MACHINE LEARNING FOR COMPUTATIONAL FINANCE
ASSIGNMENT 3

KARTHIK IYER

***Problem 1**: Matrix Norms*

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

$$A = U\Sigma V^\top,$$

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. $\text{diag}(\Sigma) = [\sigma_1, \dots, \sigma_n]^\top \triangleq \sigma(A)$, and $\sigma_1 \geq \dots \geq \sigma_n \geq 0$. Here are some useful properties,

- $\|A\|_F^2 = \text{trace}(A^\top A)$.
- $\text{trace}(AB) = \text{trace}(BA)$.
- $\|A\|_2 = \sup_{\|x\|_2\leq 1} \|Ax\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2$.

Show the following result:

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

Solution: By the first property listed above, $\|A\|_F^2 = \text{trace}(A^\top A)$. Since $A = U\Sigma V^\top$, $\text{trace}(A^\top A) = \text{trace}(V\Sigma^2 V^\top)$. By the second property listed above, $\text{trace}(V\Sigma^2 V^\top) = \text{trace}(VV^\top\Sigma^2) = \text{trace}(\Sigma^2)$.

Note that $\text{trace}(\Sigma^2) = \sum_{i=1}^n \sigma_i^2$. And since $\|\sigma(A)\|^2 = \sum_{i=1}^n \sigma_i^2$, it follows that

$\|A\|_F^2 = \|\sigma(A)\|_2^2 = \sum_{i=1}^n \sigma_i^2$. ■

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

Solution: We first note that for any vector $y \in \mathbb{R}^n$, if $U \in \mathbb{R}^{m\times n}$ is such that U has orthornormal columns, then $||Uy||_2 = ||y||_2$. Hence for any vector $x \in \mathbb{R}^n$, if the SVD of $A = U\Sigma V^\top$, then $||Ax||_2 = \|U\Sigma V^\top x\|_2 = \|\Sigma V^\top x\|_2$. (The assumption that U has orthonormal columns is indeed justified since $m > n$ and we are looking at the reduced SVD.)

Next, note that $||x||_2 = 1$ iff $||V^\top x||_2 = 1$. Hence,
$\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2 = \sup_{\|x\|_2=1} \|\Sigma V^\top x\|_2 = \sup_{\|V^\top x\|_2=1} \|\Sigma V^\top x\|_2 = \sup_{\|y\|_2=1} \|\Sigma y\|_2$.

Since Σ is a diagonal matrix with positive entries, $\|\Sigma\|_2 = \sigma_1$ (the largest singular value of A). Moreover, it is clear that $\sigma_1 = \|\sigma(A)\|_\infty$ and hence $\|A\|_2 = \|\sigma(A)\|_\infty = \sigma_1$. ■

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

Solution: Since $A = U\Sigma V^\top$, $A^\top A = V\Sigma^2 V^\top$. If we define $B = V\Sigma V^\top$, then $B^2 = A^\top A$. (Since the choice of such a V in the singular value decomposition is not unique, the choice of B is also not unique.) $\|A\|_* = \text{trace}\left(\sqrt{A^\top A}\right) = \text{trace}(B) = \text{trace}(V\Sigma V^T) = \text{trace}(VV^\top\Sigma) = \text{trace}(\Sigma)$. (The second to last equality follows since $\text{trace}(XY) = \text{trace}(YX)$.)

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

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

Consider optimization problem,

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

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

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

Solution: By Question 1 above, $\frac{1}{2}\|R - X\|_F^2 = \text{trace}((R - X^\top)(R - X))$.

Before we proceed, let us collect two facts without proof.

<u>Fact 1:</u>

$\nabla_X \text{trace}(AXB) = A^\top B^\top$

<u>Fact 2:</u>

$\nabla_X \text{trace}(AX^T B) = BA$

Hence,

$$\begin{aligned}\nabla f(X) &= \frac{1}{2}\nabla_X\left(\text{trace}(RR^\top) - \text{trace}(RX^\top) - \text{trace}(XR^\top) - \text{trace}(XX^\top)\right)\\ &= X - R \text{ (Using Fact 1 and Fact 2)}\end{aligned}$$

Clearly, $\nabla^2 f(X) = I$ and thus F is a β-smooth function with $\beta = ||\nabla^2 f||_2 = ||I||_2 = 1$. ■

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

Proof. $\text{prox}_{\alpha g}(Z) = \text{argmin}_U\left(\lambda||U||_* + \frac{1}{2\alpha}\|U - Z\|_F^2\right) = \text{argmin}_U\left(||U||_* + \frac{1}{2\alpha\lambda}\|U - Z\|_F^2\right)$.

To solve this, take the singular value decomposition, $Z = U\sigma(Z)V^\top$ and apply soft-thresholding to the singular values (Theorem 2.1 in [1]) to get $\text{prox}_{\alpha g}(Z) = U\hat{\Sigma}V^\top$, where $\hat{\Sigma}$ is a $n \times n$ diagonal matrix with ith diagonal entry equal to $\hat{\sigma}_i$ where

$$\hat{\sigma}_i = \begin{cases}\sigma_i - \lambda\alpha, & \text{when } \sigma_i > \alpha\lambda\\ 0 \text{ otherwise}\end{cases}$$

Here σ_i denotes the ith singular value of A. ■

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

Solution: See the accompnying Jupyter notebook for the code. For the given value of $\lambda = 30$, we obtain the rank of the optimal solution to be 29. ■

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

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

which we denote as X_k.

Proof. Let u_i denote the ith column of U and let v_i denote the ith column of V where U and V are such that $A = U\Sigma V^\top$ (SVD of A.) It is well known that we can write A as a sum of rank 1 matrices of the form $A = \sum_{i=1}^n \sigma_i u_i v_i^\top$.

Let us define $A_k = \sum_{i=1}^k \sigma_i u_i v_i^\top$. It is clear that A_k has rank k. We claim that among all matrices B of rank atmost k,

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

Let us prove this claim. Assume B minimizes $||A - B||_F^2$ among all matrices of rank atmost k. Let V be the vector space spanned by the rows of B. The dimension of V is at most k. Since B minimizes $||A - B||_F^2$, it must be that each row of B is the projection of corresponding row of A on to V. If not, replace each row of B with the projection of the corresponding row of A onto V. This will not change V and thus the rank of B but it reduces $||A - B||_F^2$. Since each row of B is the projection of the corresponding row of A, it follows that $||A - B||_F^2$ is the sum of squared distances of rows of A to V. Since A_k minimizes the sum of squared distance of rows of A to any k-dimensional subspace , it follows that $||A - A_k|| \leq ||A - B||_F$ and our claim is justified.

Hence $X_k = \sum_{i=1}^k \sigma_i u_i v_i^\top$, where $R = U\Sigma V^\top$. Moreover, $\frac{1}{2}||X_k - R||_F^2 = \sum_{i=k+1}^n \sigma_i^2$. ∎

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

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

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

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

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

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

Solution: Using the facts we mentioned in Problem 2 and the characterization of the Frobenius norm in terms of trace,
$||R - BF||_F^2 = \operatorname{trace}(RR^\top) - \operatorname{trace}(RF^\top B^\top) - \operatorname{trace}(BFR^\top) + \operatorname{trace}(BFF^\top B^\top)$. Using the facts mentioned in **Problem 2**, we can conclude that

$$\nabla_B f(B, F) = (BF - R)F^\top.$$

Notice that $||\nabla_B^2 f(B, F)||_2 = ||F^\top F||_2$. Appealing again to the facts mentioned in Problem 2 and characterization of the Frobenius norm in terms of trace,

$$\nabla_F f(B, F) = B^\top(BF - R).$$

Notice that $||\nabla_F^2 f(B, F)||_2 = ||B^\top B||_2$. ∎

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

Proof. The PALM algorithm is implemented in the accompanying Jupyter notebook. ■

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

Proof. It makes more sense to look at the percent change. We therefore compute the quantity $\frac{\|X_k - B_*F_*\|}{\|X_k\|}$ and it turns out that there is only about 6% change in the answers for the two algorithms. ■

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,

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

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

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

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

Solution:

$$\begin{aligned} \nabla_B \rho_\kappa(\bar{R} - BF) &= \nabla_B \rho_\kappa(\bar{R} - IBF) \\ &= -I^\top \nabla \rho_\kappa(\bar{R} - IBF)F^\top \\ &= -\nabla \rho_\kappa(\bar{R} - BF)F^\top. \end{aligned}$$

Similarly,

$$\begin{aligned} \nabla_F \rho_\kappa(\bar{R} - BF) &= \nabla_F \rho_\kappa(\bar{R} - BFI) \\ &= -B^\top \nabla \rho_\kappa(\bar{R} - BFI)I^\top \\ &= -B^\top \nabla \rho_\kappa(\bar{R} - BF). \end{aligned}$$

■

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

Proof. Note that $\text{prox}_{\alpha g}(Z) = \text{argmin}_U \left(g(U) + \frac{1}{2\alpha}\|U - Z\|_F^2 \right) = \text{argmin}_{U : U^\top U = I} \left(\frac{1}{2\alpha}\|U - Z\|_F^2 \right)$. Consider the SVD of Z of the form $Z = U\Sigma V^\top$ where $U^\top U = VV^\top = I$. Finding the prox is now equivalent to finding the closest (in Frobenius norm) orthogonal matrix to Z.

Since the Frobenus norm, like the 2-norm) is unitary equivalent,

$$\|Z - Q\|_F = \|U\Sigma V^T - Q\|_F = \|\Sigma - U^T Q V\|_F.$$

This is the same as saying that we want to minimize

$$\|\Sigma - Q\|_F.$$

over all orthogonal matrices Q. We have,

$$\begin{aligned}\|\Sigma - Q\|_F^2 &= \sum_k (\Sigma_{kk} - Q_{kk})^2 + \sum_{j\neq k} Q_{kj}^2 \\ &= \sum_k (\Sigma_{kk}^2 + Q_{kk}^2 - 2\Sigma_{kk}Q_{kk}) + \sum_{j\neq k} Q_{kj}^2 \\ &= \sum_k (\Sigma_{kk}^2 - 2\Sigma_{kk}Q_{kk}) + \sum_{j,k} Q_{kj}^2 \\ &= \mathrm{Tr}(\Sigma^2) + \mathrm{Tr}(Q^TQ) - 2\sum_k \Sigma_{kk}Q_{kk} \\ &= \mathrm{Tr}(\Sigma^2) + n - 2\sum_k \Sigma_{kk}Q_{kk}\end{aligned}$$

To minimize this quantity over Q, since the entries of Σ are non-negative and $Q_{kk} \in [-1, 1]$, we need to choose $Q_{kk} = 1$ for all k, which makes $Q = I$ (Note that columns of Q are othonormal and hence are unit Euclidean vectors. If any one component of a unit Eucidean vector is 1, then all other components are forced to be 0.) Hence the minimum is

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

(I learnt this argument from [2].) ■

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

See the accompanying Jupyter notebook for the code.

(4) After obtaining 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.

References

[1] https://arxiv.org/pdf/0810.3286.pdf
[2] https://math.stackexchange.com/questions/2215359/showing-that-matrix-q-uvt-is-the-nearest-orthogonal-matrix-to-a