Name: Uzair Akbar ID: 03697290

Convex Optimization for ML & CV

Weekly Exercises 7 – Duality

Exercise 2

Let $\mathcal{L}(u) = \langle u, c \rangle + \varepsilon \cdot J(u)$

Now, from the definition of convex conjugate:

$$\Rightarrow \mathcal{L}^*(v) := \sup_u \langle u, v \rangle - \mathcal{L}(u)$$

$$\Rightarrow = \sup_u \langle u, v \rangle - \langle u, c \rangle - \varepsilon \cdot J(u)$$

$$\Rightarrow = \sup_u \langle u, v - c \rangle - \varepsilon \cdot J(u)$$

$$\Rightarrow = \varepsilon \cdot \sup_u \langle u, \frac{v-c}{\varepsilon} \rangle - J(u)$$

$$\therefore \mathcal{L}^*(v) = \varepsilon \cdot J^*\left(\frac{v-c}{\varepsilon}\right)$$

Exercise 3

First we derive the convex conjugates F^* and G^*, and subdifferentials

$$\because F^*(u,v) = \sup_{(\omega,x)} \left\langle \begin{bmatrix} \omega \\ x \end{bmatrix}, \begin{bmatrix} u \\ v \end{bmatrix} \right\rangle - F(\omega, x)$$

$$= \sup_{(\omega,x)} \begin{bmatrix} \omega^T & x^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} - F(\omega, x)$$

$$= \sup_{(\omega,x)} \omega^T u + x^T v - F(\omega, x)$$

$$= \sup_{(\omega,x)} \omega^T u + x^T v - \delta\{(\omega,x) = (\mu,\nu)\}$$

$$= \sup_{(\omega,x)=(\mu,\nu)} \omega^T u + x^T v$$

$$\therefore F^*(u,v) = \mu^T u + \nu^T v = \left\langle \begin{bmatrix} \mu \\ \nu \end{bmatrix}, \begin{bmatrix} u \\ v \end{bmatrix} \right\rangle$$

Therefore,

$$\partial F^*(u,v) = \{\nabla F^*(u,v)\} = [\mu^T \ \ \nu^T]$$

Similarly,

$$\because G^*(X) := \sup_Y \langle X, Y \rangle_F - G(Y)$$

$$= \sup_Y \sum_{j=1}^{n} X_j^T Y_j - G(Y)$$

$$= \sup_Y \left(\sum_{j=1}^{n} \sum_{i=1}^{m} X_{ij} Y_{ij} \right) - G(Y)$$

$$= \sup_Y \left\{ \left(\sum_{j=1}^{n} \sum_{i=1}^{m} X_{ij} Y_{ij} \right) - \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \left[C_{ij} Y_{ij} + \varepsilon Y_{ij} (\log Y_{ij} - 1) + \delta(Y_{ij} \geq 0) \right] \right) \right\}$$

$$= \sup_Y \sum_{i=1}^{m} \sum_{j=1}^{n} \left\{ X_{ij} Y_{ij} - C_{ij} Y_{ij} - \varepsilon Y_{ij} (\log Y_{ij} - 1) - \delta(Y_{ij} \geq 0) \right\}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \sup_{Y_{ij}} \left\{ X_{ij} Y_{ij} - C_{ij} Y_{ij} - \varepsilon Y_{ij} (\log Y_{ij} - 1) - \delta(Y_{ij} \geq 0) \right\}$$

$$\Rightarrow G^*(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} \sup_{Y_{ij} \geq 0} \left\{ X_{ij} Y_{ij} - C_{ij} Y_{ij} - \varepsilon Y_{ij} (\log Y_{ij} - 1) \right\}$$

Now we find the optimal value Y_{ij}^* element-wise to solve the supremum.

$$\Rightarrow \left. \frac{d\left(X_{ij} Y_{ij} - C_{ij} Y_{ij} - \varepsilon Y_{ij} (\log Y_{ij} - 1) \right)}{d Y_{ij}} \right|_{Y_{ij} = Y_{ij}^*} = 0$$

$$\Rightarrow X_{ij} - C_{ij} - \varepsilon (\log Y_{ij}^* - 1) - \varepsilon = 0$$

$$\Rightarrow Y_{ij}^* = \exp\left(\frac{X_{ij} - C_{ij}}{\varepsilon}\right) \geq 0$$

Plugging this back into $G^*(X)$, we get

$$\Rightarrow G^*(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} Y_{ij}^* - C_{ij} X_{ij}^* - \varepsilon Y_{ij}^* (\log Y_{ij}^* - 1)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \left(X_{ij} - C_{ij} - \varepsilon(\log Y_{ij}^* - 1)\right) Y_{ij}^*$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \left\{ \left(X_{ij} - C_{ij} - \varepsilon \left\{ (X_{ij} - C_{ij})/\varepsilon - 1 \right\}\right) \times \exp\left((X_{ij} - C_{ij})/\varepsilon\right) \right\}$$

$$\therefore G^*(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} \varepsilon \cdot \exp\left(\frac{X_{ij} - C_{ij}}{\varepsilon}\right)$$

Now, before calculating $\partial G(X)$, consider the following norm-cone:

$$N_{\mathbb{R}_+}(x) = \partial \delta_{\mathbb{R}_+}(x) = \left\{ p \in \mathbb{R} \;\middle|\; \begin{matrix} p = 0, & x > 0 \\ p \geq 0, & x \leq 0 \end{matrix} \right\}, \; x \in \mathbb{R}.$$

Now we proceed with computing $\partial G(x)$. We do this element-wis

$$\because G(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(C_{ij} X_{ij} + \varepsilon X_{ij} (\log X_{ij} - 1) + \delta_{\mathbb{R}_+}(X_{ij}) \right)$$

$$\Rightarrow \partial G(X) = \left\{ Y \in \mathbb{R}^{m \times n} \;\middle|\; Y_{ij} \in \partial\left(C_{ij} X_{ij} + \varepsilon X_{ij} (\log X_{ij} - 1) + \delta_{\mathbb{R}_+}(X_{ij}) \right) \right\}$$

$$\Rightarrow \partial G(X) = \left\{ Y \in \mathbb{R}^{m \times n} \,\middle|\, Y_{ij} \in \left(C_{ij} + \varepsilon \log X_{ij} + N_{\mathbb{R}_+}(X_{ij}) \right) \right\}$$

$$\therefore \partial G(X) = \left\{ Y \in \mathbb{R}^{m \times n} \,\middle|\, \begin{array}{ll} Y_{ij} = C_{ij} + \varepsilon \log X_{ij} & , X_{ij} > 0 \\ Y_{ij} \in \mathbb{R} & , X_{ij} \leq 0 \end{array} \right\}$$

We can now use the convex conjugates to state the dual formulation, and use the calculated subgradients to state the optimality conditions.

$$D^* := \max_{(p,q)} \left\{ -G^*\left(-K^T \begin{bmatrix} p \\ q \end{bmatrix}\right) - F^*(p,q) \right\}, (p,q) \in \mathbb{R}^m \times \mathbb{R}^n$$

From the Fenchel-Rockafellar duality, (U^*, p^*, q^*) is the optimal primal-dual tuple iff:

$$\begin{cases} K U^* \in \partial F^*(p^*, q^*) & \text{— (i)} \\ -K^T \begin{bmatrix} p^* \\ q^* \end{bmatrix} \in \partial G(U^*) & \text{— (ii)} \end{cases}$$

where $U^* \in \mathbb{R}^{m \times n}$, $(p^*, q^*) \in \mathbb{R}^m \times \mathbb{R}^n$ are the primal and dual veriables respectively.

(i) can be re-written as:

$$U^* \mathbb{1}_n = \mu \quad , \quad U^{*T} \mathbb{1}_m = \nu$$

(ii) can be re-written as:

$$p_i^* + q_j^* = -C_{ij} - \varepsilon \log U_{ij}^* \quad , \text{ if } U_{ij}^* > 0$$

Exercise 1

Part 1.

Let $X \in \mathbb{R}^{m \times n}$, $\|X\|_{2,1} \leq 1$. We have, for any $Y \in \mathbb{R}^{m \times n}$,

$$\begin{aligned}
\langle X, Y \rangle_F &= \sum_{i=1}^{n} \langle X_i, Y_i \rangle \\
&\leq \sum_{i=1}^{n} |X_i| \cdot |Y_i| \\
&\leq \sum_{i=1}^{n} |X_i| \cdot \max_{1 \leq j \leq n} |Y_j| \\
&= \|X\|_{2,1} \cdot \|Y\|_{2,\infty}
\end{aligned}$$

This implies that

$$\begin{aligned}
\langle X, Y \rangle_F - \|Y\|_{2,\infty} &\leq \|X\|_{2,1} - \|Y\|_{2,\infty} \\
&= (\|X\|_{2,1} - 1) \cdot \|Y\|_{2,\infty} \leq 0
\end{aligned}$$

Since $\langle X, 0 \rangle_F - \|0\|_{2,\infty} = 0$, we get

$$f_1^*(X) = \sup_{Y \in \mathbb{R}^{m \times n}} \langle X, Y \rangle_F - \|Y\|_{2,\infty} = 0$$

Now let $\|X\|_{2,1} > 1$. Define $Y \in \mathbb{R}^{m \times n}$ so that the i-th column Y_i of Y, $1 \leq i \leq n$ is given as $Y_i := \dfrac{X_i}{\|X_i\|_2}$, which implies $\|Y\|_{2,\infty} = 1$

We get

$$\langle X, Y\rangle_F = \sum_{i=1}^{n} \|X_i\|_2 = \|X\|_{2,1}$$

For $\alpha > 0$, we get

$$\langle X, \alpha Y\rangle_F - \|\alpha Y\|_{2,\infty} = \alpha\overbrace{(\|X\|_{2,1} - 1)}^{>1}$$

Therefore,

$$f_1^{*}(X) = \sup_{Y \in \mathbb{R}^{m\times n}} \langle X, Y\rangle_F - \|Y\|_{2,\infty} = \infty$$

Altogether,

$$f_1^{*}(X) = \delta_{\|\cdot\|_{2,1} \leq 1}(X)$$

Part 2.

We have $f_2 = f_1^{*}$ and since f_1 is closed, proper and convex, we have

$$f_2^{*} = f_1^{**} = f_1$$