Convex Optimization

Review Session 2

Question 1 (A pot-pourri of short n' sweet questions)

1. [1] Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex, $f : \mathbb{R}^n \to \mathbb{R}$ is concave, $\text{dom}f = \text{dom}g = \mathbb{R}^n$, and for all x, $g(x) \leq f(x)$. Show that there exists an affine function h such that for all x, $g(x) \leq h(x) \leq f(x)$. In other words, if a concave function g is an underestimator of a convex function f, then we can fit an affine function between f and g.

2. [2] Recall that the Kullback-Leibler divergence between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n_{++}$ is given by

$$D_{\text{kl}}(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} \{u_i \log(u_i/v_i) - u_i + v_i\}$$

   Show that it is convex, and so that

$$D_{\text{kl}}(\mathbf{u}, \mathbf{v}) \geq 0 \text{ for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n_{++}$$

   with equality if and only if $\mathbf{u} = \mathbf{v}$.

3. [3] In general, the product or ratio of two convex functions is not convex. However, show that

   - If f and g are both convex, positive and increasing or decreasing, then fg is convex.
   - If f and g are both concave and positive, and one function is increasing while the other is decreasing, then fg is concave.
   - If f is convex, increasing and positive and g is concave, decreasing and positive, then f/g is convex.

Solution

1. First note that $\text{intepi}(f)$ and[1] $\text{hypo}(g)$ are non-empty (because the domain of each function is $\mathbb{R}^n$) and do not intersect (since $g(x) \leq f(x)$). As such, these sets can be separated by a hyperplane. That hyperplane is precisely the affine function that lies between f and g.

2. Checking convexity over $\mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ is elementary[2]. To

[1]Problem 3.12 in Boyd
[2]Problem 3.13 in Boyd
[3]Problem 3.32 in Boyd

[1]Recall that the *hypograph* of g, denoted $\text{hypo}(g)$ is the set that lies *below* a given function.

[2]One way of seeing this is calculating the Hessian of $f(x, y) = x \log(x/y)$. You'll find it is

$$\nabla^2 f(x, y) = \begin{pmatrix} 1/x & -1/y \\ -1/y & x/y^2 \end{pmatrix}$$

with resulting eigenvalues

$$\lambda = \left(0, \frac{1}{x} + \frac{x}{y^2}\right)$$

both of which are clearly positive over $\mathbb{R}^n_{++}$, making $f(x, y)$ convex. Or, you could just note that $f(x, y)$ is the perspective function of $f(x) = -\log(x)$ which is itself convex. Either way, once the convexity of $f(x, y)$ is established, the convexity of the KL-divergence follows since it's then simply a sum of convex functions.

prove our inequality, consider that $f(\mathbf{v}) = \sum_{i=1}^{n} v_i \log v_i$ is strictly convex for strictly positive $\mathbf{v}$[3]Thus, by the first order conditions for convexity,

[3]To show this, consider that if $f(x) = x \log x$, then $f''(x) = 1/x$. This is clearly strictly positive for strictly positive x.

$$f(\mathbf{u}) > f(\mathbf{v}) + \nabla f(\mathbf{v})^\top (\mathbf{u} - \mathbf{v})$$

for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}_{++}^n$ with $\mathbf{u} \neq \mathbf{v}$. Feeding in $f(\mathbf{v})$ into each side of the inequality, we obtain

$$\begin{aligned}\sum_{i=1}^{n} u_i \log u_i &> \sum_{i=1}^{n} v_i \log v_i + \sum_{i=1}^{n} (\log v_i + 1)(u_i - v_i) \\ &= \sum_{i=1}^{n} (u_i \log v_i + u_i - v_i)\end{aligned}$$

Re-arranging gives the required result.

3. To prove the first statement, consider that f and g are both positive and convex. Thus, for any $\mathbf{x}$ and $\mathbf{y}$, and any $\theta \in [0, 1]$ and $\mathbf{z} = \theta\mathbf{x} + (1-\theta)\mathbf{y}$, we have

$$\begin{aligned}f(\mathbf{z})g(\mathbf{z}) &\leq (\theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}))\,(\theta g(\mathbf{x}) + (1-\theta) g(\mathbf{y})) \\ &= \theta f(\mathbf{x}) g(\mathbf{x}) + (1-\theta) f(\mathbf{y}) g(\mathbf{y}) + \theta(1-\theta)\,(f(\mathbf{y} - f(\mathbf{x}))\,(g(\mathbf{x}) - g(\mathbf{y}))\end{aligned}$$

The last term is negative if f and g are either both increasing or both decreasing. Since this is the case, we find that fg is indeed convex.

For second statement, simply reverse the inequalities above. For the third, note that $1/g$ is convex, positive and increasing, so that the result follows from the first statement.

■ □ ■

Question 2 (**Support and Indicator Functions)

Consider a set $\mathscr{S}$. We define the following two functions

- The *indicator function*, $I_{\mathscr{S}} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined as

$$I_{\mathscr{S}}(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in \mathscr{S} \\ \infty & \mathbf{x} \notin \mathscr{S} \end{cases}$$

- The *support function*, $\sigma_{\mathscr{S}} : \mathbb{R}^n \cup \{\infty\}$ is defined as

$$\sigma_{\mathscr{S}}(\mathbf{a}) = \sup_{\mathbf{x} \in \mathscr{S}} \{\mathbf{a} \cdot \mathbf{x}\} = \sup_{\mathbf{x} \in \mathbb{R}^n} \{\mathbf{a} \cdot \mathbf{x} - I_{\mathscr{S}}(\mathbf{x})\}$$

Let $\bar{\mathscr{S}}_C$ be the smallest closed convex set containing $\mathscr{S}$, and for the purposes of this question, assume $\mathscr{S} \subseteq \mathbb{R}^n$. Show that

$$\bar{\mathscr{S}}_C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} \le \sigma_{\mathscr{S}}(\mathbf{a}) \text{ for all } \mathbf{a} \in \mathbb{R}^n\} \tag{1}$$

and that

$$I_{\bar{\mathscr{S}}_C}(\mathbf{x}) = \sup_{\mathbf{a} \in \mathbb{R}^n} \{\mathbf{a} \cdot \mathbf{x} - \sigma_{\mathscr{S}}(\mathbf{a})\}$$

Solution

Before we begin, let's take a step back to review session 1, when we looked at dual cones and polar sets. We saw that both these constructs basically contained all the hyperplanes that separated a given set from every point outside the set. However, because of the particular structure of these dual sets (in particular, we were constraining ourselves to hyperplanes $\mathbf{a} \cdot \mathbf{x} \le b$ with *either* 0 RHS or strictly non-zero RHS), we were constrained to describing either cones or sets containing the origin in their interior.

This hardly satisfactory – surely there must be a way to describe *all* sets based on their supporting hyperplanes. How might we do this? We saw that the reason dual cones and polar sets couldn't really do the job was because of our constraints on the RHS of the hyperplanes. A pretty good place to begin, therefore, might be to relax that structure and allow hyperplanes with *any* right-hand-side. This is precisely what we'll be doing here.

In particular, consider the definition of the support function. All it does is take a vector $\mathbf{a}$ (that defines a hyperplane) and then tries to find the RHS that will fit that hyperplane as snugly as possible against the set (by maximizing $\mathbf{a} \cdot \mathbf{x}$ for all points in the set). Thus, for each $\mathbf{a}$, the support function tells us that the smallest halfspace with gradient $\mathbf{a}$ that contains the set is

$$\mathbf{a} \cdot \mathbf{x} \le \sigma_{\mathscr{S}}(\mathbf{a})$$

Little wonder, therefore, that if we know σ, we can reconstruct the closest closed convex set that contains $\mathscr{S}$ – σ basically tells us every halfspace that contains $\mathscr{S}$; so of course, their intersection (given by equation 1) gives $\bar{\mathscr{S}}_C$. We have found a way to encode every hyperplane that supports *any* convex set!

Now that we've done the hard work of understanding what's going on, let's prove everything rigorously!

To begin with, we'll show that

$$\sigma_{\mathscr{S}}(\mathbf{a}) = \sigma_{\bar{\mathscr{S}}_C}(\mathbf{a})$$

(Recall that $\bar{\mathscr{S}}_C$ is the smallest closed convex set containing $\mathscr{S}$).

- $\boxed{\sigma_{\mathscr{S}}(\mathbf{a}) \le \sigma_{\bar{\mathscr{S}}_C}(\mathbf{a})}$ This side is easier. Recall that the definition of the support function involves a supremum over all $\mathbf{x}$ in the set in question. Since we know that $\mathscr{S} \subseteq \bar{\mathscr{S}}_C$, it is clear that when we take $\sup_{\mathbf{x}\in\bar{\mathscr{S}}_C}$, we're including every $\mathbf{x}$ that we include when we take $\sup_{\mathbf{x}\in\mathscr{S}}$, and more. Thus, the former statement must be larger.

- $\boxed{\sigma_{\mathscr{S}}(\mathbf{a}) \ge \sigma_{\bar{\mathscr{S}}_C}(\mathbf{a})}$ Imagine that $\sigma_{\bar{\mathscr{S}}_C}(\mathbf{a}) = k < \infty$. [4] This means that there is a $\bar{\mathbf{x}}_C \in \bar{\mathscr{S}}_C$ such that

$$\mathbf{a} \cdot \bar{\mathbf{x}}_C = k$$

This means that for any small $\delta > 0$, there exists a $\mathbf{x}_C \in \mathscr{S}_C$ (notice we're now considering a set that might be open) with [5]

$$\mathbf{a} \cdot \mathbf{x}_C \ge k - \delta \qquad (2)$$

However, since $\mathbf{x}_C \in \mathscr{S}_C$, it must be a finite convex combination of points in $\mathscr{S}$. Thus, we can write

$$\mathbf{x}_C = \sum_{i=1}^{n} \lambda_i \mathbf{x}^i \qquad \lambda_i \in [0,1], \sum \lambda_i = 1, \mathbf{x}^i \in \mathscr{S} \qquad (3)$$

And so feeding 3 into 2, we find that for any small $\delta > 0$,

$$\sum_{i=1}^{n} \lambda_i (\mathbf{a} \cdot \mathbf{x}^i) \ge k - \delta$$

Now, clearly,

$$k - \delta \le \sum_{i=1}^{n} \lambda_i (\mathbf{a} \cdot \mathbf{x}^i) \le \left(\sum_{i=1}^{n} \lambda_i \right) \max_i (\mathbf{a} \cdot \mathbf{x}^i) = \max_i (\mathbf{a} \cdot \mathbf{x}^i)$$

And so there must be at least one $\mathbf{x}^i \in \mathscr{S}$ such that

$$\mathbf{a} \cdot \mathbf{x}^i \ge k - \delta$$

Thus, letting $\delta \to 0$, we find that

$$\sigma_{\mathscr{S}}(\mathbf{a}) \ge k = \sigma_{\bar{\mathscr{S}}_C}(\mathbf{a})$$

[4] The proof for the case in which the support function is infinite is similar, and in fact slightly easier.

[5] For the pedants out there, here's a proof – consider that since $\bar{\mathbf{x}}_C \in \bar{\mathscr{S}}_C$, then for any $\delta > 0$, there exists a ball $\mathscr{B} = B_{\delta/\|\mathbf{a}\|}(\bar{\mathbf{x}}_C)$ such that $\mathscr{B} \cap \mathscr{S}_C \neq \varphi$. Take some $\mathbf{x}_C \in \mathscr{B} \cap \mathscr{S}_C$. Consider, then, that if we let $\mathbf{d} = \bar{\mathbf{x}}_C - \mathbf{x}_C$, we have $\|\mathbf{d}\| \le \delta/\|\mathbf{a}\|$. As such,

$$\begin{aligned} \mathbf{a} \cdot \mathbf{x}_C &= \mathbf{a} \cdot \mathbf{x} - \mathbf{a} \cdot \mathbf{d} \\ &= k - \mathbf{a} \cdot \mathbf{d} \\ &\ge k - \|\mathbf{a}\| \frac{\delta}{\|\mathbf{a}\|} \\ &= k - \delta \end{aligned}$$

Having established that $\sigma_{\mathscr{S}}(\mathbf{a}) = \sigma_{\bar{\mathscr{S}}_C}(\mathbf{a})$, we can restrict ourselves to closed convex sets $\mathscr{C}$. We need to show that

$$\mathscr{C} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} \le \sigma_{\mathscr{C}}(\mathbf{a}) \text{ for all } \mathbf{a} \in \mathbb{R}^n\}$$

We can re-write this as

$$\mathscr{C} = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} \le \sup_{\mathbf{x}\in\mathscr{C}} \{\mathbf{a} \cdot \mathbf{x}\} \text{ for all } \mathbf{a} \in \mathbb{R}^n \right\}$$

In what follows, we will be referring to the set on the right-hand-side of this equation as "RHS". So we need to prove that $\mathscr{C} = \text{RHS}$. We do this in two steps

- $\boxed{\mathscr{C} \subseteq \text{RHS}}$ This is somewhat obvious. If $\mathbf{c} \in \mathscr{C}$, then it's obvious that $\mathbf{a} \cdot \mathbf{c} \leq \sup_{\mathbf{x} \in \mathscr{C}} \{\mathbf{a} \cdot \mathbf{x}\}$ for all $\mathbf{a}$, because $\mathbf{x} = \mathbf{c}$ is a valid assignment in the optimization problem. Thus, if we take the supremum over *all* possible $\mathbf{x}$, we must get something bigger.

- $\boxed{\text{RHS} \subseteq \mathscr{C}}$ Slightly harder – let's prove it the standard way. Suppose $\mathbf{c} \notin \mathscr{C}$. We need to show that $\mathbf{c} \notin \text{RHS}$. To do this, note that since $\mathbf{c} \notin \mathscr{C}$, and $\mathscr{C}$ is assumed to be convex and closed, there is a strictly separating hyperplane $\mathbf{a}$ such that

  $$\mathbf{a} \cdot \mathbf{c} > \mathbf{a} \cdot \mathbf{x} \text{ for all } \mathbf{x} \in \mathscr{C}$$

  Taking the supremum over $\mathbf{x}$ on the RHS,

  $$\mathbf{a} \cdot \mathbf{c} > \sup_{\mathbf{x} \in \mathscr{C}} \{\mathbf{a} \cdot \mathbf{x}\}$$

  We have therefore found at least one $\mathbf{a}$ for which $\mathbf{a} \cdot \mathbf{c} > \sup_{\mathbf{x} \in \mathscr{C}} \{\mathbf{a} \cdot \mathbf{x}\}$. Thus, by definition, we must have $\mathbf{c} \notin \text{RHS}$.

Our last step is to prove that

$$I_{\bar{\mathscr{I}}_C}(\mathbf{x}) = \sup_{\mathbf{a} \in \mathbb{R}^n} \{\mathbf{a} \cdot \mathbf{x} - \sigma_{\mathscr{S}}(\mathbf{a})\}$$

We showed above that $\mathscr{C}$ was precisely that set of points whose dot product with $\mathbf{a}$ is $\leq \sigma_{\mathscr{C}}(\mathbf{a})$ for all $\mathbf{a}$. Thus,

- If $\mathbf{x} \in \mathscr{C}$, then $\mathbf{a} \cdot \mathbf{x} \leq \sigma_{\mathscr{C}}(\mathbf{a})$ for all $\mathbf{a}$, and so the maximum the supremum on the right-hand side could reach is 0 (when $\mathbf{a} \cdot \mathbf{x} = \sigma_{\mathscr{C}}(\mathbf{a})$). And we know that it *will* reach 0, because it does it at $\mathbf{a} = \mathbf{0}$.

- If $\mathbf{x} \notin \mathscr{C}$, there *must* be some $\mathbf{a}$ for which $\mathbf{a} \cdot \mathbf{x} > \sigma_{\mathscr{S}}(\mathbf{a}) = \sup_{\mathbf{x} \in \mathscr{C}} \{\mathbf{a} \cdot \mathbf{x}\}$. By multiplying $\mathbf{a}$ by some arbitrarily large constant λ, we can make this difference as large as we want, and therefore make $\sup_{\mathbf{a} \in \mathbb{R}^n} \{\mathbf{a} \cdot \mathbf{x} - \sigma_{\mathscr{S}}(\mathbf{a})\}$ infinite.

Thus, we have shown that

$$\sup_{\mathbf{a} \in \mathbb{R}^n} \{\mathbf{a} \cdot \mathbf{x} - \sigma_{\mathscr{S}}(\mathbf{a})\} = \begin{cases} 0 & \mathbf{x} \in \mathscr{S} \\ \infty & \mathbf{x} \notin \mathscr{S} \end{cases}$$

It is therefore equal to the indicator function of $\mathscr{C}$.

■ □ ■

Question 3 (***Conjugate functions) ____

Recall that the *conjugate function* of a proper function f is defined by

$$f^*(\mathbf{x}) = \sup_{\mathbf{y}} \{\mathbf{x} \cdot \mathbf{y} - f(\mathbf{y})\}$$

Answer the following questions

- Show that f^* is convex, whether f is convex or not.
- Show that f^{**} is the largest closed convex function that lies below f.
- Prove Frenchel's Inequality

$$\mathbf{a} \cdot \mathbf{x} \leq f(\mathbf{x}) + f^*(\mathbf{a})$$

and use it to prove Young's Inequality, which states that for any non-negative real numbers a and b, and any $p, q > 0$, with $p^{-1} + q^{-1} = 1$

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Solution

Let's first spend some time understanding what conjugate functions are. All it appears you're doing when you work out $f(\mathbf{a})$ is taking a line with gradient $\mathbf{a}$ that passes through the origin, and trying to find the point at which that line and the curve $f(\mathbf{x})$ are as close to each other as possible. Simple logic[6] should quickly convince you that this occurs at the point $\mathbf{y}$ for which $\nabla f(\mathbf{y}) = \mathbf{a}$. Another way of looking at the quantity $f^*(\mathbf{a})$, therefore, is that we generate it by taking a hyperplane with gradient $\mathbf{a}$, and push it as snugly as possible to the curve. The intercept of the hyperplane is then $-f^*(\mathbf{a})$. This is illustrated in figure 1.

[6] Or, if you don't like logic, simply differentiate $\mathbf{a} \cdot \mathbf{y} - f(\mathbf{y})$ with respect to $\mathbf{y}$ and set it to 0:

$$\mathbf{a} - \nabla f(\mathbf{y}) = 0$$

So we see that the supremum in the convex conjugate is indeed obtained at the point $\mathbf{y}$ at which $\nabla f(\mathbf{y}) = \mathbf{a}$

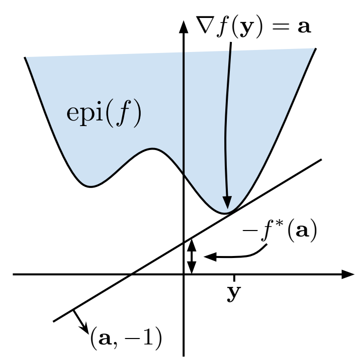

Figure 1: The convex conjugate $f^*(\mathbf{a})$ takes a hyperplane with gradient $\mathbf{a}$ and tries to fit it as snugly as possible below the function f. Another way to look at this is that the convex conjugate is none other than the support function of the epigraph of f at the point $(\mathbf{a}, -1)^\top$. Note for future reference that the equation of the line illustrated in this diagram (ie: the tangent hyperplane with gradient $\mathbf{a}$) is precisely $y = \mathbf{a} \cdot \mathbf{x} + (-f^*(\mathbf{x}))$.

Now consider that concept – taking a hyperplane and pushing it up again a set. That's something we've seen before when we studied support functions. Is there a relation between convex conjugates and support functions? Indeed there is – the convex conjugate of f is in fact precisely the support function of the epigraph of f, as follows

$$\begin{aligned} f^*(\mathbf{a}) &= \sup_{\mathbf{y}} \{\mathbf{a} \cdot \mathbf{y} - f(\mathbf{y})\} \\ &= \sup_{\mathbf{y}} \left\{ \begin{bmatrix} \mathbf{a} \\ -1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{y} \\ f(\mathbf{y}) \end{bmatrix} \right\} \\ &= \sup_{\mathbf{z} \in \text{epi}(f)} \left\{ \begin{bmatrix} \mathbf{a} \\ -1 \end{bmatrix} \cdot \mathbf{z} \right\} \\ &= \sigma_{\text{epi}(f)} \left(\begin{bmatrix} \mathbf{a} \\ -1 \end{bmatrix} \right) \end{aligned}$$

where for any function $f : \mathbb{R}^n \to \mathbb{R}$, we define

$$\text{epi}(f) = \{(\mathbf{x}, t) : \mathbf{x} \in \mathbb{R}^n, f(\mathbf{x} \leq t\} \subseteq \mathbb{R}^{n+1}$$

This is also illustrated in figure 1.

Let's now consider the statement that f^{**} is the largest closed convex function smaller than f. Why does that make sense intuitively? First, consider that

$$f^{**}(\mathbf{x}) = \sup_{\mathbf{a}} \{\mathbf{a} \cdot \mathbf{x} - f^*(\mathbf{a})\}$$

But consider figure 1 again. We mentioned in the caption there that $\mathbf{a} \cdot \mathbf{x} - f^*(\mathbf{a})$ is precisely the height at $\mathbf{x}$ of the tangent hyperplane to f with gradient $\mathbf{a}$. Thus, we can write this as

$$f^{**}(\mathbf{x}) = \sup_{\text{Supporting hyperplanes of epi}(f)} \{\text{Height of hyperplane at } \mathbf{x}\}$$

As illustrated in figure 2, this is clearly equal to the value at $\mathbf{x}$ of the largest closed convex function that lies below f.

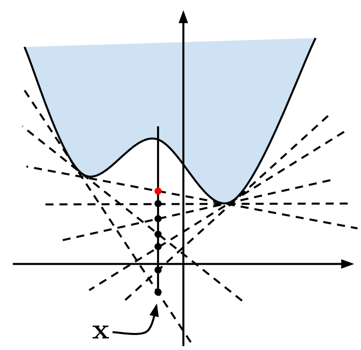

Figure 2: This diagram illustrates a number of hyperplanes that support epi(f). The heights of those hyperplanes at the point $\mathbf{x}$ are indicated by blobs. Clearly, the highest such blob over all hyperplanes (in red in the diagram) is none other than the value of the largest closed convex function that lies below f.

Finally, consider Frenchel's Inequality. In light of everything we've seen, it's almost trivial. Re-arranging it, we obtain

$$\mathbf{a} \cdot \mathbf{x} - f^*(\mathbf{a}) \leq f(\mathbf{x})$$

We have already see that the LHS is the hyperplane with gradient $\mathbf{p}$ supporting the epigraph of f. Since it's a supporting hyperplane, it's obvious it lies below f everywhere, hence the inequality.

OK – now it's time to prove everything rigorously! First, the fact f^* is convex. Consider two vectors $\mathbf{p}$ and $\mathbf{q}$, and a $\lambda \in [0, 1]$. Then construct $\mathbf{r} = \lambda\mathbf{p} + (1-\lambda)\mathbf{q}$. We find that

$$\begin{aligned}
f^*(\mathbf{r}) &= f^*(\lambda\mathbf{p} + (1-\lambda)\mathbf{q}) \\
&= \sup_{\mathbf{y}} \{[\lambda\mathbf{p} + (1-\lambda)\mathbf{q}] \cdot \mathbf{y} - f(\mathbf{y})\} \\
&= \sup_{\mathbf{y}} \{\lambda(\mathbf{p} \cdot \mathbf{y}) + (1-\lambda)(\mathbf{q} \cdot \mathbf{y}) - f(\mathbf{y})\} \\
&= \sup_{\mathbf{y}} \{\lambda[\mathbf{p} \cdot \mathbf{y} - f(\mathbf{y})] + (1-\lambda)[\mathbf{q} \cdot \mathbf{y} - f(\mathbf{y})]\} \\
&\stackrel{7}{\leq} \lambda \sup_{\mathbf{y}} \{\mathbf{p} \cdot \mathbf{y} - f(\mathbf{y})\} + (1-\lambda) \sup_{\mathbf{y}} \{\mathbf{q} \cdot \mathbf{y} - f(\mathbf{y})\} \\
&= \lambda f^*(\mathbf{p}) + (1-\lambda) f^*(\mathbf{q})
\end{aligned}$$

[7]This step follows because in the previous line, we are maximizing the entire expression ove ra *single* $\mathbf{y}$. In other words, we pick *one* $\mathbf{y}$ that must maximize the whole thing. In this line, we are splitting the spremum, and we allow each individual expression to be maximized by its own, individual $\mathbf{y}$. Clearly, I could *choose* to make these two $\mathbf{y}$'s equal, but I also have more freedom to do something else. Thus, the result must be at least larger.

We made no assumption on the convexity of f in the above. Thus, f^* is indeed convex regardless of whether f is.

Now, let $\tilde{f}(\mathbf{x})$ be the largest closed convex function that lies below f. Let's prove that $f^{**} = \tilde{f}$. We do this in two parts

- $\boxed{f^{**}(\mathbf{x}) \leq \tilde{f}(\mathbf{x})}$ Consider that

$$f^{**}(\mathbf{x}) = \sup_{\mathbf{z}} \left\{\mathbf{z} \cdot \mathbf{x} - \sup_{\mathbf{y}} (\mathbf{z} \cdot \mathbf{y} - f(\mathbf{y}))\right\} \qquad (4)$$

Let's consider two cases:

- If $\mathbf{x}$ is such that $f(\mathbf{x}) = \tilde{f}(\mathbf{x})$, then simply set $\mathbf{y} = \mathbf{x}$ in the infemum. The RHS then becomes $f(\mathbf{x})$. Clearly, the *actual* result will be smaller, because we are finding the inf over *all* $\mathbf{y}$. Thus, $f^{**}(\mathbf{x}) \leq \tilde{f}(\mathbf{x})$.
- If $\mathbf{x}$ is such that $f(\mathbf{x}) > \tilde{f}(\mathbf{x})$, then this means that $f(\mathbf{x})$ overestimates the convex function $\tilde{f}$ at that point.

  Let's now define the set $\mathscr{O}$, which contains all $(\mathbf{a}, \mathbf{b}, \lambda)$ such that

  * $\mathbf{x}$ lies on the line between $\mathbf{a}$ and $\mathbf{b}$, and the vector λ describes how far along that line $\mathbf{x}$ lies. In other words,
  $$\mathbf{x} = \lambda\mathbf{a} + (1-\lambda)\mathbf{b}$$
  * f overestimates $\tilde{f}$ on the entire line between $\mathbf{a}$ and $\mathbf{b}$. In other words
  $$f(\mathbf{x}) > \mu f(\mathbf{a}) + (1-\mu)\mathbf{b} \text{ for all } \mu \in [0,1]$$

  Since $f(\mathbf{x}) > \tilde{f}(\mathbf{x})$, $\mathscr{O}$ must be non-empty.

  Clearly, $\tilde{f}$ will try to 'fix' this lack of convexity, and it will do it by taking the value of the lowest straight line between $(\mathbf{a}, f(\mathbf{a}))$ and $(\mathbf{b}, f(\mathbf{b}))$. Thus, letting
  $$(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\lambda}) = \operatorname{arginf}_{(\mathbf{a},\mathbf{b},\lambda)\in\mathscr{O}} \{\lambda f(\mathbf{a}) + (1-\lambda) f(\mathbf{b})\}$$
  We have that
  $$\tilde{f}(\mathbf{x}) = \tilde{\lambda} f(\tilde{\mathbf{a}}) + (1-\tilde{\lambda}) f(\tilde{\mathbf{b}}) \tag{5}$$
  Now consider equation 4 again. We'll put the following value of $\mathbf{y}$ in there
  $$\mathbf{y} = \begin{cases} \tilde{\mathbf{a}} & \text{if } \mathbf{z}\cdot(\tilde{\mathbf{b}} - \tilde{\mathbf{a}}) \leq f(\tilde{\mathbf{b}}) - f(\tilde{\mathbf{a}}) \\ \tilde{\mathbf{b}} & \text{if } \mathbf{z}\cdot(\tilde{\mathbf{b}} - \tilde{\mathbf{a}}) \geq f(\tilde{\mathbf{b}}) - f(\tilde{\mathbf{a}}) \end{cases}$$
  By multiplying the first condition by $(1-\lambda)$ and the second condition $-\lambda$, we find that this is equivalent to
  $$\mathbf{y} = \begin{cases} \mathbf{a} & \text{if } \mathbf{z}\cdot(\mathbf{x} - \tilde{\mathbf{a}}) \leq (1-\tilde{\lambda})\,(f(\mathbf{b}) - f(\tilde{\mathbf{a}})) \\ \mathbf{b} & \text{if } \mathbf{z}\cdot(\mathbf{x} - \tilde{\mathbf{b}}) \leq \tilde{\lambda}\left(f(\tilde{\mathbf{a}}) - f(\tilde{\mathbf{b}})\right) \end{cases}$$
  Feeding into equation 4, the RHS becomes $\tilde{f}(\mathbf{x})$, as given in equation 5. Thus, we find that $f^{**} \leq \tilde{f}$.

- $\boxed{f^{**}(\mathbf{x}) \geq \tilde{f}(\mathbf{x})}$ This side is significantly harder! Suppose that it is not true – that $f^{**}(\mathbf{x}) < f(\mathbf{x})$ at a given point $\mathbf{x}$. By the strictly Separating Hyperplane Theorem, applied, to the epigraph of $\tilde{f}$, there exists a vector $\mathbf{g}$ and a number γ such that

$$[\mathbf{y}, \alpha] \cdot \begin{bmatrix} \mathbf{g} \\ \gamma \end{bmatrix} < [\mathbf{x}, f^{**}(\mathbf{x})] \cdot \begin{bmatrix} \mathbf{g} \\ \gamma \end{bmatrix} \qquad \forall\, [\mathbf{y}, \alpha] \in \text{epi}(\tilde{f})$$

Now, consider that

- We must have $\gamma \leq 0$. Otherwise, we would be able to increase α above to infinity[8] and make the LHS arbitrarily large. This is impossible since the RHS is finite.
- We must have $\gamma \neq 0$, because if we had $\gamma = 0$, the inequality would read

$$\mathbf{y} \cdot \mathbf{g} < \mathbf{x} \cdot \mathbf{g} \quad \forall \mathbf{y}$$

  This is clearly nonsense; feeding $\mathbf{y} = \mathbf{x}$ into the LHS violates the inequality.

[8] Recall that the definition of epi($\tilde{f}$) admits any $[\mathbf{y}, \alpha]$ such that $\alpha \geq f(\mathbf{y})$. Thus, increasing α to infinity keeps the vector in the epigraph.

Thus,

$$\gamma < 0$$

As such, we can divide both sides of the inequality by $|\gamma|$, denote $\tilde{\mathbf{z}} = \mathbf{g}/\gamma$, and obtain

$$[\mathbf{y}, \alpha] \cdot \begin{bmatrix} \tilde{\mathbf{z}} \\ -1 \end{bmatrix} < [\mathbf{x}, f^{**}(\mathbf{x})] \cdot \begin{bmatrix} \tilde{\mathbf{z}} \\ -1 \end{bmatrix} \qquad \forall\, [\mathbf{y}, \alpha] \in \text{epi}(\tilde{f})$$

Expanding the dot product, assuming $\alpha = f(\mathbf{y})$ (which keeps $[\mathbf{y}, \alpha] \in \text{epi}(\tilde{f})$), and taking the supremum over the RHS

$$\sup_{\mathbf{y}} \{\mathbf{y} \cdot \tilde{\mathbf{z}} - f(\mathbf{y})\} < \mathbf{x} \cdot \tilde{\mathbf{z}} - f^{**}(\mathbf{x})$$

Re-arranging, we find that

$$f^{**}(\mathbf{x}) < \tilde{\mathbf{z}} \cdot \mathbf{x} - \sup_{\mathbf{y}} \{\tilde{\mathbf{z}} \cdot \mathbf{y} - f(\mathbf{y})\}$$

We have therefore found a value of $\tilde{\mathbf{z}}$ for which the RHS is greater than $f^{**}(\mathbf{x})$. This, however, is a contradiction, because $f^{**}(\mathbf{x})$ is given by taking the supremum of the RHS over all $\mathbf{z}$ (see equation 4).

Written in full, Frenchel's Inequality states that

$$\mathbf{p} \cdot \mathbf{x} \leq f(\mathbf{x}) + \sup_{\mathbf{y}} \{\mathbf{p} \cdot \mathbf{y} - f(\mathbf{y})\}$$

Setting $\mathbf{y} = \mathbf{x}$ in the supremum, the RHS becomes $\mathbf{p} \cdot \mathbf{x}$. Clearly, this is less than or equal to what the RHS *could* attain at its supremum, and this proves the inequality.

To prove Young's Inequality, consider the function

$$f(x) = \frac{x^p}{p}$$

Its convex conjugate is given by

$$f^*(x) = \sup_y \left\{ xy - \frac{y^p}{p} \right\}$$

Differentiating with respect to y and setting to 0, we find that this occurs when

$$x - y^{p-1} = 0 \Rightarrow y = x^{1/(p-1)}$$

Feeding this back into the above, we find that

$$f^*(x) = x^{\frac{1}{p-1}+1} - \frac{1}{p} x^{\frac{p}{p-1}}$$

Re-arranging

$$f^*(x) = x^{\frac{p}{p-1}} - \frac{1}{p} x^{\frac{p}{p-1}}$$

Now, note that since $p^{-1} + q^{-1} = 1$, we have that $q = \frac{p}{p-1}$. Thus

$$f^*(x) = x^q \left(1 - \frac{1}{p}\right)$$

Thus

$$f^*(x) = \frac{x^q}{q}$$

Feeding this into Fenchel's Inequality directly gives Young's Inequality.

■ □ ■

Question 4 (***Risk Measures)

[4]Consider a set of possible future scenarios $\{1, \ldots, n\}$. A *portfolio* is defined by a vector $\mathbf{x} \in \mathbb{R}^n$. If scenario i occurs, the profit of the portfolio is x_i (if $x_i < 0$, we interpret this as a loss). We define $\mathbf{1}$ to be 'cash' – a portfolio that is worth 1 in every future scenario. We define $\mathbf{0}$ to be the empty portfolio – one that is worth nothing in every future scenario.

We wish to quantify the level of risk of a portfolio. To this end, we say that a function $\rho : \mathbb{R}^n \to \mathbb{R}$ is a *coherent risk measure* if it satisfies the four axioms below. A coherent risk measure can be interpreted as a *capital requirement* – the amount of money an agent might prudently set aside in order to cover potential losses associated with portfolio $\mathbf{x}$.

Monotonicity If $\mathbf{x} \leq \mathbf{y}$, then $\rho(\mathbf{x}) \geq \rho(\mathbf{y})$. In other words, if one portfolio is more profitable than another in every scenario, it should require less capital.

[4]From the DRO Qualifying Exam, 2010

Cash Invariance For any $t \in \mathbb{R}$, $\rho(\mathbf{x} + t\mathbf{1}) = \rho(\mathbf{x}) - t$. In other words, injecting a quantity of cash into the portfolio commensurately reduces the capital requirement.

Sub-additivity $\rho(\mathbf{x} + \mathbf{y}) \leq \rho(\mathbf{x}) + \rho(\mathbf{y})$. In other words, diversification is good – the combination of two portfolios' capital requirement should be no more than the sum of the individual capital requirements.

Positive Homogeneity For any $\lambda \in \mathbb{R}$ with $\lambda \geq 0$, $\rho(\lambda\mathbf{x}) = \lambda\rho(\mathbf{x})$. In other words, the capital requirement scales individually with the size of the portfolio.

Part A

Suppose that ρ is a coherent risk measure. Prove that it is convex and that $\rho(t\mathbf{1}) = -t$ for all $t \in \mathbb{R}$.

Solution

Consider two portfolios $\mathbf{x}, \mathbf{y}$, and a $\lambda \in [0, 1]$. Let $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$. Then

$$\begin{aligned}
\rho(\mathbf{z}) &= \rho\left(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\right) \\
&\leq_{\text{Axiom 3}} \rho(\lambda\mathbf{x}) + \rho((1-\lambda)\mathbf{y}) \\
&=_{\text{Axiom 4}} \lambda\rho(\mathbf{x}) + (1-\lambda)\rho(\mathbf{y})
\end{aligned}$$

Thus, ρ is indeed convex.

Consider further that

$$\begin{aligned}
\rho(t\mathbf{1}) &= \rho(\mathbf{0} + t\mathbf{1}) \\
&=_{\text{Axiom 2}} \rho(\mathbf{0}) - t \\
&=_{\text{Axiom 4}} 0\rho(\mathbf{0}) - t \\
&= -t
\end{aligned}$$

Part B

An *acceptance set* $\mathscr{A} \subset \mathbb{R}^n$ is a collection of portfolios that satisfies

- $\mathscr{A}$ is a closed convex cone.
- $-\mathbf{1} \notin \mathscr{A}$.
- $\mathbb{R}^n_+ \subset \mathscr{A}$.

Suppose that $\mathscr{A}$ is an acceptance set. For $\mathbf{x} \in \mathbb{R}^n$, define

$$\rho(\mathbf{x}) = \inf\left\{t \in \mathbb{R} : t\mathbf{1} + \mathbf{x} \in \mathscr{A}\right\} \tag{6}$$

Prove that ρ is a coherent risk measure.

Here, $\rho(\mathbf{x})$ can be interpreted as the the minimum quantity of cash which must be injected into the portfolio $\mathbf{x}$ to drive it into some set of desirable portfolios.

Solution

Let's consider each of the four axioms

Monotonicity : If $\mathbf{x} \leq \mathbf{y}$, then $\mathbf{y} = \mathbf{x} + \mathbf{z}$, with $\mathbf{z} \in \mathbb{R}^n_+$. Now, consider that

$$\begin{aligned} \rho(\mathbf{y}) &= \inf\{t : t\mathbf{1} + \mathbf{y} \in \mathscr{A}\} \qquad (7) \\ &= \inf\{t : t\mathbf{1} + \mathbf{x} + \mathbf{z} \in \mathscr{A}\} \end{aligned}$$

Now, consider the vector $t\mathbf{1} + \mathbf{x} + \mathbf{z}$ when $t = \rho(\mathbf{x})$. In that case

- We know that $t\mathbf{1} + \mathbf{x} \in \mathscr{A}$, by definition of $\rho(\mathbf{x})$.
- We also have $\mathbf{z} \in \mathscr{A}$ because $\mathbb{R}^n_+ \subset \mathscr{A}$.

Thus, $\rho(\mathbf{x})\mathbf{1} + \mathbf{x} + \mathbf{z}$ is the sum of two vectors in the cone $\mathscr{A}$, and is therefore in the cone itself[9]. As such, we see that $t = \rho(\mathbf{x})$ is feasible in the problem in line 7. Thus, $\rho(\mathbf{y})$ must be less than or equal to $\rho(\mathbf{x})$.

[9] To see why, consider that for a convex cone $\mathscr{A}$, if $\mathbf{x}, \mathbf{y} \in \mathscr{A}$, then by the cone property $2\mathbf{x}, 2\mathbf{y} \in \mathscr{A}$, and by convexity $\frac{1}{2}2\mathbf{x} + \frac{1}{2}2\mathbf{y} = \mathbf{x} + \mathbf{y} \in \mathscr{A}$.

Cash invariance : Consider that

$$\begin{aligned} \rho(\mathbf{x} + t\mathbf{1}) &= \inf\{\tau \in \mathbb{R} : \tau\mathbf{1} + \mathbf{x} + t\mathbf{1} \in \mathscr{A}\} \\ &= \inf\{\tau \in \mathbb{R} : (\tau + t)\mathbf{1} + \mathbf{x} \in \mathscr{A}\} \\ &= \inf\{\gamma - t \in \mathbb{R} : \gamma\mathbf{1} + \mathbf{x} \in \mathscr{A}\} \\ &= \inf\{\gamma \in \mathbb{R} : \gamma\mathbf{1} + \mathbf{x} \in \mathscr{A}\} - t \\ &= \rho(\mathbf{x}) - t \end{aligned}$$

Sub-additivity : Consider that

$$\begin{aligned} \rho(\mathbf{x}) + \rho(\mathbf{y}) &= \inf_{t_1}\{t_1 : t_1\mathbf{1} + \mathbf{x} \in \mathscr{A}\} \\ &\qquad + \inf_{t_2}\{t_2 : t_2\mathbf{1} + \mathbf{x} \in \mathscr{A}\} \\ &\geq^{10} \inf_{t_1+t_2}\{t_1 + t_2 : (t_1 + t_2)\mathbf{1} + \mathbf{x} + \mathbf{y} \in \mathscr{A}\} \\ &= \inf_{w}\{w : w\mathbf{1} + \mathbf{x} + \mathbf{y} \in \mathscr{A}\} \\ &= \rho(\mathbf{x} + \mathbf{y}) \end{aligned}$$

[10] To see why, consider that if we have a t_1 and a t_2 feasible for the line above, then $t_1\mathbf{1} + \mathbf{x} \in \mathscr{A}$ and $t_2\mathbf{1} + \mathbf{y} \in \mathscr{A}$. As such, as we saw in the previous sidenote, the sum of these two vectors, $(t_1+t_2)\mathbf{1}+\mathbf{x}+\mathbf{y} \in \mathscr{A}$. Thus, the solution to the previous line is feasible in this line, so by allowing the joint optimization, we can only do better and therefore *reduce* our optimal value.

Positive homogeneity : Consider that

$$\begin{aligned}
\rho(\lambda \mathbf{x}) &= \inf\{t : t\mathbf{1} + \lambda\mathbf{x} \in \mathscr{A}\} \\
&= \inf\left\{t : \lambda\left[\frac{t}{\lambda}\mathbf{1} + \mathbf{x}\right] \in \mathscr{A}\right\} \\
&\overset{11}{=} \inf\left\{t : \frac{t}{\lambda}\mathbf{1} + \mathbf{x} \in \mathscr{A}\right\} \\
&= \inf\{\lambda w : w\mathbf{1} + \mathbf{x} \in \mathscr{A}\} \\
&= \lambda\rho(\mathbf{x})
\end{aligned}$$

[11] Here, we are using the fact that if $\mathscr{A}$ is a cone, $\lambda\mathbf{x} \in \mathscr{A} \Leftrightarrow \mathbf{x} \in \mathscr{A}$.

ρ therefore satisfies all four axioms of a coherent risk measure.

Part C

Suppose that ρ is a coherent risk measure. Define the set of portfolios

$$\mathscr{A} = \{\mathbf{x} \in \mathbb{R}^n : \rho(\mathbf{x}) \leq 0\}$$

Prove that $\mathscr{A}$ is an acceptance set, and that with this choice of acceptance set, ρ can be written in the form given in equation 6.

Solution

Let us show that $\mathscr{A}$ satisfies all the properties of an acceptance set outlined in the previous part

$\mathscr{A}$ **is a cone** : First, consider that

$$\mathbf{x} \in \mathscr{A} \Leftrightarrow \rho(\mathbf{x}) \leq 0$$

However, by the definition of a coherent risk measure, $\rho(\alpha\mathbf{x}) = \alpha\rho(\mathbf{x})$ for all $\alpha \geq 0$. Thus, for any $\alpha \geq 0$,

$$\mathbf{x} \in \mathscr{A} \Leftrightarrow \rho(\mathbf{x}) \leq 0 \Leftrightarrow \rho(\alpha\mathbf{x}) \leq 0 \Leftrightarrow \alpha\mathbf{x} \in \mathscr{A}$$

Thus, $\mathscr{A}$ is indeed a cone.

$\mathscr{A}$ **is convex** : Consider $\mathbf{x}, \mathbf{y} \in \mathscr{A}$. By definition of $\mathscr{A}$, we have that $\rho(\mathbf{x}), \rho(\mathbf{y}) \leq 0$. Now consider $\lambda \in [0,1]$ and consider $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$. Since ρ is convex, we have that

$$\rho(\mathbf{z}) \leq \lambda\rho(\mathbf{x}) + (1-\lambda)\rho(\mathbf{y}) \leq 0$$

Thus, $\mathbf{z} \in \mathscr{A}$, and $\mathscr{A}$ is convex.

$-\mathbf{1} \notin \mathscr{A}$: We showed in part (A) that $\rho(-\mathbf{1}) = 1 > 0$. Thus, $-\mathbf{1} \notin \mathscr{A}$.

$\mathbb{R}^n_+ \subset \mathscr{A}$: We showed in part (A) that $\rho(\mathbf{0}) = 0$. Now take an $\mathbf{x} \in \mathbb{R}^n_+$. Clearly, $\mathbf{0} \leq \mathbf{x}$. As such, by Axiom 1, $\rho(\mathbf{x}) \leq \rho(\mathbf{0}) = 0$, and so $\mathbf{x} \in \mathscr{A}$.

$\mathscr{A}$ **is closed** : This boils down to whether ρ is continuous. This can be shown using positive homogeneity (axiom 4).

Now, having established that $\mathscr{A}$ is an acceptance cone, let's see what happens when we use this acceptance set in expression 6 for $\rho(\mathbf{x})$

$$\begin{aligned}
\tilde{\rho}(\mathbf{x}) &= \inf\{t \in \mathbb{R} : t\mathbf{1} + \mathbf{x} \in \mathscr{A}\} \\
&= \inf\{t \in \mathbb{R} : \rho(t\mathbf{1} + \mathbf{x}) \leq 0\} \\
&= \inf\{t \in \mathbb{R} : \rho(\mathbf{x}) - t \leq 0\} \\
&= \inf\{t \in \mathbb{R} : \rho(\mathbf{x}) \leq t\} \\
&= \rho(\mathbf{x})
\end{aligned}$$

Part D

Define $\mathscr{P}$ to be the probability simplex over a set of future scenarios

$$\mathscr{P} = \{\mathbf{q} \in \mathbb{R}^n : \mathbf{q} \geq \mathbf{0}, \mathbf{1} \cdot \mathbf{q} = 1\}$$

Given a distribution $\mathbf{q} \in \mathscr{P}$ and a portfolio $\mathbf{x} \in \mathbb{R}^n$, define $\mathbb{E}_{\mathbf{q}}[\mathbf{x}] = \mathbf{q} \cdot \mathbf{x}$ to be the expected profit of $\mathbf{x}$ under the distribution $\mathbf{q}$.

Suppose that ρ is a coherent risk measure. Prove that there exists a closed convex subset $\mathscr{Q} \subseteq \mathscr{P}$ of probability distributions so that, for all $\mathbf{x} \in \mathbb{R}^n$,

$$\rho(\mathbf{x}) = \sup_{\mathbf{q} \in \mathscr{Q}} \mathbb{E}_{\mathbf{q}}[-\mathbf{x}]$$

Here, $\rho(\mathbf{x})$ can be interpreted as the worst-case expected loss of the portfolio $\mathbf{x}$, over some set of probability distributions.

Solution

Consider some coherent risk measure ρ, and consider its convex conjugate

$$\rho^*(\mathbf{a}) = \sup_{\mathbf{y}} \{\mathbf{a} \cdot \mathbf{y} - \rho(\mathbf{y})\}$$

Now, note that

- If any component of $\mathbf{a}$ is positive, increasing that component of y can only make the expression in the supremum larger. Indeed, increasing that component of y decreases ρ (by monotonicity) and certainly increases $\mathbf{a} \cdot \mathbf{y}$.

  Thus, for any $\mathbf{a}$ with any positive component, $\rho^*(\mathbf{a})\infty$.

- Now assume that all components of $\mathbf{a}$ are negative, and let the sum of the components of a be denoted by a_+.

Now consider what happens what happens to the expression in the supremum if we add $\alpha\mathbf{1}$ to it, for some $\alpha \in \mathbb{R}$

$$\begin{aligned}\mathbf{a}\cdot(\mathbf{y}+\alpha\mathbf{1})-\rho(\mathbf{y}+\mathbf{1}) &= \mathbf{a}\cdot\mathbf{y}+\alpha a_{+}-\rho(\mathbf{y})+\alpha \\ &= \mathbf{a}\cdot\mathbf{y}-\rho(\mathbf{y})+\alpha(1+a_{+})\end{aligned}$$

Thus, if $a_{+} \in (0,-1)$, we can blow this expression up to infinity by making α arbitrarily large and positive, and if $a_{+} \in (-1,-\infty]$, we can blow this expression up to infinity by making α arbitrarily large and negative. Either way, unless $a_{+}=-1$, $\rho^{*}(\mathbf{a})\infty$.

We therefore have that

$$\rho^{*}(\mathbf{a}) = \begin{cases} 0 & \text{for } \mathbf{a}\in\tilde{\mathscr{Q}} \\ \infty & \text{otherwise}\end{cases}$$

where, so far, all we know about $\tilde{\mathscr{Q}}$ is that

$$\tilde{\mathscr{Q}} \subseteq \{\mathbf{x}:\mathbf{x}\leq\mathbf{0},\mathbf{1}\cdot\mathbf{x}=-1\}=-\mathscr{P}$$

Let us now show that $\mathscr{Q}$ is convex and closed

Convex : Take two vectors $\mathbf{q}_1,\mathbf{q}_2 \in se\tilde{t}Q$. This means that for each of these vectors, there is no $\mathbf{y}$ that makes the expression in the supremum above strictly positive[12]. Now consider a $\lambda\in[0,1]$ and a $\mathbf{z}=\lambda\mathbf{q}_1+(1-\lambda)\mathbf{q}_2$. We then have

$$\begin{aligned}\mathbf{a}\cdot\mathbf{z}-\rho(\mathbf{z}) &= \mathbf{a}\cdot(\lambda\mathbf{q}_1+(1-\lambda)\mathbf{q}_2)-\rho\left(\lambda\mathbf{q}_1+(1-\lambda)\mathbf{q}_2\right) \\ &\leq \mathbf{a}\cdot(\lambda\mathbf{q}_1+(1-\lambda)\mathbf{q}_2)-\rho\left(\lambda\mathbf{q}_1\right)+\rho\left((1-\lambda)\mathbf{q}_2\right) \\ &= \lambda\left(\mathbf{a}\cdot\mathbf{q}_1+\rho(\mathbf{q}_1)\right)+(1-\lambda)\left(\mathbf{a}\cdot\mathbf{q}_2+\rho(\mathbf{q}_2)\right) \\ &\leq 0\end{aligned}$$

Thus, $\mathbf{z}\in\tilde{\mathscr{Q}}$ and $\tilde{\mathscr{Q}}$ is convex.

[12]Suppose there was a $\mathbf{y}$ for which

$$\mathbf{q}_1\cdot\mathbf{y}-\rho(\mathbf{y})>0$$

Then by positive homogeneity, we could multiply $\mathbf{y}$ by an arbitrarily large positive constant α to obtain $\alpha\left(\mathbf{q}_1\cdot\mathbf{y}-\rho(\mathbf{y})\right)\to\infty$. This would make $\rho(\mathbf{q}_1)=\infty$, contradicting $\mathbf{q}_1\in\tilde{\mathscr{Q}}$. A similar argument applies to $\mathbf{q}_2$.

Closed : All the inequalities defining $\tilde{\mathscr{Q}}$ are not strict. It can therefore easily be shown that any sequence of vectors in $\tilde{\mathscr{Q}}$ tends to a vector in $\tilde{\mathscr{Q}}$.

Finally, consider the convex conjugate of $\rho^{*}(\mathbf{a})$. Since we established that ρ is a convex function, this biconjugate will give us the original function back. Thus

$$\begin{aligned}\rho(\mathbf{x}) &= \rho^{**}(\mathbf{x}) \\ &= \sup_{\mathbf{a}}\left\{\mathbf{a}\cdot\mathbf{x}-I_{\tilde{\mathscr{Q}}}(\mathbf{a})\right\} \\ &= \sup_{\mathbf{a}\in\tilde{\mathscr{Q}}}\{\mathbf{a}\cdot\mathbf{x}\} \\ &= \sup_{\mathbf{q}\in\mathscr{Q}}\{-\mathbf{q}\cdot\mathbf{x}\} \\ &= \sup_{\mathbf{q}\in\mathscr{Q}}\mathbb{E}_{\mathbf{q}}[-\mathbf{x}]\end{aligned}$$

where $\mathscr{Q} = -\tilde{\mathscr{Q}}$. Since we know $-\mathscr{Q} \subset -\mathscr{P}$ and was closed and convex, we must also have that $\mathscr{Q} \subset \mathscr{P}$ and that $\mathscr{Q}$ is closed and convex. This proves our proposition.

■ □ ■