EE 150 - Applications of Convex Optimization in Signal Processing and Communications
Dr. Andre Tkacenko, JPL
Third Term 2011-2012

Homework Set #3 - Solutions

1. (a) Note that $\mathbf{x}$ is closer to $\mathbf{x}_0$ than to $\mathbf{x}_\ell$ in the Euclidean norm sense if and only if we have the following.

$$\begin{aligned} ||\mathbf{x}-\mathbf{x}_0||_2 \leq ||\mathbf{x}-\mathbf{x}_\ell||_2 &\iff ||\mathbf{x}-\mathbf{x}_0||_2^2 \leq ||\mathbf{x}-\mathbf{x}_\ell||_2^2 \,, \\ &\iff (\mathbf{x}-\mathbf{x}_0)^T(\mathbf{x}-\mathbf{x}_0) \leq (\mathbf{x}-\mathbf{x}_\ell)^T(\mathbf{x}-\mathbf{x}_\ell) \,, \\ &\iff \mathbf{x}^T\mathbf{x} - 2\mathbf{x}_0^T\mathbf{x} + \mathbf{x}_0^T\mathbf{x}_0 \leq \mathbf{x}^T\mathbf{x} - 2\mathbf{x}_\ell^T\mathbf{x} + \mathbf{x}_\ell^T\mathbf{x}_\ell \,, \\ &\iff 2(\mathbf{x}_\ell-\mathbf{x}_0)^T\mathbf{x} \leq \mathbf{x}_\ell^T\mathbf{x}_\ell - \mathbf{x}_0^T\mathbf{x}_0 \,. \end{aligned}$$

Note that the last condition above defines a halfspace for every ℓ. Thus, we can express $\mathcal{V}$ as $\mathcal{V} = \{\mathbf{x} : \mathbf{A}\mathbf{x} \preceq \mathbf{b}\}$ with

$$\mathbf{A} = 2\begin{bmatrix} \mathbf{x}_1-\mathbf{x}_0 \\ \vdots \\ \mathbf{x}_K-\mathbf{x}_0 \end{bmatrix} = \begin{bmatrix} 2(\mathbf{x}_1-\mathbf{x}_0) \\ \vdots \\ 2(\mathbf{x}_K-\mathbf{x}_0) \end{bmatrix} , \ \mathbf{b} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{x}_1 - \mathbf{x}_0^T\mathbf{x}_0 \\ \vdots \\ \mathbf{x}_K^T\mathbf{x}_K - \mathbf{x}_0^T\mathbf{x}_0 \end{bmatrix} .$$

(b) Suppose we have $\mathcal{P} = \{\mathbf{x} : \mathbf{A}\mathbf{x} \preceq \mathbf{b}\}$ with $\mathbf{A} \in \mathbb{R}^{K\times n}$ and $\mathbf{b} \in \mathbb{R}^K$. Then, as $\mathcal{P}$ is assumed to have a nonempty interior, we can choose any point $\mathbf{x}_0 \in \{\mathbf{x} : \mathbf{A}\mathbf{x} \prec \mathbf{b}\}$, and then construct K points $\mathbf{x}_\ell$ by taking the mirror image of $\mathbf{x}_0$ with respect to the hyperplanes $\{\mathbf{x} : \mathbf{a}_\ell^T\mathbf{x} = b_\ell\}$. In other words, we choose $\mathbf{x}_\ell$ of the form $\mathbf{x}_\ell = \mathbf{x}_0 + \lambda\mathbf{a}_\ell$, where λ is chosen in such a way that the distance of $\mathbf{x}_\ell$ to the hyperplane defined by $\mathbf{a}_\ell^T\mathbf{x} = b_\ell$ is equal to the distance of $\mathbf{x}_0$ to the hyperplane. This leads to the following condition:

$$b_\ell - \mathbf{a}_\ell^T\mathbf{x}_0 = \mathbf{a}_\ell^T\mathbf{x}_\ell - b_\ell \,.$$

Substituting $\mathbf{x}_\ell = \mathbf{x}_0 + \lambda\mathbf{a}_\ell$ into the above condition and solving for λ yields

$$\lambda = \frac{2\left(b_\ell - \mathbf{a}_\ell^T\mathbf{x}_0\right)}{||\mathbf{a}_\ell||_2^2} \,.$$

Thus, if we choose

$$\mathbf{x}_\ell = \mathbf{x}_0 + \frac{2\left(b_\ell - \mathbf{a}_\ell^T\mathbf{x}_0\right)}{||\mathbf{a}_\ell||_2^2}\mathbf{a}_\ell \,, \ \ell = 1, \ldots, K \,,$$

then the polyhedron $\mathcal{P}$ is the Voronoi region of $\mathbf{x}_0$ with respect to $\mathbf{x}_1, \ldots, \mathbf{x}_K$.

(c) A polyhedral decomposition of $\mathbb{R}^n$ can not always be described as Voronoi regions generated by a set of points $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. A counterexample in $\mathbb{R}^2$ is shown below in Figure 1. In this figure, $\mathbb{R}^2$ is decomposed into 4 polyhedra $\mathcal{P}_1, \ldots, \mathcal{P}_4$ by 2 hyperplanes $\mathcal{H}_1, \mathcal{H}_2$. Suppose we arbitrarily pick $\mathbf{x}_1 \in \mathcal{P}_1$ and $\mathbf{x}_2 \in \mathcal{P}_2$. Then $\mathbf{x}_3 \in \mathcal{P}_3$ must be the mirror image of $\mathbf{x}_1$ and $\mathbf{x}_2$ with respect to $\mathcal{H}_2$ and $\mathcal{H}_1$, respectively. However, the mirror image of $\mathbf{x}_1$ with respect to $\mathcal{H}_2$ lies in $\widetilde{\mathcal{P}}_1$, and the mirror image of $\mathbf{x}_2$ with respect to $\mathcal{H}_1$ lies in $\widetilde{\mathcal{P}}_2$, so it is impossible to find such an $\mathbf{x}_3$.

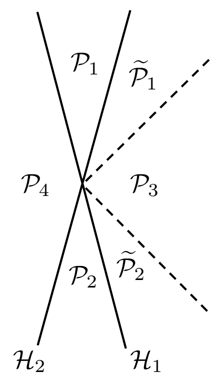

Figure 1: Polyhedral decomposition/Voronoi region partition counterexample in $\mathbb{R}^2$.

2. Recall that a set is convex if and only if its intersection with an arbitrary line of the form $\{\widehat{\mathbf{x}} + t\mathbf{v} : t \in \mathbb{R}\}$ is convex. We will use this property for both parts of this problem.

(a) Note that we have

$$(\widehat{\mathbf{x}} + t\mathbf{v})^T \mathbf{A} (\widehat{\mathbf{x}} + t\mathbf{v}) + \mathbf{b}^T (\widehat{\mathbf{x}} + t\mathbf{v}) + c = \alpha t^2 + \beta t + \gamma \,,$$

where we have

$$\alpha \triangleq \mathbf{v}^T\mathbf{A}\mathbf{v}\,,\ \beta \triangleq \mathbf{b}^T\mathbf{v} + 2\widehat{\mathbf{x}}^T\mathbf{A}\mathbf{v}\,,\ \gamma \triangleq c + \mathbf{b}^T\widehat{\mathbf{x}} + \widehat{\mathbf{x}}^T\mathbf{A}\widehat{\mathbf{x}}\,.$$

The intersection of $\mathcal{C}$ with the line defined by $\widehat{\mathbf{x}}$ and $\mathbf{v}$ is the set

$$\left\{\widehat{\mathbf{x}} + t\mathbf{v} : \alpha t^2 + \beta t + \gamma \le 0\right\}\,,$$

which is convex if $\alpha \ge 0$. This is true for any $\mathbf{v}$ if $\mathbf{v}^T\mathbf{A}\mathbf{v} \ge 0$ for all $\mathbf{v}$, i.e., $\mathbf{A} \succeq \mathbf{0}$. To show that the converse is false, consider the following counterexample. Suppose $\mathbf{A} = -1$, $\mathbf{b} = 0$, and $c = -1$. Then, $\mathbf{A} \not\succeq \mathbf{0}$, but we have

$$\mathcal{C} = \left\{x \in \mathbb{R} : -x^2 - 1 \le 0\right\} = \left\{x \in \mathbb{R} : x^2 + 1 \ge 0\right\} = \mathbb{R}\,,$$

which is clearly convex.

(b) Suppose that we define α, β, and γ as in the previous part of the solution. Note that we now also have

$$\mathbf{g}^T (\widehat{\mathbf{x}} + t\mathbf{v}) + h = \delta t + \epsilon\,,$$

where we have

$$\delta \triangleq \mathbf{g}^T\mathbf{v}\,,\ \epsilon \triangleq \mathbf{g}^T\widehat{\mathbf{x}} + h\,.$$

Without loss of generality, we can assume that $\widehat{\mathbf{x}} \in \mathcal{H}$, i.e., $\epsilon = 0$. The intersection of $\mathcal{C} \cap \mathcal{H}$ with the line defined by $\widehat{\mathbf{x}}$ and $\mathbf{v}$ is

$$\left\{\widehat{\mathbf{x}} + t\mathbf{v} : \alpha t^2 + \beta t + \gamma \le 0\,,\ \delta t = 0\right\}\,.$$

If $\delta = \mathbf{g}^T\mathbf{v} \neq 0$, then the intersection is the singleton $\{\widehat{\mathbf{x}}\}$ if $\gamma \le 0$, or it is empty otherwise. In either case, it is a convex set. If $\delta = \mathbf{g}^T\mathbf{v} = 0$, then the set reduces to

$$\left\{\widehat{\mathbf{x}} + t\mathbf{v} : \alpha t^2 + \beta t + \gamma \le 0\right\}\,,$$

which is convex if $\alpha \geq 0$. Therefore, $\mathcal{C} \cap \mathcal{H}$ is convex if

$$\mathbf{g}^T \mathbf{v} = 0 \Longrightarrow \mathbf{v}^T \mathbf{A} \mathbf{v} \geq 0 \,. \tag{1}$$

But this is true if there exists a λ such that $\left(\mathbf{A} + \lambda \mathbf{g}\mathbf{g}^T\right) \succeq \mathbf{0}$. In this case, (1) holds, because then we have

$$\mathbf{v}^T \mathbf{A} \mathbf{v} = \mathbf{v}^T \left(\mathbf{A} + \lambda \mathbf{g}\mathbf{g}^T\right) \mathbf{v} \geq 0 \,,$$

for all $\mathbf{v}$ satisfying $\mathbf{g}^T \mathbf{v} = 0$.

To show the converse is false, consider the following counterexample. Suppose we take

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \,, \; \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \,, \; c = -1 \,, \; \mathbf{g} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \,, \; h = 0 \,.$$

Then we clearly have

$$\mathbf{A} + \lambda \mathbf{g}\mathbf{g}^T = \begin{bmatrix} \lambda & 0 \\ 0 & -1 \end{bmatrix} \not\succeq \mathbf{0} \,,$$

for any choice of λ. In this case, though, we have

$$\mathcal{C} = \left\{(x_1, x_2) \in \mathbb{R}^2 : -x_2^2 - 1 \leq 0\right\} = \mathbb{R}^2 \,,$$

and so we have $\mathcal{C} \cap \mathcal{H} = \mathcal{H} = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_1 = 0\right\}$, which is clearly convex.

3. We first note that the constraints $p_k \geq 0$, $k = 1, \ldots, n$ define halfspaces, and $\sum_{k=1}^n p_k = 1$ defines a hyperplane, so $\mathcal{P}$ is a polyhedron, and hence, a convex set.

(a) The condition here is

$$\alpha \leq \sum_{k=1}^{n} p_k f(a_k) \leq \beta \,,$$

which is equivalent to two linear inequalities, and as such is convex in $\mathbf{p}$.

(b) The condition here is

$$\Pr\left\{X > \alpha\right\} = \sum_{k : a_k > \alpha} p_k \leq \beta \,,$$

which is equivalent to a linear inequality, and as such is convex in $\mathbf{p}$.

(c) The condition here is equivalent to

$$\sum_{k=1}^{n} p_k \left(|a_k|^3 - \alpha \,|a_k|\right) \leq 0 \,,$$

which is equivalent to a linear inequality, and as such is convex in $\mathbf{p}$.

(d) The condition here is

$$\sum_{k=1}^{n} p_k a_k^2 \leq \alpha \,,$$

which is equivalent to a linear inequality, and as such is convex in $\mathbf{p}$.

(e) The condition here is

$$\sum_{k=1}^{n} p_k a_k^2 \geq \alpha \,,$$

which is equivalent to a linear inequality, and as such is convex in $\mathbf{p}$.

(f) The condition here is

$$\mathrm{Var}(X) = E\big[X^2\big] - (E[X])^2 = \sum_{k=1}^{n} p_k a_k^2 - \left(\sum_{k=1}^{n} p_k a_k\right)^2 \leq \alpha\,,$$

which is not convex in general. As a counterexample, take $n = 2$, $a_1 = 0$, $a_2 = 1$, and $\alpha = \frac{1}{5}$. Then, $(p_1, p_2) = (1, 0)$ and $(p_1, p_2) = (0, 1)$ are two points which satisfy $\mathrm{Var}(X) = 0 \leq \frac{1}{5} = \alpha$, but the convex combination $(p_1, p_2) = \left(\frac{1}{2}, \frac{1}{2}\right)$ does not ($\mathrm{Var}(X) = \frac{1}{4} > \frac{1}{5} = \alpha$ here).

(g) The condition here is

$$\sum_{k=1}^{n} p_k a_k^2 - \left(\sum_{k=1}^{n} p_k a_k\right)^2 \geq \alpha \iff \sum_{k=1}^{n}\sum_{\ell=1}^{n} p_k a_k a_\ell p_\ell - \sum_{k=1}^{n} a_k^2 p_k + \alpha \leq 0\,.$$

By defining the following quantities:

$$\mathbf{a} \triangleq \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \ \mathbf{A} \triangleq \mathbf{a}\mathbf{a}^T\,, \ \mathbf{b} \triangleq \begin{bmatrix} a_1^2 \\ \vdots \\ a_n^2 \end{bmatrix},$$

it follows that the condition is equivalent to

$$\mathbf{p}^T\mathbf{A}\mathbf{p} - \mathbf{b}^T\mathbf{p} + \alpha \leq 0\,.$$

However, this defines a convex set, i.e., it is convex in $\mathbf{p}$, since $\mathbf{A} = \mathbf{a}\mathbf{a}^T \succeq \mathbf{0}$ (see the results of part (a) of the previous problem).

(h) For emphasize the dependence of the first quartile on $\mathbf{p}$, let us denote $Q_1(X) = f(\mathbf{p})$. From Figure 1 of the homework set, we have $f(\mathbf{p}) = a_2$. Using this figure, it can be seen that the condition $f(\mathbf{p}) \geq \alpha$ is equivalent to

$$F_X(\beta) < \frac{1}{4} \text{ for all } \beta < \alpha\,.$$

If $\alpha \leq a_1$, this is always true. Otherwise, define $k \triangleq \max\{\ell : a_\ell < \alpha\}$. This is a fixed integer, independent of $\mathbf{p}$. The constraint $f(\mathbf{p}) \geq \alpha$ holds if and only if

$$F_X(a_k) = \sum_{\ell=1}^{k} p_\ell < \frac{1}{4}\,.$$

This is a strict linear inequality in $\mathbf{p}$, which defines an open halfspace. As such, it is convex in $\mathbf{p}$.

(i) Using the notation defined in the previous part of the solution, it follows that the condition $f(\mathbf{p}) \leq \alpha$ is equivalent to

$$F_X(\beta) \geq \frac{1}{4} \text{ for all } \beta \geq \alpha\,.$$

This can be expressed as a linear inequality as follows:

$$\sum_{\ell=k+1}^{n} p_\ell \geq \frac{1}{4}\,.$$

Here, if $\alpha \leq a_1$, we define $k = 0$. As such, this condition is convex in $\mathbf{p}$.

4. (a) Following the hint, we have

$$\begin{aligned} g(t) &= \operatorname{tr}\Big((\mathbf{Z}+t\mathbf{V})^{-1}\Big) = \operatorname{tr}\left(\left(\mathbf{Z}^{\frac{1}{2}}\left(\mathbf{I}_n + t\mathbf{Z}^{-\frac{1}{2}}\mathbf{V}\mathbf{Z}^{-\frac{1}{2}}\right)\mathbf{Z}^{\frac{1}{2}}\right)^{-1}\right), \\ &= \operatorname{tr}\left(\mathbf{Z}^{-\frac{1}{2}}\left(\mathbf{I}_n + t\mathbf{Z}^{-\frac{1}{2}}\mathbf{V}\mathbf{Z}^{-\frac{1}{2}}\right)^{-1}\mathbf{Z}^{-\frac{1}{2}}\right) = \operatorname{tr}\left(\mathbf{Z}^{-1}\left(\mathbf{I}_n + t\mathbf{Z}^{-\frac{1}{2}}\mathbf{V}\mathbf{Z}^{-\frac{1}{2}}\right)^{-1}\right). \end{aligned}$$

For sake of simplicity, let us define $\mathbf{A} \triangleq \mathbf{Z}^{-\frac{1}{2}}\mathbf{V}\mathbf{Z}^{-\frac{1}{2}} \in \mathbb{S}^n$ and let $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ denote an eigenvalue decomposition of $\mathbf{A}$, where $\mathbf{Q}$ is an $n \times n$ unitary matrix of eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of eigenvalues of $\mathbf{A}$. Substituting this into the expression above yields

$$\begin{aligned} g(t) &= \operatorname{tr}\Big(\mathbf{Z}^{-1}(\mathbf{I}_n + t\mathbf{A})^{-1}\Big) = \operatorname{tr}\left(\mathbf{Z}^{-1}\left(\mathbf{I}_n + t\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\right)^{-1}\right), \\ &= \operatorname{tr}\Big(\mathbf{Z}^{-1}\left(\mathbf{Q}(\mathbf{I}_n + t\mathbf{\Lambda})\mathbf{Q}^T\right)^{-1}\Big) = \operatorname{tr}\Big(\mathbf{Z}^{-1}\mathbf{Q}(\mathbf{I}_n + t\mathbf{\Lambda})^{-1}\mathbf{Q}^T\Big), \\ &= \operatorname{tr}\Big(\left(\mathbf{Q}^T\mathbf{Z}\mathbf{Q}\right)(\mathbf{I}_n + t\mathbf{\Lambda})^{-1}\Big) = \sum_{k=1}^{n} \frac{\left[\mathbf{Q}^T\mathbf{Z}\mathbf{Q}\right]_{k,k}}{1 + t\lambda_k}. \end{aligned}$$

Now note that from the last equality that $\mathbf{g}(t)$ can be expressed as a positive weighted sum of convex functions $\frac{1}{1+t\lambda_k}$. Hence, $g(t)$ is convex.

(b) In this part, we will use the same eigenvalue decomposition of $\mathbf{A} \triangleq \mathbf{Z}^{-\frac{1}{2}}\mathbf{V}\mathbf{Z}^{-\frac{1}{2}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ as was done in the previous part. Here, we have

$$\begin{aligned} g(t) &= (\det(\mathbf{Z}+t\mathbf{V}))^{\frac{1}{n}} = \left(\det\left(\mathbf{Z}^{\frac{1}{2}}\left(\mathbf{I}_n + t\mathbf{Z}^{-\frac{1}{2}}\mathbf{V}\mathbf{Z}^{-\frac{1}{2}}\right)\mathbf{Z}^{\frac{1}{2}}\right)\right)^{\frac{1}{n}}, \\ &= \left(\det\left(\mathbf{Z}^{\frac{1}{2}}\right)\det\left(\mathbf{I}_n + t\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\right)\det\left(\mathbf{Z}^{\frac{1}{2}}\right)\right)^{\frac{1}{n}}, \\ &= \left((\det(\mathbf{Z}))^{\frac{1}{2}}\det\left(\mathbf{I}_n + t\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\right)(\det(\mathbf{Z}))^{\frac{1}{2}}\right)^{\frac{1}{n}}, \\ &= \left(\det(\mathbf{Z})\det\left(\mathbf{Q}(\mathbf{I}_n + t\mathbf{\Lambda})\mathbf{Q}^T\right)\right)^{\frac{1}{n}}, \\ &= (\det(\mathbf{Z}))^{\frac{1}{n}}\left(\det(\mathbf{Q})\det(\mathbf{I}_n + t\mathbf{\Lambda})\det\left(\mathbf{Q}^T\right)\right)^{\frac{1}{n}}, \\ &= (\det(\mathbf{Z}))^{\frac{1}{n}}(\det(\mathbf{I}_n + t\mathbf{\Lambda}))^{\frac{1}{n}}, \\ &= (\det(\mathbf{Z}))^{\frac{1}{n}}\left(\prod_{k=1}^{n}(1 + t\lambda_k)\right)^{\frac{1}{n}}. \end{aligned}$$

From the last equality, we have shown that $g(t)$ can be expressed as a product of a positive constant and the geometric mean of $(1 + t\lambda_k)$ for $k = 1, \ldots, n$, which is concave in t. Hence, the net result is that $g(t)$ is concave.

***5.** First note that we have

$$P_{\mathbf{p}}(\mathcal{C}) = 1 - P_{\mathbf{p}}\left(\widetilde{\mathcal{C}}\right),$$

where $\widetilde{\mathcal{C}} \triangleq \{1, \ldots, n\} \setminus \mathcal{C}$ is the *complement* of $\mathcal{C}$ in the set $\{1, \ldots, n\}$. Hence we have

$$P_{\mathbf{p}}(\mathcal{C}) - P_{\mathbf{q}}(\mathcal{C}) = -P_{\mathbf{p}}\left(\widetilde{\mathcal{C}}\right) + P_{\mathbf{q}}\left(\widetilde{\mathcal{C}}\right) = -\left(P_{\mathbf{p}}\left(\widetilde{\mathcal{C}}\right) - P_{\mathbf{q}}\left(\widetilde{\mathcal{C}}\right)\right),$$

and so $d_{\mathrm{mp}}(\mathbf{p},\mathbf{q})$ can be expressed equivalently as

$$d_{\mathrm{mp}}(\mathbf{p},\mathbf{q}) = \max\left\{P_{\mathbf{p}}(\mathcal{C}) - P_{\mathbf{q}}(\mathcal{C}) : \mathcal{C} \subseteq \{1,\ldots,n\}\right\} .$$

As $d_{\mathrm{mp}}(\mathbf{p},\mathbf{q})$ is the maximum of 2^n linear functions of $(\mathbf{p},\mathbf{q})$ (since the number of events in the set $\mathcal{S} \triangleq \{1,\ldots,n\}$ is the cardinality of the *power set* $\mathcal{P}(\mathcal{S})$, which is 2^n), it follows that $d_{\mathrm{mp}}(\mathbf{p},\mathbf{q})$ is convex.

To simplify the expression for $d_{\mathrm{mp}}(\mathbf{p},\mathbf{q})$, let us identify a subset $\mathcal{C}$ that maximizes

$$P_{\mathbf{p}}(\mathcal{C}) - P_{\mathbf{q}}(\mathcal{C}) = \sum_{k\in\mathcal{C}} (p_k - q_k) . \tag{2}$$

We claim that such an optimal subset is given by

$$\mathcal{C}^\star \triangleq \{k \in \{1,\ldots,n\} : p_k > q_k\} . \tag{3}$$

To show this, first note that the indices for which $p_k = q_k$ clearly do not matter. Thus, we will ignore these indices and assume without loss of generality that for each index, either $p_k > q_k$ or $p_k < q_k$. In other words, we need only compute (2) over relevant indices for which either $p_k > q_k$ or $p_k < q_k$. Now consider any other subset $\mathcal{C}$. If there is an element ℓ in $\mathcal{C}^\star$ but not $\mathcal{C}$, then by adding ℓ to $\mathcal{C}$, we increase $(P_{\mathbf{p}}(\mathcal{C}) - P_{\mathbf{q}}(\mathcal{C}))$ by $(p_\ell - q_\ell) > 0$, and so $\mathcal{C}$ could not have been optimal. Conversely, suppose that $\ell \in \mathcal{C} \setminus \mathcal{C}^\star$, so $(p_\ell - q_\ell) < 0$. If we remove ℓ from $\mathcal{C}$, we would increase $(P_{\mathbf{p}}(\mathcal{C}) - P_{\mathbf{q}}(\mathcal{C}))$ by $(q_\ell - p_\ell) > 0$, and so $\mathcal{C}$ could not have been optimal. This proves that $\mathcal{C}^\star$ as defined in (3) is optimal for maximizing the expression in (2) and so we have

$$d_{\mathrm{mp}}(\mathbf{p},\mathbf{q}) = \sum_{k:p_k>q_k} (p_k - q_k) . \tag{4}$$

To express $d_{\mathrm{mp}}(\mathbf{p},\mathbf{q})$ in terms of $||\mathbf{p}-\mathbf{q}||_1$, note that we have

$$\sum_{k:p_k>q_k} (p_k - q_k) + \sum_{k:p_k\le q_k} (p_k - q_k) = \sum_{k=1}^{n} (p_k - q_k) = \sum_{k=1}^{n} p_k - \sum_{k=1}^{n} q_k = \mathbf{1}^T\mathbf{p} - \mathbf{1}^T\mathbf{q} = 1 - 1 = 0 ,$$

and so

$$\sum_{k:p_k>q_k} (p_k - q_k) = -\sum_{k:p_k\le q_k} (p_k - q_k) .$$

Substituting this into (4) yields the following:

$$\begin{aligned} d_{\mathrm{mp}}(\mathbf{p},\mathbf{q}) &= \frac{1}{2}\sum_{k:p_k>q_k} (p_k - q_k) + \frac{1}{2}\sum_{k:p_k>q_k} (p_k - q_k) , \\ &= \frac{1}{2}\sum_{k:p_k>q_k} (p_k - q_k) - \frac{1}{2}\sum_{k:p_k\le q_k} (p_k - q_k) , \\ &= \frac{1}{2}\sum_{k=1}^{n} |p_k - q_k| . \end{aligned}$$

Hence, we conclude

$$d_{\mathrm{mp}}(\mathbf{p},\mathbf{q}) = \frac{1}{2}\,||\mathbf{p}-\mathbf{q}||_1 .$$

From this, it is very clear that $d_{\mathrm{mp}}(\mathbf{p},\mathbf{q})$ is convex.

The best way to interpret this result is as an interpretation of the ℓ_1-norm for probability distributions. It states that the ℓ_1-distance between two probability distributions is twice the maximum difference in probability, over all events, of the distributions.