Name: Gosuddin Siddiqi

AMATH 521 Homework Set 3

Due: Tuesday November 14th, on Canvas.

Problem 1*: Matrix Norms*

Consider any full rank matrix $A \in \mathbb{R}^{m \times n}$ (assume $m > n$), with SVD,

$$A = U\Sigma V^{\mathsf{T}},$$

where $U \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}n \times n$ and $\Sigma \in \mathbb{R}^{n\times n}$ is a diagonal matrix with positive entries. $\text{diag}(\Sigma) = [\sigma_1, \ldots, \sigma_n]^{\mathsf{T}} \triangleq \sigma(A)$, and $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$.

Here are some useful properties,

- $\|A\|_{\mathsf{F}}^2 = \text{trace}(A^{\mathsf{T}}A)$.
- $\text{trace}(AB) = \text{trace}(BA)$.
- $\|A\|_2 = \sup_{\|x\|_2 \leq 1} \|Ax\|_2 = \sup_{\|x\|_2 = 1} \|Ax\|_2$.

Show the following result:

(1) $\|A\|_{\mathsf{F}}^2 = \|\sigma(A)\|_2^2 = \sum_{i=1}^n \sigma_i^2$.

We know that,
$\|A\|_{\mathsf{F}}^2 = \text{trace}(A^{\mathsf{T}}A) \ldots (1)$
$\text{trace}(AB) = \text{trace}(BA) \ldots (2)$

Here, $A = U\Sigma V^{\mathsf{T}}$
We can rewrite, $\|A\|_{\mathsf{F}}^2$ using (1) as $\text{trace}(V\Sigma^2V^T)$

$\text{trace}(V\Sigma^2V^T) = \text{trace}(VV^T\Sigma^2) = \text{trace}(\Sigma^2)$

We know that $\text{trace}(\Sigma^2) = \Sigma_{i=1}^n \sigma_i^2$

Gathering all we get, $\|A\|_{\mathsf{F}}^2 = \|\sigma(A)\|_2^2 = \sum_{i=1}^n \sigma_i^2$

(2) $\|A\|_2 = \|\sigma(A)\|_\infty = \sigma_1$.

We know that $\|Uy\|_2 = \|y\|_2$, where U has orthonormal columns.

Let SVD of $A = U\Sigma V^T$, where U has orthonormal columns.

For a vector $x \in \mathbb{R}^n$, $\|Ax\|_2 = \|U\Sigma V^T x\|_2 = \|\Sigma V^T x\|_2 \ldots (1)$

Now, when $\|V^T x\|_2 = 1$, $\|x\|_2 = 1$.

Thus our above equation (1) becomes,
$\|A\|_2 = sup_{\|x\|_2=1}\|Ax\|_2 = sup_{\|V^t x\|_2=1}\|\Sigma V^T x\|_2 = sup_{\|y\|_2=1}\|\Sigma y\|_2$

Also, since Σ has all diagonal positive entries only, $\|\Sigma\|_2 = \sigma_1$

This proves our equation, $\|A\|_2 = \|\sigma(A)\|_\infty = \sigma_1$

(3) $\|A\|_* = \text{trace}\left(\sqrt{A^\mathsf{T} A}\right) = \|\sigma(A)\|_1 = \sum_{i=1}^n \sigma_i$, where $\sqrt{A^\mathsf{T} A}$ denotes any symmetric positive semidefinite matrix B such that $B^2 = A^\mathsf{T} A$.

We know that $A = U\Sigma V^T$, $A^T A = V\Sigma^2 V^T$. $B^2 = A^T A$, if we define $B = V\Sigma V^T$, by eigenvalue-eigenvector decomposition.

Thus, $\|A\|_* = \text{trace}(\sqrt{(A^T A)}) = \text{trace}(\sqrt{(B^2)}) = \text{trace}(\sqrt{(V\Sigma^2 V^T)}) = \text{trace}(\sqrt{(VV^T\Sigma^2)}) = \text{trace}(\sqrt{(\Sigma^2)}) = \text{trace}(\Sigma)$

$\text{trace}(\Sigma) = \sum_{i=1}^n \sigma_i^2 = \|\sigma(A)\|_1$

Hence, proved.

Problem 2: *Low Rank Matrix Factorization (Convex)*

Consider optimization problem,

$$\min_X \frac{1}{2}\|R - X\|_\mathsf{F}^2 + \lambda\|X\|_*$$

Denote $f(X) = \frac{1}{2}\|R - X\|_\mathsf{F}^2$ and $g(X) = \lambda\|X\|_*$.

(1) Calculate $\nabla f(X)$. Is f a β-smooth function? If it is what is the β?

We can write, $\frac{1}{2}\|R - X\|_\mathsf{F}^2 = \text{trace}((R - X^T)(R - X))$

When we differentiate w.r.t X, we obtain
$\nabla_X \text{trace}(AXB) = A^T B^T and \nabla_X \text{trace}(AX^T B) = BA$

Thus, $\nabla f(x) = \frac{1}{2}(\text{trace}(RR^T) - \text{trace}(RX^T) - \text{trace}(XR^T) + \text{trace}(XX^T))$

On simplifying, $= X - R$

$\nabla^2 f(x) = I$ and when $\nabla f(x) = F(X)$
$F(X) - F(Y) = (X - R) - (Y - R) = (X - Y)I.$

This implies that f is β-smooth with $\beta = 1$

(2) What is $\text{prox}_{\alpha g}(Z)$?

$\text{prox}_{\alpha g}(Z) = argmin_U(\lambda \|U\|_* + \frac{1}{2\alpha}\|U - Z\|_F^2) = argmin_U(\lambda \|U\|_* + \frac{1}{2\alpha\lambda}\|U - Z\|_F^2)$

Compute the SVD of Z. $Z = U\sigma(Z)V^T = U\hat{\Sigma}V^T$. Let $\hat{\Sigma}$ be a diagonal matrix having positive entires equal to $\hat{\sigma_i}$.

with this we can write the $\text{prox}_{\alpha g}(Z)$ as, $\hat{\sigma_i} =$

$$\begin{cases} \sigma_i - \lambda\alpha, & when \sigma_i > \alpha\lambda \\ \sigma_i + \lambda\alpha, & when \sigma_i < -\alpha\lambda \\ 0, & otherwise \end{cases}.$$

(3) Implement proximal gradient descent in **Jupyter Notebook** to solve the problem. For given λ report the rank k of optimal solution X_*.

The rank k of the optimal solution is 29.

(4) Based on the rank we find above, what is the solution of the following problem,

$$\min_X \frac{1}{2}\|R - X\|_{\mathsf{F}}^2, \quad \text{s.t. rank}(X) = k,$$

which we denote as X_k.

Let SVD of $A = U\Sigma V^T$. For ith columns of U and V, we can represent individual elements as a sum. Thus $A = \Sigma_{i=1}^n \sigma_i u_i v_i^T$. This can be generalized to rank k as $A_k = \Sigma_{i=1}^k \sigma_i u_i v_i^T$

Let us consider a matrix B, with rank k,then

$\|A - A_k\|_F \leq \|A - B\|_F$.

The above equation will be minimized when the B is of rank at the most of k because our matrix A has k singular values and rest as zeros.

So the corresponding vector V of matrix B, will also has a maximum rank of k. Since B minimizes $\|A - B\|_F$, each row of will be the projection of B onto A. This implies that $\|A - B\|_F$ is euclidean distance from A to V. Thus A_k minimizes euclidean distance for rank k dimensions,our consideration holds true.

Thus, $X_k = \Sigma_{i=1}^{k} \sigma_i u_i v_i^T$, R = $U \Sigma V^T$

Problem 3: *Low Rank Matrix Factorization (Nonconvex)*

Consider the low rank variation of the low rank matrix factorization,

$$\min_{B,F} \frac{1}{2}\|R - BF\|_{\mathsf{F}}^2, \quad \text{s.t. } B \in \mathbb{R}^{m \times k},\ F \in \mathbb{R}^{k \times n}.$$

where k is the rank you obtain from **Problem 2** (3).

Denote $f(B, F) = \frac{1}{2}\|R - BF\|_{\mathsf{F}}^2$.

(1) Calculate $\nabla_B f(B, F)$ and $\nabla_F f(B, F)$.

As seen in Problem 2 (1), we can show that

$\|R - BF\|_{\mathsf{F}}^2 = \text{trace}(RR^T) - \text{trace}(RB^T F^T) - \text{trace}(BFR^T) + \text{trace}(BFF^T B^T)$

When we differentiate w.r.t X, we obtain
$\nabla_X \text{trace}(AXB) = A^T B^T and \nabla_X \text{trace}(AX^T B) = BA$

Using these properties, we get, when we differentiate w.r.t B,

$\nabla_B f(B, F) = (BF - R)F^T$ and $\|\nabla_B^2 f(B, F)\|_2 = \|F^T F\|_2$

Similarly when we differentiate w.r.t F,

$\nabla_F f(B, F) = B^T(BF - R)$ and $\|\nabla_F^2 f(B, F)\|_2 = \|B^T B\|_2$

(2) Implement PALM algorithm in **Jupyter Notebook** solve the problem.

(3) Denote B_* and F_* the optimal solution you obtain from the algorithm, compare X_k with $B_* F_*$. Report $\|X_k - B_* F_*\|_{\mathsf{F}} / \|X_k\|_F$.

The change is just 0.0030 after 30000 iterations, which shows stability.

Problem 4: *Robust Risk Decomposition*

Consider R as our return matrix. Sometimes due to accidental activity R will contain outliers. In order to eliminate this random effect when we do the risk decomposition, we need to use robust formulation,

$$\min_{B,F} \rho_\kappa(\bar{R} - BF), \quad \text{s.t. } B^{\mathsf{T}} B = I$$

where ρ_κ is the Huber penalty and $\bar{R}$ is the centered return matrix.
Denote $f(B,F) = \rho_\kappa(\bar{R} - BF)$ and

$$g(B) = \delta(B \mid B^{\mathsf{T}} B = I) = \begin{cases} 0, & B^{\mathsf{T}} B = I \\ \infty, & B^{\mathsf{T}} B \neq I \end{cases}$$

(1) What is $\nabla_B f(B,F)$ and $\nabla_F f(B,F)$?

Differentiating w.r.t. B,
$\nabla_B \rho_\kappa(\bar{R} - BF) = \nabla_B \rho_\kappa(\bar{R} - IBF)$
$= -I^T \nabla \rho_\kappa(\bar{R} - IBF) F^T$
$= -\nabla \rho_\kappa(\bar{R} - BF) F^T$

Differentiating w.r.t. F,
$\nabla_F \rho_\kappa(\bar{R} - BF) = \nabla_F \rho_\kappa(\bar{R} - BFI)$
$= -B^T \nabla \rho_\kappa(\bar{R} - BFI)$
$= -B^T \nabla \rho_\kappa(\bar{R} - BF)$

(2) What is $\text{prox}_{\alpha g}(Z)$? Hint: there is a closed form solution that uses the SVD.

$\text{prox}_{\alpha g}(Z) = argmin_U(\frac{1}{2\alpha}\|U - Z\|_F^2)$. Let SVD of $Z = U \Sigma V^T$, We assume $U^T U = V V^T = I$
This is equivalent to,
$\|Z - Q\|_F = \|U \Sigma V^T - Q\|_F = \|\Sigma - U^T Q V\|_F = \|\Sigma - U^T Q V\|_F$

Minimizing all Orthogonal matrices Q, we get
$\|\Sigma - U^T QV\|_F^2 = \Sigma_k(\Sigma_{kk} - Q_{kk})^2 + \Sigma_{j\neq k} Q_{jk}^2$

$= \Sigma_k(\Sigma_{kk}^2 + Q_{kk}^2 - 2\Sigma_{kk}Q_{kk}) + \Sigma_{j\neq k} Q_{jk}^2$

$= \Sigma_k(\Sigma_{kk}^2 - 2\Sigma_{kk}Q_{kk}) + \Sigma_{j,k} Q_{jk}^2$

$= Tr(\Sigma^2) + Tr(Q^T Q) - 2\Sigma_{kk}Q_{kk})$

$= Tr(\Sigma^2) + n - 2\Sigma_{kk}Q_{kk})$

All the entries of Σ are diagonal and positive and from $Q \in [-1, 1]$. To find the minimum value, we should concentrate on values of $Q = 1$.

Thus, $\|\Sigma - I\|_F = \|U\Sigma V^T - UV^T\|_F = \|Z - UV^T\|_F$

(3) Implement PALM algorithm in **Jupyter Notebook** over S&P 500 data.

(4) After obtain the optimal solution B_*, F_*, calculate the SVD of $B_*F_* = U\Sigma V$ and report the first column of U which corresponding to the most risky portfolio.