2\sigma_j^2)\bar{v}_{j,t}. \end{aligned}$$

Equivalently, with $w_{j,t} := \bar{u}_{j,t} + i\bar{v}_{j,t}$ and $a_j := \eta\sigma_j$,

$$w_{j,t+1} = (1 + ia_j - a_j^2) w_{j,t}.$$

Therefore the exact squared contraction factor in mode j is

$$|1 + ia_j - a_j^2|^2 = (1 - a_j^2)^2 + a_j^2 = 1 - a_j^2 + a_j^4.$$

Thus

$$\|z_{t+1}\|^2 \leq \rho_\eta^2 \|z_t\|^2, \qquad \rho_\eta^2 := \max_{1 \leq j \leq s} \left(1 - \eta^2\sigma_j^2 + \eta^4\sigma_j^4\right).$$

If $0 < \eta < 1/\|C\|_{\mathrm{op}}$, then $0 < \eta\sigma_j < 1$ for every j, and hence

$$1 - \eta^2\sigma_j^2 + \eta^4\sigma_j^4 < 1.$$

Since C is nonsingular, all $\sigma_j > 0$, so $\rho_\eta < 1$. Therefore EG converges linearly to $(0, 0)$:

$$\|z_t\| \leq \rho_\eta^t \|z_0\|, \qquad \rho_\eta = \max_{1 \leq j \leq s} \sqrt{1 - \eta^2\sigma_j^2 + \eta^4\sigma_j^4} < 1.$$