<u>WP</u> a) def of subgrad: g of $f(x)$ @ x:

g s.t. $\quad f(y) - f(x) - \langle g, y-x \rangle \geq 0 \quad \forall\, y \in \mathrm{dom}(f)$

$\Rightarrow \quad Z = UV^T + W$ must satisfy:

$$\|Y\|_* - \|X\|_* - \langle UV^T + W, Y - X \rangle \geq 0 \quad \forall\, Y \in \mathbb{R}^{m \times n}$$

- rewriting:

$$\|Y\|_* - \|X\|_* - \underbrace{\langle UV^T + W, Y \rangle}_{\leq \|Y\|_* \text{ by given lemma } (2)} + \overbrace{\langle UV^T, X \rangle}^{= \|X\|_* \ (1)} + \cancelto{\emptyset}{\langle W, X \rangle}$$

- by definition of W

<u>Proof (1):</u>

$$\langle UV^T, X \rangle = \langle UV^T, U\Sigma V^T \rangle = \mathrm{Tr}(V\Sigma U^T U V^T) = \mathrm{Tr}(V\Sigma V^T)$$

- $V\Sigma V^T$ is a symm. matrix M,

$$\therefore \mathrm{Tr}(V\Sigma V^T) = \sum_i \lambda_i^M = \sum_j \sigma_j^X = \|X\|_*$$

<u>Proof (2):</u>

by formulation of dual norm:

$$\|Y\|_* = \max_{\{Z \,|\, \|Z\| \leq 1\}} \langle Z, Y \rangle \geq \langle Z_0, Y \rangle \quad \text{s.t. } \|Z_0\| \leq 1$$

- the dual norm of the nuclear norm is the spectral norm, i.e, maximal SV.

- additionally, $\|UV^T + W\|_{\text{spectral}} = 1$ as UV^T and W are complementary, and $\|W\|_{\text{spectral}} < 1$ while $\|UV^T\|_{\text{spectral}} = 1$

$$\Rightarrow \quad \langle UV^T + W \rangle \leq \max_{\{Z \,|\, \|Z\| \leq 1\}} \langle Z, Y \rangle = \|Y\|_*$$

$$\Rightarrow \quad \|Y\|_* - \|X\|_* - \langle UV^T + W, Y - X \rangle \geq \|Y\|_* - \|Y\|_* - \|X\|_* + \|X\|_* = \emptyset$$

$\square$

Convex OPT PS4 Justin Lewis

WP a) i) $D_\phi(x,y) = \Phi(x) - \phi(y) - \langle \nabla \phi(y), x-y \rangle$

$$\phi(x) = \sum_i x_i \log(x_i)$$

$$D_\phi(x,y) = \sum_i x_i \log(x_i) - \sum_i y_i \log(y_i) - \sum_i (1+\log(y_i))(x_i - y_i)$$

$$= \sum_i x_i \log\left(\frac{x_i}{y_i}\right)$$

$$= \boxed{KL(x \| y)}$$

ii) $$x^{t+1} = \underset{x \in \mathcal{X}}{\arg\min} \left(f(x^t) + \langle \nabla f(x^t), x - x^t \rangle + \frac{D_\phi(x, x^t)}{\eta_t} \right) \quad (*)$$

$\mathcal{X}$: prob. simplex

- from previous hw:

$$\underset{x \in \mathcal{X}}{\arg\min} \left(D_\phi(x \| y) \right) = \frac{y}{\|y\|_1}$$

- so rewriting $(*)$:

$$\Rightarrow \underset{x \in \mathcal{X}}{\arg\min} \left\{ \sum_i \log\left(\exp(\eta_t \cdot \nabla f(x^t)_i)\right) \cdot x_i + x_i \log\left(\frac{x_i}{x^t_i}\right) \right\}$$

$$= \underset{x \in \mathcal{X}}{\arg\min} \left(\sum_i x_i \log\left(\frac{x_i}{x^t_i \cdot \exp(-\eta_t \nabla f(x^t)_i)}\right)\right)$$

$$\Rightarrow \boxed{x_i^{t+1} = \frac{x^t_i \cdot \exp(-\eta_t \nabla f(x^t)_i)}{\sum_{j=1}^{n} x^t_j \exp(-\eta_t \nabla f(x^t)_j)}}$$

- by using result from previous hw

(which is essentially the same method suggested)

WP d)

- as Q is symmetric, it has real eigenvalues with an orthonormal basis of eigenvectors

- let $\mathbb{V}_+$, $\mathbb{V}_-$, and $\mathbb{V}_0$ be the set of eigenvectors with associated eigenvalues that are posi, neg, and $\emptyset$ (respectively).

$$\therefore \quad x = \sum_{i=1}^{|\mathbb{V}_+|} \alpha_i V_+[i] + \sum_{j=1}^{|\mathbb{V}_-|} \alpha_j V_-[j] + \sum_{k=1}^{|\mathbb{V}_0|} \alpha_k V_0[k] \quad \forall x \quad (1)$$

- now, $x^{t+1} = x^t - \eta \nabla f(x^t)$
$$= x^t - \eta 2 Q x^t$$

- notice, that following the gradient @ x which is in one of the distinct subspaces keeps x in that subspace:

e.g.: $x^t = V_-[0] \implies x^{t+1} = x^t - 2\eta Q x^t$
$$= V_-[0] - 2\eta \lambda(V_-[0]) \cdot V_-[0]$$
$$= V_-[0] \underbrace{(1 - 2\eta\lambda(V_-[0]))}_{>1 \ \forall \eta > 0}$$

- moreover, for $x \in \text{span}\{\mathbb{V}_-\}$, grad. update for any $\eta > 0$ causes a dilation of the iterates.

- for $x \in \text{span}\{\mathbb{V}_0\}$, the iterates are stationary as the eigenvalues associated are $\emptyset$. (2)

- lastly, for $x \in \text{span}\{\mathbb{V}_+\}$, the iterates shrink to $\emptyset$ for suff. small η, oscillate, or oscillate and diverge for suff. large η. (3)

Proof (2):

$$x^{t+1} = V_0[0](1 - 2\eta \cancelto{0}{\lambda(V_-[0])}) = x^t$$

Proof (3):

$$x^{t+1} = V_+[0] \underbrace{(1 - 2\eta\lambda(V_+[0]))}$$

cases: a) in $(-1, 1)$
b) equal to -1
c) in $(-\infty, -1)$

- putting all of this together, the process diverges for all $\boxed{x_0 \in \mathbb{R}^n \setminus \text{span}\{\mathbb{V}_+, \mathbb{V}_0\}}$ for any $\eta > 0$

$$\implies f(x^{t+1}) = (x^{t+1})^T Q x^{t+1}$$
$$= \underbrace{(x^{t+1}_{\mathbb{V}_+})^T Q (x^{t+1}_{\mathbb{V}_+})}_{\xrightarrow{t} 0 \text{ for small } \eta} + \underbrace{(x^{t+1}_{\mathbb{V}_0})^T Q (x^{t+1}_{\mathbb{V}_0})}_{= x_0^T Q x_0 \ \forall t} + \underbrace{(x^{t+1}_{\mathbb{V}_-})^T Q (x^{t+1}_{\mathbb{V}_-})}_{\xrightarrow{t} -\infty \ \forall \eta > 0 \ (4)}$$

Proof (4): consider $x^0_{\mathbb{V}_-} = V_-[0]$, $x^N_{\mathbb{V}_-} = \beta^N V_-[0]$, $\beta > 0$ $(\forall \eta > 0)$

$$\implies \text{last term} = \beta^N V_-[0]^T Q \beta^N V_-[0] = \underset{\substack{\Uparrow \\ \text{negative}}}{\lambda} \cdot \beta^{2N} \cancelto{1}{\|V_-[0]\|_2^2}$$