First, recall the definition of the convex conjugate (also known as the Fenchel dual) of an arbitrary function f.

Definition 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary function. Define the convex conjugate $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

$$f^*(\boldsymbol{z}) \triangleq \sup_{\boldsymbol{x}\in\mathbb{R}^n} \langle \boldsymbol{z}, \boldsymbol{x}\rangle - f(\boldsymbol{x}). \tag{1}$$

Note that f^* can take the value $+\infty$ if the difference in (1) can always be made larger. (E.g. consider $f(x) = |x|$.) Also, f^* is always convex regardless of f, since it is the pointwise supremum of affine functions. However, for convex f there is also a close connection between the convex conjugate and the subdifferential, which could be useful for the homework.

Proposition 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and f^* its conjugate. Then for any $\boldsymbol{x}$, $\boldsymbol{z}$ we have*

$$\langle \boldsymbol{x}, \boldsymbol{z}\rangle \le f(\boldsymbol{x}) + f^*(\boldsymbol{z}), \tag{2}$$

with equality if and only if $\boldsymbol{z} \in \partial f(\boldsymbol{x})$.

Intuitively, looking at Figure 1, you can see that the conjugate represents the maximum amount you can shift an affine function $\boldsymbol{x} \mapsto \langle \boldsymbol{z}, \boldsymbol{x}\rangle$ down, while still remaining above or equal to f *somewhere*. Moreover, Proposition 1 tells us that for those $\boldsymbol{x}$ achieving the supremum in (1) for $f^*(\boldsymbol{z})$, we must have $\boldsymbol{z} \in \partial f(\boldsymbol{x})$. I.e., $\boldsymbol{z}$ is a subgradient for points $\boldsymbol{x}$ that are the last to lose contact as you shift $\boldsymbol{x} \mapsto \langle \boldsymbol{z}, \boldsymbol{x}\rangle$.

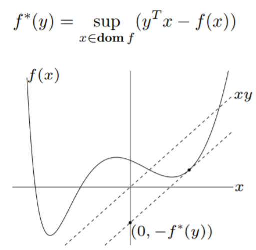

Figure 1: Illustration of the convex conjugate.

An unrelated but possibly also useful property from linear algebra is the Von-Neumann trace inequality.

Proposition 2 (Von-Neumann trace inequality)**.** *Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{D\times N}$. Assume without loss of generality that $r = \operatorname{rank}(\boldsymbol{X}) \ge \operatorname{rank}(\boldsymbol{Y}) = r'$. 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*

$$\langle \boldsymbol{X}, \boldsymbol{Y}\rangle \le \langle \Sigma_X, \Sigma_Y\rangle, \tag{3}$$

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

$$\langle \widehat{\boldsymbol{X}}, \Sigma_Y\rangle = \sum_{i=1}^{r'} \sigma_{Y\,i}\,\boldsymbol{p}_i^\top\Sigma_X\boldsymbol{q}_i, \tag{4}$$

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

$$\max_{\boldsymbol{P},\boldsymbol{Q}} \quad \sum_{i=1}^{r'} \sigma_{Y\,i}\,\boldsymbol{p}_i^\top\Sigma_X\boldsymbol{q}_i \quad \text{s.t.} \quad \boldsymbol{P}^\top\boldsymbol{P} = \boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}, \tag{5}$$

whose optimal value is an (achievable) upper bound for (4). By induction on r', one can show that the maximizers $(\boldsymbol{P}^*, \boldsymbol{Q}^*)$ of (5) 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. $\square$