Homework 1 Solution: Mathematics of Deep Learning (EN 580.745)

Instructor: René Vidal, Biomedical Engineering, Johns Hopkins University

1. Properties of the ℓ_1 norm. Let $\boldsymbol{x} \in \mathbb{R}^n$ and recall the ℓ_1 norm $\|\boldsymbol{x}\|_1 \triangleq \sum_i |x_i|$.

(a) **(10 points)** The subdifferential $\partial f(\boldsymbol{x})$ of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x}$ is defined to be

$$\partial f(\boldsymbol{x}) \triangleq \{\boldsymbol{z} \in \mathbb{R}^n \mid f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{z}, \boldsymbol{y} - \boldsymbol{x} \rangle \text{ for all } \boldsymbol{y} \in \mathbb{R}^n\}. \tag{1}$$

Prove that

$$\partial \|\boldsymbol{x}\|_1 = \{\operatorname{sign}(\boldsymbol{x}) + \boldsymbol{w} \mid \operatorname{supp}(\boldsymbol{w}) \subseteq \operatorname{supp}(\boldsymbol{x})^c,\ \max_i |w_i| \leq 1\}, \tag{2}$$

where $\operatorname{sign}\colon \mathbb{R}^n \to \{-1, 0, 1\}$ denotes the sign function, and $\operatorname{supp}(\boldsymbol{x}) \subseteq [n]$ denotes the support of $\boldsymbol{x}$, that is the set of indices where $\boldsymbol{x}$ is non-zero.

(b) **(10 points)** Define the proximal operator $\operatorname{prox}_f \colon \mathbb{R}^n \to \mathbb{R}^n$ of a convex function f to be

$$\operatorname{prox}_f(\boldsymbol{x}) \triangleq \operatorname*{arg\,min}_{\boldsymbol{a} \in \mathbb{R}^n} \frac{1}{2}\|\boldsymbol{a} - \boldsymbol{x}\|_2^2 + f(\boldsymbol{a}). \tag{3}$$

Let $\tau > 0$. Prove that $\operatorname{prox}_{\tau\|\cdot\|_1}(\boldsymbol{x}) = \mathcal{S}_\tau(\boldsymbol{x})$, where $\mathcal{S}_\tau$ is the elementwise soft-thresholding operator

$$\mathcal{S}_\tau(x_i) \triangleq \operatorname{sign}(x_i) \max(|x_i| - \tau, 0). \tag{4}$$

Solution.

(a) We'll first prove the statement for the special case $n = 1$. Fix $x \in \mathbb{R}$ and suppose $z = \operatorname{sign}(x) + w$, with w as in (2). Choose an arbitrary y and note that

$$|x| + (\operatorname{sign}(x) + w)(y - x) = (\operatorname{sign}(x) + w)y \leq |y|, \tag{5}$$

where in the first equality we used $\operatorname{sign}(x)x = |x|$ and $wx = 0$, and in the second inequality we used $|\operatorname{sign}(x) + w| \leq 1$. This proves the $\supseteq$ containment in (2) using the definition of the subdifferential.

For the other direction, note first that if $x \neq 0$, then $|\cdot|$ is differentiable at x with $\frac{d}{dx}|x| = \operatorname{sign}(x)$. So suppose that $x = 0$ and assume $z \in \partial|x|$. Then by definition we must have $|y| \geq zy$ for all y, which implies $|z| \leq 1$.

This completes the proof for the case $n = 1$. To show the general case, note that $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |\boldsymbol{e}_i^\top \boldsymbol{x}|$, where $\boldsymbol{e}_i$ is the ith standard basis element. Then by the summation and affine transformation rules of subdifferential calculus[1], we know

$$\partial\|\boldsymbol{x}\|_1 = \sum_{i=1}^n \boldsymbol{e}_i \partial|\boldsymbol{e}_i^\top \boldsymbol{x}|. \tag{6}$$

By the special case $n = 1$, the RHS of (6) is equal to the RHS of (2).

(b) The objective in (3) is the sum of a convex and strongly convex function, hence is strongly convex and a unique minimizer exists. Denote this minimizer $\boldsymbol{a}^*$. The (necessary and sufficient) first order optimality condition states that $\boldsymbol{a}^*$ must satisfy $\boldsymbol{x} - \boldsymbol{a}^* \in \tau\partial\|\boldsymbol{a}^*\|_1$. By part (a), this implies $\boldsymbol{x} - \boldsymbol{a}^* = \tau(\operatorname{sign}(\boldsymbol{a}^*) + \boldsymbol{w})$, for $\boldsymbol{w}$ as in (2). We can now check that setting $\boldsymbol{a}^* = \mathcal{S}_\tau(\boldsymbol{x})$ satisfies this condition

$$x_i - \mathcal{S}_\tau(x_i) = \begin{cases} \tau \operatorname{sign}(\mathcal{S}_\tau(x_i)) & |x_i| > \tau \\ x_i & \text{o.w.} \end{cases} \tag{7}$$

□

[1] https://web.stanford.edu/class/ee364b/lectures/subgradients_slides.pdf

2. Properties of the nuclear norm. Let $\boldsymbol{X} \in \mathbb{R}^{D \times N}$ be a matrix of rank r. Recall the nuclear norm $\|\boldsymbol{X}\|_* \triangleq \sum_{i=1}^r \sigma_i(\boldsymbol{X})$, where $\sigma_i(\boldsymbol{X})$ denotes the ith singular value of $\boldsymbol{X}$. Let $\boldsymbol{X} = \boldsymbol{U}\Sigma\boldsymbol{V}^\top$ be the compact SVD, so that $\boldsymbol{U} \in \mathbb{R}^{D \times r}$, $\Sigma \in \mathbb{R}^{r \times r}$, and $\boldsymbol{V} \in \mathbb{R}^{N \times r}$. Recall also the spectral norm $\|\boldsymbol{X}\|_2 = \sigma_1(\boldsymbol{X})$.

(a) **(10 points)** Prove that [2]

$$\partial\|\boldsymbol{X}\|_* = \{\boldsymbol{U}\boldsymbol{V}^\top + \boldsymbol{W} \mid \boldsymbol{U}^\top\boldsymbol{W} = \boldsymbol{0}, \boldsymbol{W}\boldsymbol{V} = \boldsymbol{0}, \|\boldsymbol{W}\|_2 \le 1\}. \tag{8}$$

(b) **(10 points)** Let $\tau > 0$. Prove that $\operatorname{prox}_{\tau\|\cdot\|_*}(\boldsymbol{X}) = \boldsymbol{U}\mathcal{S}_\tau(\Sigma)\boldsymbol{V}^\top$, where $\mathcal{S}_\tau$ is as in (4). Note that when generalizing (3) to matrices, the squared Frobenius norm $\frac{1}{2}\|\boldsymbol{A} - \boldsymbol{X}\|_F^2$ is used in place of the squared ℓ_2 norm ($\|\boldsymbol{X}\|_F \triangleq (\sum_{ij} X_{ij}^2)^{\frac{1}{2}}$).

Solution.

(a) Let $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{V}^\top + \boldsymbol{W}$ with $\boldsymbol{W}$ as in (8), and fix an arbitrary $\boldsymbol{Y}$. Then we have

$$\|\boldsymbol{X}\|_* + \langle \boldsymbol{Z}, \boldsymbol{Y} - \boldsymbol{X}\rangle = \langle \boldsymbol{Z}, \boldsymbol{Y}\rangle \le \|\boldsymbol{Z}\|_2\|\boldsymbol{Y}\|_* \le \|\boldsymbol{Y}\|_*, \tag{9}$$

where the first equality uses $\langle \boldsymbol{Z}, \boldsymbol{X}\rangle = \|\boldsymbol{X}\|_*$, the second inequality uses the Von-Neumann trace inequality and Hölder's inequality, and the last inequality uses $\|\boldsymbol{Z}\|_2 \le 1$, which holds by the definition (8). This proves the $\supseteq$ containment.

For the other direction, fix $\boldsymbol{Z} \in \partial\|\boldsymbol{X}\|_*$. Let $\|\cdot\|_*^*$ denote the convex conjugate of $\|\cdot\|_*$

$$\|\boldsymbol{Z}\|_*^* \triangleq \sup_{\boldsymbol{Y}} \langle \boldsymbol{Z}, \boldsymbol{Y}\rangle - \|\boldsymbol{Y}\|_*. \tag{10}$$

Note that again by the Von-Neumann trace inequality and Hölder's inequality, we can say

$$\|\boldsymbol{Z}\|_*^* \le \sup_{\boldsymbol{Y}} (\|\boldsymbol{Z}\|_2 - 1)\|\boldsymbol{Y}\|_* = \begin{cases} \infty & \|\boldsymbol{Z}\|_2 > 1 \\ 0 & \text{o.w.} \end{cases} \tag{11}$$

Furthermore, the upper bound can always be achieved by choosing $\boldsymbol{Y}$ to be rank one with top-1 left and right singular vectors equal to those of $\boldsymbol{Z}$. So, (11) is in fact an equality. Now by a fundamental property of the convex conjugate, we know that $\langle \boldsymbol{Z}, \boldsymbol{X}\rangle = \|\boldsymbol{X}\|_* + \|\boldsymbol{Z}\|_*^*$, which implies $\|\boldsymbol{Z}\|_*^* = 0$, hence $\|\boldsymbol{Z}\|_2 \le 1$ and $\langle \boldsymbol{Z}, \boldsymbol{X}\rangle = \|\boldsymbol{X}\|_*$. Applying the trace inequality once more, we know $\langle \boldsymbol{Z}, \boldsymbol{X}\rangle = \|\boldsymbol{X}\|_*$ only if $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{V}^\top + \boldsymbol{W}$ as in (8).

(b) As in problem 1(b), we know the unique minimizer $\boldsymbol{A}^*$ of the proximal optimization problem uniquely satisfies the optimality condition $\boldsymbol{X} - \boldsymbol{A}^* \in \tau\partial\|\boldsymbol{A}^*\|_*$. Indeed, $\boldsymbol{U}\mathcal{S}_\tau(\Sigma)\boldsymbol{V}^\top$ satisfies this condition

$$\boldsymbol{X} - \boldsymbol{U}\mathcal{S}_\tau(\Sigma)\boldsymbol{V}^\top = \boldsymbol{U}\begin{pmatrix} \tau\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \Sigma_1 \end{pmatrix}\boldsymbol{V}^\top = \boldsymbol{U}_0(\tau\boldsymbol{I})\boldsymbol{V}_0^\top + \boldsymbol{U}_1\Sigma_1\boldsymbol{V}_1^\top, \tag{12}$$

where $\boldsymbol{U} = [\boldsymbol{U}_0\ \boldsymbol{U}_1]$, $\boldsymbol{V} = [\boldsymbol{V}_0\ \boldsymbol{V}_1]$, $\operatorname{diag}(\Sigma)^\top = [\operatorname{diag}(\Sigma_0)^\top\ \operatorname{diag}(\Sigma_1)^\top]$ with Σ_0 containing all singular values greater than τ.

□

3. Low rank matrix factorization. Let $\boldsymbol{Y} \in \mathbb{R}^{D \times N}$ be of rank r. Consider the matrix factorization problem

$$\underset{\boldsymbol{U},\boldsymbol{V}}{\text{minimize}} \quad F(\boldsymbol{U}, \boldsymbol{V}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{V}^\top\|_F^2 + \frac{\tau}{2}(\|\boldsymbol{U}\|_F^2 + \|\boldsymbol{V}\|_F^2), \tag{13}$$

where $\boldsymbol{U} \in \mathbb{R}^{D \times r}$, $\boldsymbol{V} \in \mathbb{R}^{N \times r}$, and $\tau > 0$. This is in fact a regularized variant of the problem studied in (Baldi & Hornik, 1989). It is also closely related to the nuclear norm proximal operator from Problem 2(b), as we will see.

(a) **(10 points)** Let $\boldsymbol{Y} = \boldsymbol{P}\Sigma\boldsymbol{Q}^\top$ be the compact SVD. Define

$$\widehat{\boldsymbol{U}} = \boldsymbol{P}\mathcal{S}_\tau(\Sigma)^{\frac{1}{2}} \qquad \widehat{\boldsymbol{V}} = \boldsymbol{Q}\mathcal{S}_\tau(\Sigma)^{\frac{1}{2}}, \tag{14}$$

where the square root of a diagonal matrix is applied elementwise. Prove that $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}})$ is a critical point of (13).

[2] *Hint:* the Von-Neumann trace inequality could be useful.

(b) **(bonus)** Prove that $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}})$ is in fact a global minimizer.

Problem (13) and the nuclear norm proximal operator are therefore closely related, in the sense that the global minimizer $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}})$ satisfies $\widehat{\boldsymbol{U}}\widehat{\boldsymbol{V}}^\top = \mathrm{prox}_{\tau\|\cdot\|_*}(\boldsymbol{Y})$.

Solution.

(a) First, compute the gradients with respect to $\boldsymbol{U}, \boldsymbol{V}$ to find the first order optimality conditions

$$\nabla_{\boldsymbol{U}} F(\boldsymbol{U}, \boldsymbol{V}) = (\boldsymbol{U}\boldsymbol{V}^\top - \boldsymbol{Y})\boldsymbol{V} + \tau\boldsymbol{U} = \boldsymbol{0} \qquad \nabla_{\boldsymbol{V}} F(\boldsymbol{U}, \boldsymbol{V}) = (\boldsymbol{U}\boldsymbol{V}^\top - \boldsymbol{Y})^\top\boldsymbol{U} + \tau\boldsymbol{V} = \boldsymbol{0} \tag{15}$$

$$\Rightarrow \boldsymbol{U}(\boldsymbol{V}^\top\boldsymbol{V} + \tau\boldsymbol{I}) = \boldsymbol{Y}\boldsymbol{V} \qquad \Rightarrow \boldsymbol{V}(\boldsymbol{U}^\top\boldsymbol{U} + \tau\boldsymbol{I}) = \boldsymbol{Y}^\top\boldsymbol{U}. \tag{16}$$

Now, note that

$$\widehat{\boldsymbol{U}}(\widehat{\boldsymbol{V}}^\top\widehat{\boldsymbol{V}} + \tau\boldsymbol{I}) = \boldsymbol{P}\mathcal{S}_\tau(\Sigma)^{\frac{1}{2}}(\mathcal{S}_\tau(\Sigma) + \tau\boldsymbol{I}) = \boldsymbol{P}\Sigma\mathcal{S}_\tau(\Sigma)^{\frac{1}{2}} = \boldsymbol{Y}\widehat{\boldsymbol{V}}, \tag{17}$$

where the middle equality holds because

$$\max(\sigma_i - \tau, 0)^{\frac{1}{2}}(\max(\sigma_i - \tau, 0) + \tau) = \sigma_i \max(\sigma_i - \tau, 0)^{\frac{1}{2}}. \tag{18}$$

Similarly,

$$\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{U}}^\top\widehat{\boldsymbol{U}} + \tau\boldsymbol{I}) = \boldsymbol{Q}\mathcal{S}_\tau(\Sigma)^{\frac{1}{2}}(\mathcal{S}_\tau(\Sigma) + \tau\boldsymbol{I}) = \boldsymbol{Q}\Sigma\mathcal{S}_\tau(\Sigma)^{\frac{1}{2}} = \boldsymbol{Y}^\top\widehat{\boldsymbol{U}}. \tag{19}$$

Thus, $\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}$ satisfy the first order optimality conditions.

(b) We first prove the equivalence of the nuclear norm with its so-called "variational form".

Lemma 1 (Variational form of the nuclear norm)**.** *Let $\boldsymbol{Y}$ be as defined above. Then*

$$\|\boldsymbol{Y}\|_* = \min_{\substack{\boldsymbol{U}\in\mathbb{R}^{D\times r}, \boldsymbol{V}\in\mathbb{R}^{N\times r} \\ \boldsymbol{Y}=\boldsymbol{U}\boldsymbol{V}^\top}} \frac{1}{2}(\|\boldsymbol{U}\|_F^2 + \|\boldsymbol{V}\|_F^2). \tag{20}$$

In particular, the minimum in (20) exists.

Proof. Let $\boldsymbol{P}$, Σ, $\boldsymbol{Q}$ also be as defined above. Then note that setting $\boldsymbol{U} = \boldsymbol{P}\Sigma^{\frac{1}{2}}$ and $\boldsymbol{V} = \boldsymbol{Q}\Sigma^{\frac{1}{2}}$ satisfies $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{V}^\top$ and $\|\boldsymbol{Y}\|_* = \frac{1}{2}(\|\boldsymbol{U}\|_F^2 + \|\boldsymbol{V}\|_F^2)$. This proves

$$\|\boldsymbol{Y}\|_* \geq \inf_{\substack{\boldsymbol{U},\boldsymbol{V} \\ \boldsymbol{Y}=\boldsymbol{U}\boldsymbol{V}^\top}} \frac{1}{2}(\|\boldsymbol{U}\|_F^2 + \|\boldsymbol{V}\|_F^2). \tag{21}$$

It remains to show the corresponding upper bound. So choose any $\boldsymbol{U}, \boldsymbol{V}$ such that $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{V}^\top$. Let $\boldsymbol{U}' = \boldsymbol{P}^\top\boldsymbol{U}$ and $\boldsymbol{V}' = \boldsymbol{Q}^\top\boldsymbol{V}$. Then $\Sigma = \boldsymbol{U}'\boldsymbol{V}'^\top$. Moreover, since $\boldsymbol{P}$ and $\boldsymbol{Q}$ have orthonormal columns, $\|\boldsymbol{U}\|_F^2 \geq \|\boldsymbol{U}'\|_F^2$ and $\|\boldsymbol{V}\|_F^2 \geq \|\boldsymbol{V}'\|_F^2$. Now, let σ_j be the jth singular value, and let $\boldsymbol{u}_j'$, $\boldsymbol{v}_j'$ be the jth rows of $\boldsymbol{U}'$, $\boldsymbol{V}'$. Then we have

$$\sigma_j = \boldsymbol{u}_j'^\top\boldsymbol{v}_j' \leq \|\boldsymbol{u}_j'\|_2\|\boldsymbol{v}_j'\|_2 \leq \frac{1}{2}(\|\boldsymbol{u}_j'\|_2^2 + \|\boldsymbol{v}_j'\|_2^2), \tag{22}$$

where the first upper bound is Cauchy-Schwarz, the second is the AM-GM inequality. It follows that

$$\|\boldsymbol{Y}\|_* = \sum_j \sigma_j \leq \frac{1}{2}\sum_j(\|\boldsymbol{u}_j'\|_2^2 + \|\boldsymbol{v}_j'\|_2^2) = \frac{1}{2}(\|\boldsymbol{U}'\|_F^2 + \|\boldsymbol{V}'\|_F^2), \tag{23}$$

completing the proof of the lemma. □

By Lemma 1, we now know that for all $\boldsymbol{U}, \boldsymbol{V}$

$$\frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{V}^\top\|_F^2 + \tau\|\boldsymbol{U}\boldsymbol{V}^\top\|_* \leq F(\boldsymbol{U}, \boldsymbol{V}). \tag{24}$$

Thus,

$$\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{X}\|_F^2 + \tau\|\boldsymbol{X}\|_* \leq \min_{\boldsymbol{U},\boldsymbol{V}} F(\boldsymbol{U}, \boldsymbol{V}). \tag{25}$$

Since the LHS of (25) is the nuclear norm proximal optimization problem, and by construction $\widehat{\boldsymbol{U}}\widehat{\boldsymbol{V}}'^{\top} = \operatorname{prox}_{\tau\|\cdot\|_*}(\boldsymbol{Y})$, it follows that $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}})$ is a global minimizer of (13).

□

Proposition 1 (Von-Neumann trace inequality)**.** *Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{D\times N}$. Assume without loss of generality that $r = \operatorname{rank}(\boldsymbol{X}) \geq \operatorname{rank}(\boldsymbol{Y}) = r'$. Let $\boldsymbol{X} = \boldsymbol{U}_X \Sigma_X \boldsymbol{V}_X^\top$ and $\boldsymbol{Y} = \boldsymbol{U}_Y \Sigma_Y \boldsymbol{V}_Y^\top$ be their respective compact **rank-**r SVDs. Also let $\boldsymbol{U}'_X \Sigma'_X \boldsymbol{V}'^{\top}_X$ and $\boldsymbol{U}'_Y \Sigma'_Y \boldsymbol{V}'^{\top}_Y$ contain just the top r' singular vectors. Then*

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle \leq \langle \Sigma_X, \Sigma_Y \rangle, \tag{26}$$

with equality if and only if $\boldsymbol{U}'^{\top}_Y \boldsymbol{U}'_X = \boldsymbol{V}'^{\top}_Y \boldsymbol{V}'_X = \Pi \in \mathbb{R}^{r'\times r'}$, for a permutation matrix Π satisfying $\langle \Pi^\top \Sigma'_X \Pi, \Sigma'_Y \rangle = \langle \Sigma_X, \Sigma_Y \rangle$. If all singular values are unique, $\Pi = \boldsymbol{I}$.

Proof. Let $\boldsymbol{P} = \boldsymbol{U}_Y^\top \boldsymbol{U}_X$, $\boldsymbol{Q} = \boldsymbol{V}_Y^\top \boldsymbol{V}_X$, and $\widehat{\boldsymbol{X}} = \boldsymbol{P}\Sigma_X \boldsymbol{Q}^\top$. Note that both $\boldsymbol{P}$ and $\boldsymbol{Q}$ are $r \times r$ orthogonal matrices. By the cyclic property of trace, we have $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \langle \widehat{\boldsymbol{X}}, \Sigma_Y \rangle$. Furthermore, note that

$$\langle \widehat{\boldsymbol{X}}, \Sigma_Y \rangle = \sum_{i=1}^{r'} \sigma_{Y\,i}\, \boldsymbol{p}_i^\top \Sigma_X \boldsymbol{q}_i, \tag{27}$$

where the $\boldsymbol{p}_i$, $\boldsymbol{q}_i$ are rows of $\boldsymbol{P}$, $\boldsymbol{Q}$ respectively. Now consider the optimization problem

$$\max_{\boldsymbol{P},\boldsymbol{Q}} \quad \sum_{i=1}^{r'} \sigma_{Y\,i}\, \boldsymbol{p}_i^\top \Sigma_X \boldsymbol{q}_i \quad \text{s.t.} \quad \boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}, \tag{28}$$

whose optimal value is an (achievable) upper bound for (27). By induction on r', one can show that the maximizers $(\boldsymbol{P}^*, \boldsymbol{Q}^*)$ of (28) satisfy $\boldsymbol{p}_i^* = \boldsymbol{q}_i^* = \boldsymbol{e}_{\pi(i)} \in \mathbb{R}^r$ for $i = 1, \ldots, r'$ and some permutation $\pi \colon [r'] \to [r']$ such that $\sigma_{Y\,i}\, \boldsymbol{p}_i^{*\top} \Sigma_X \boldsymbol{q}_i^* = \sigma_{Y\,i} \sigma_{X\,\pi(i)} = \sigma_{Y\,i} \sigma_{X\,i}$. Taking $\Pi = [\boldsymbol{e}_{\pi(1)} \cdots \boldsymbol{e}_{\pi(r')}] \in \mathbb{R}^{r'\times r'}$, this also means $\langle \Pi^\top \Sigma'_X \Pi, \Sigma'_Y \rangle = \langle \Sigma_X, \Sigma_Y \rangle$. Moreover, note that if all singular values are unique, π must be the identity. □