EECS 127/227AT **Fall 2020**
Solutions for Discussion 5

1 A simple constrained optimization problem

Consider the optimization problem

$$\begin{aligned} &\min_{x_1,x_2\in\mathbb{R}} f(x_1,x_2) \\ &\text{subject to } 2x_1+x_2\geq 1, \\ &\qquad x_1+3x_2\geq 1, \\ &\qquad x_1\geq 0,\ x_2\geq 0. \end{aligned}$$

(a) Sketch the feasible set.

Solution:
See Figure 1.

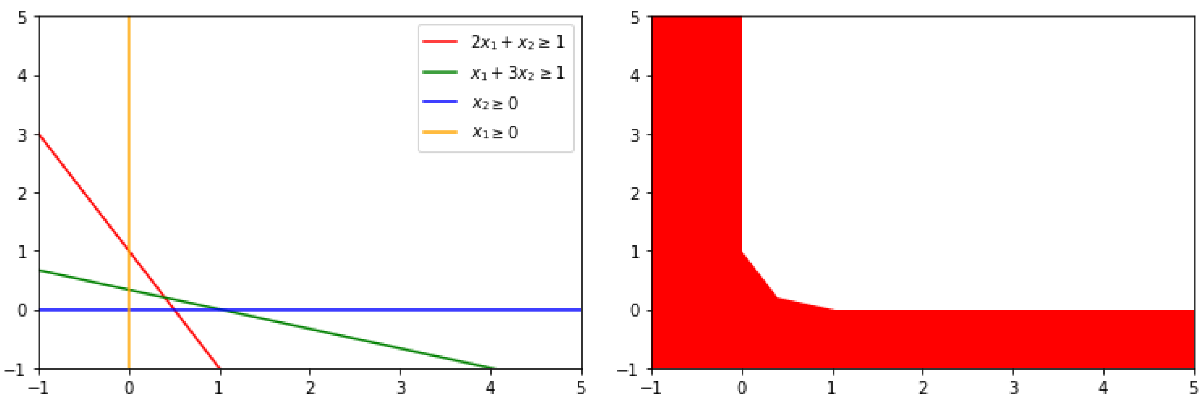

Figure 1: The feasible set is in white on the figure on the right.

(b) For each of the following objective functions, give the optimal set or the optimal value.

i. $f(x_1,x_2)=x_1+x_2$.
Solution:
Using the heat map of the objective function (see figure 2) it seems that the solution is such that $x_1^\star=\frac{2}{5}$ and $x_2^\star=\frac{1}{5}$.
One can verify the optimality of (x_1^*,x_2^*) using the first order convexity condition:

$$\nabla f(\frac{2}{5},\frac{1}{5})^\top((x_1,x_2)-(\frac{2}{5},\frac{1}{5}))\geq 0,\quad \forall (x_1,x_2)\in\mathcal{X},$$

Where $\mathcal{X}$ is the feasible set. Note that here $\nabla f(x_1,x_2)=\begin{bmatrix}1 & 1\end{bmatrix}^T$, at all (x_1,x_2) (here we are thinking of f as having the domain $\mathbb{R}^2$, so we can talk about its gradient even when we are outside the feasible set or on the boundary of the feasible set). The inequality follows from the fact that $x_1+x_2\geq\frac{3}{5}$ for all feasible (x_1,x_2), which comes from adding 2 times the first constraint and the second constraint.
To check that (x_1^*,x_2^*) is the unique optimum point in the feasible set, note that from every other feasible point there is a direction in the feasible set in which x_1+x_2 strictly decreases.

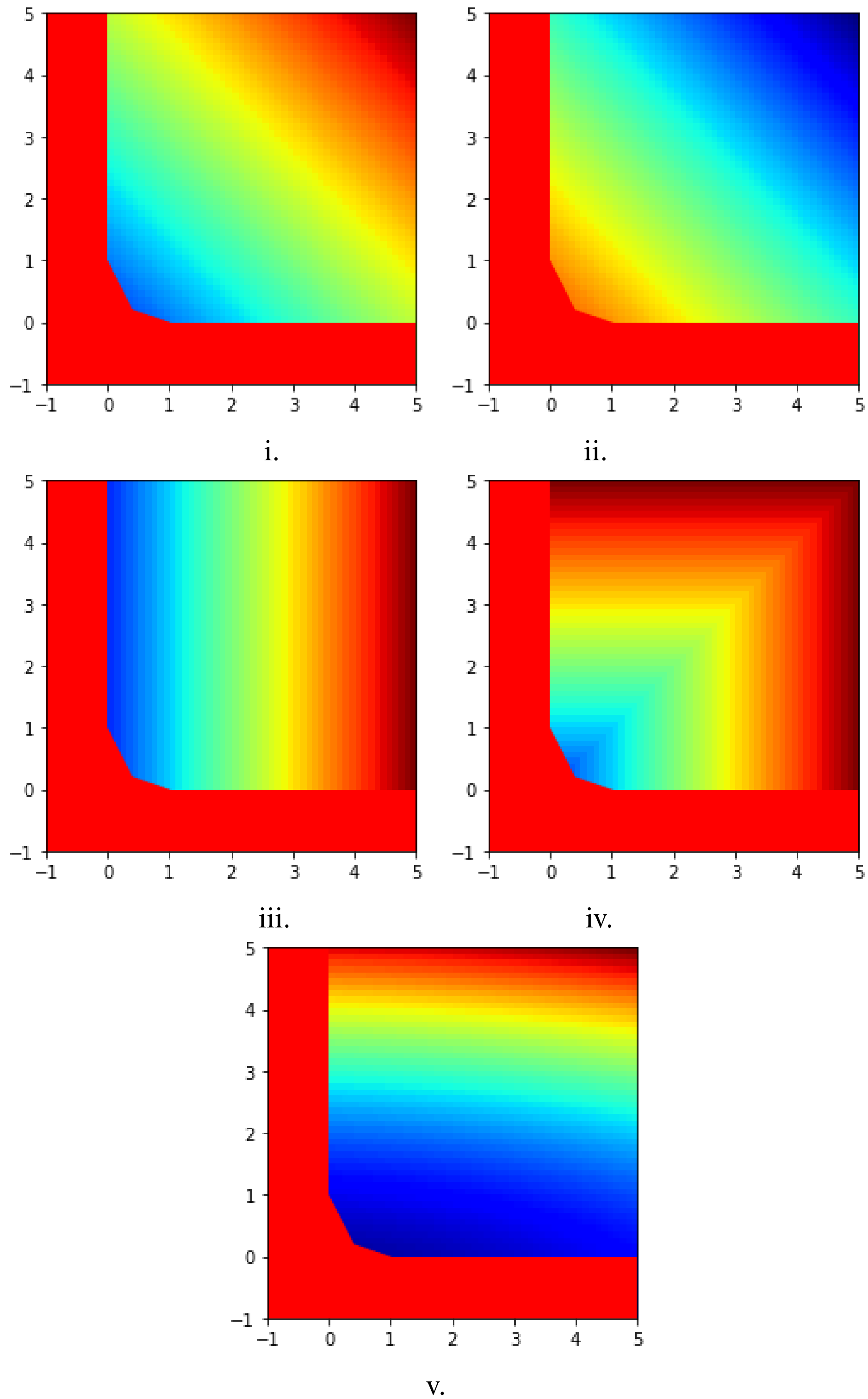

Figure 2: Solution of 2(b)i.: $x^\star = (\frac{2}{5}, \frac{1}{5})$; 2(b)ii. the objective is unbounded below; solutions of 2(b)iii.: $x^\star \in \{(0, x_2) : x_2 \geq 1\}$; solution of 2(b)iv.: $x^\star = (\frac{1}{4}, \frac{1}{4})$; solution of 2(b)v.: $x^\star = (\frac{2}{5}, \frac{1}{5})$. In red below the feasible set are the infeasible points, then the level sets of the objective are shown with a heat map, where the more blue points are those (x_1, x_2) with lower values for $f(x_1, x_2)$ and the more red points are those with highest values for $f(x_1, x_2)$.

ii. $f(x_1, x_2) = -x_1 - x_2$.

Solution:

Here (see figure 2) the objective function is unbounded below, since if $(x_1, x_2) = t(1, 1)$ with $t \geq 0$ then (x_1, x_2) is feasible and $-2t \to -\infty$ when $t \to \infty$.

iii. $f(x_1, x_2) = x_1$.

Solution:

The set of solutions is $S = \{x : x_1 = 0 \text{ and } x_2 \geq 1\}$ (see figure 2).

iv. $f(x_1, x_2) = \max\{x_1, x_2\}$.

Solution:

Using the drawing (see figure 2) it seems that the solution is such that

$$x_1^\star = x_2^\star = \frac{1}{3}.$$

Here, it might be hard to use the first order convexity condition, as the objective function is not differentiable (you can use sub-gradients, but learning how to do this systematically is beyond the scope of class).

Another technique is to use a slack variable (see Sec. 8.3.4.4 of the textbook of Calafiore and El Ghaoui for more about the use of slack variables). The problem is equivalent to

$$\begin{aligned} &\min_{x_1, x_2, t} t \\ \text{subject to } & t \geq x_1, t \geq x_2, \\ & 2x_1 + x_2 \geq 1, \\ & x_1 + 3x_2 \geq 1, \\ & x_1 \geq 0, \; x_2 \geq 0. \end{aligned}$$

Here we can use the first order condition to check optimality. We think of the objective function now as the function of (t, x_1, x_2) that equals t and its gradient is $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$. The triple $(t^*, x_1^*, x_2^*) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is feasible, and for every feasible (t, x_1, x_2) we have $\frac{1}{3}(t - \frac{1}{3}) + 0(x_1 - \frac{1}{3}) + 0(x_1 - \frac{1}{3}) \geq 0$, because this condition boils down to $t \geq \frac{1}{3}$ and this can be seen to hold from the inequalities $t \geq x_1$, $t \geq x_2$, and $2x_1 + x_2 \geq 1$.

We can also argue that (t^*, x_1^*, x_2^*) is the unique optimal point. The only feasible point (t, x_1, x_2) with $t = \frac{1}{3}$ is (t^*, x_1^*, x_2^*), so if there were some other optimal point (t, x_1, x_2) we would have $t > \frac{1}{3}$ at that point. For (t, x_1, x_2) to be feasible we must have $t \geq x_1$ and $t \geq x_2$. If either x_1 or x_2 were strictly less than t, we could increase it up to t, while still staying feasible and optimal, so we would conclude that (t, t, t) is also an optimal point, with $t > \frac{1}{3}$. But this is false because for sufficiently small $\epsilon > 0$ the point $(t - \epsilon, t - \epsilon, t - \epsilon)$ would still be feasible and would strictly improve the objective.

v. $f(x_1, x_2) = x_1^2 + 9x_2^2$.

Solution:

Using the drawing (see figure 2) it seems that the solution is such that $x_1^\star = \frac{1}{2}$ and $x_2^\star = \frac{1}{6}$.

This can be verified using the first order convexity condition:

$$\nabla f(\frac{1}{2}, \frac{1}{6})^\top ((x_1, x_2) - (\frac{1}{2}, \frac{1}{6})) \geq 0, \quad \forall (x_1, x_2) \in \mathcal{X},$$

where $\mathcal{X}$ is the feasible set. Indeed, here we have $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 & 18x_2 \end{bmatrix}^T$, so $\nabla f(x_1^*, x_2^*) = \begin{bmatrix} 1 & 3 \end{bmatrix}^T$, and the inequality to be checked becomes $x_1 + 3x_2 \geq 1$ for all feasible (x_1, x_2), which is true.

In fact, the inequality above holds strictly for all feasible points except for those feasible points (x_1, x_2) that are on the interval from $(\frac{2}{5}, \frac{1}{5})$ to $(1, 0)$. Thus none of the feasible points other than those in this interval could be optimal points. We can also check directly that none of these points can be an optimal point either. For this it suffices to observe that the value of the objective at each such point is strictly bigger than its value at (x_1^*, x_2^*), namely that $x_1^2 + 9x_2^2 > \frac{1}{2}$ at each such point. Indeed, the objective restricted to the line where $x_1 + 3x_2 = 1$ is $x_1^2 + (1 - x_1)^2$, which achieves its minimum at the unique value $x_1^* = \frac{1}{2}$.

2 Convex conjugates

For a function $f : \mathbb{R}^n \to \mathbb{R}$, not necessarily a convex function, with a domain $\mathrm{dom}(f)$, which we assume to be nonempty, but not necessarily a convex set, we can define its *conjugate* (also called its *convex conjugate*, *Fenchel conjugate* or *Legendre-Fenchel conjugate*), $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ via the rule

$$f^*(z) := \sup_{x \in \mathrm{dom}(f)} \left(z^T x - f(x) \right).$$

Note that f^* is an extended real valued function and does not take the value $-\infty$. Also note that it is convenient to treat f also as an extended real valued function, taking the value ∞ outside $\mathrm{dom}(f)$, and with this viewpoint we can also write

$$f^*(z) = \sup_{x \in \mathbb{R}^n} \left(z^T x - f(x) \right). \tag{1}$$

Note that, as an extended real valued function, f also does not take the value $-\infty$.

(a) We will now find the conjugate of the convex function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) := -\log x$, with $\mathrm{dom}(f) = \mathbb{R}_{++}$ in a sequence of steps. (You can assume that the logarithm is to the natural base.)

i. Verify that the given function is convex.

Solution:

First of all, we need to check that the domain is nonempty and the function does not take the value $-\infty$, which is clearly the case here. Now, the domain of the function is an open set and the function is twice differentiable at every point in its domain, so we can use the Hessian condition to check for convexity. Here this boils down to taking the second derivative, which is $1/x^2$, and noting that this is nonnegative on $\mathbb{R}_{++}$.

ii. Show that $f^*(z) = \infty$ for $z \geq 0$.

Solution:

Suppose $z \geq 0$. To find the conjugate, we need to compute

$$\sup_{x>0} \left(zx + \log x \right).$$

Note that $zx + \log x \geq \log x$ for all $x > 0$. Further, $\lim_{x\to\infty} \log x = \infty$. Hence $f^*(z) = \infty$ for $z \geq 0$.

iii. Next consider $z < 0$. Show that $\sup_{x>0} (zx + \log x)$ is achieved at $x = \frac{1}{|z|}$, and thereby show that $f^*(z) = -1 - \log |z|$.

Solution:

Suppose now that $z < 0$. To find the conjugate, we need to compute

$$\sup_{x>0} \left(zx + \log x \right).$$

We check when x solves the first derivative condition for an extremum in the function $zx + \log x$, for $x > 0$. This requires us to solve

$$(zx + \log x)' = z + \frac{1}{x} = 0,$$

and we see that the solution is at $x = \frac{1}{|z|}$. (The reason this extremum is a maximum is because the second derivative $\left(z + \frac{1}{x}\right)' = -\frac{1}{x^2}$ is negative at the extremum.) Substituting this value for x gives $f^*(z) = -1 - \log |z|$.

iv. Putting the previous parts together, determine the conjugate f^* of the given function.

Solution:

We have

$$f^*(z) = \begin{cases} -1 - \log|z| & \text{if } z < 0, \\ \infty & \text{if } z \geq 0. \end{cases}$$

It is worthwhile to make some general observations.

First of all, note that f^* is a convex lower semicontinuous function. [1] This will always be the case – for any function f with nonempty domain (even a non-convex function on a non-convex domain) its conjugate is either a convex lower semicontinuous function with nonempty domain or equal to infinity everywhere, and if f is convex (as part of which we alway require that it has a nonempty domain which is a convex set and does not take on the value $-\infty$) then its conjugate will be convex and lower semicontinuous).

Secondly, note that the domain of the conjugate f^* is the set of gradients that are expressed by f. This will also always be the case for the conjugate of a differentiable convex function (and this observation can be suitably generalized in terms of subgradients).

Thirdly, note that the conjugate of the conjugate is the original function (you can work this out for yourself). This will also always be the case if the original function is convex and lower semicontinuous, as is the case here.

(b) Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$. Find the conjugate of the function $f : \mathbb{R}^n \to \mathbb{R}$, with $\text{dom}(f) = \mathbb{R}^n$, given by $f(x) := \|x\|$.

Hint: Your answer will involve the dual norm $\|\cdot\|^*$.

Solution:

In order to follow the hint, we recall the definition of the dual norm, which is given, for all $x \in \mathbb{R}^n$, by

$$\|x\|^* = \max_{z:\|z\|=1} x^T z.$$

We also note that since $\{z \in \mathbb{R}^n : \|z\| = 1\}$ is a compact set (i.e. a closed and bounded set) and since, for any $x \in \mathbb{R}^n$, the function $z \to x^T z$ is a continuous function on this set, the maximum in the definition of the dual norm is achieved. [2]

It is also useful to observe that, being a norm, f is a convex function.

[1] A function $f : \mathbb{R}^n \to \mathbb{R}$ with domain $\text{dom}(f)$ is called lower semicontinuous if its epigraph is a closed subset of $\mathbb{R}^{n+1}$ (see Sec. 8.2.1.3 of the textbook of Calafiore and El Ghaoui). Many of the important functions we deal with in convex optimization are only defined on subsets of $\mathbb{R}^n$ and not on all of $\mathbb{R}^n$, so it is convenient to think of them via their epigraph. When we do this, it becomes important to know whether the epigraph is a closed set or not, because it gives us some idea of what happens when we take limits of sequences of values of the function. This is why lower semicontinuity is an important concept.

[2] This is called Weierstrass's theorem, see Thm. 8.5 in Sec. 8.3.2. of the textbook of Calafiore and El Ghaoui. We will just accept this as a fact.

Let us now find f^*. For any $z \in \mathbb{R}^n$, we have

$$\begin{aligned}
f^*(z) &:= \sup_{x \in \mathbb{R}^n} \left(z^T x - \|x\| \right) \\
&= \sup \left(0, \sup_{x \in \mathbb{R}^n, x \neq 0} \left(z^T x - \|x\| \right) \right) \\
&\overset{(a)}{=} \sup \left(0, \sup_{K>0} K \left(\sup_{x \in \mathbb{R}^n, \|x\|=1} \left(z^T x - 1 \right) \right) \right) \\
&= \sup \left(0, \sup_{K>0} K(\|z\|^* - 1) \right) \\
&= \begin{cases} 0 & \text{if } \|z\|^* \le 1, \\ \infty & \text{if } \|z\|^* > 1, \end{cases}
\end{aligned}$$

where in step (a) we have normalized x by its norm $\|x\|$ and pulled $\|x\|$ out as a scaling factor, which we have denoted by K.

Recall that in the theory of convex sets and functions, given a set $B \subseteq \mathbb{R}^n$ its *indicator function* $I_B : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$I_B(x) = \begin{cases} 0 & \text{if } x \in B, \\ \infty & \text{otherwise.} \end{cases}$$

We thus see that the conjugate of a norm $\|\cdot\|$ on $\mathbb{R}^n$, when the norm is viewed as a function on $\mathbb{R}^n$, is the indicator function of the closed unit ball of the dual norm.

Recall that if $1 \le p, q, \le \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, then the ℓ_q norm is the dual norm corresponding to the ℓ_p norm (in particular, the ℓ_2 norm is self-dual). The result in this part of the problem should give you some more insight into the nature of this duality.

3 Replacing containment by inequalities

Let $K \subseteq \mathbb{R}^n$. In the theory of convex sets and functions, the function

$$I_K(x) := \begin{cases} 0 & \text{if } x \in K, \\ \infty & \text{otherwise,} \end{cases}$$

is called the *indicator function* of K. Note that this terminology is not consistent with the one used in probability theory.

(a) Suppose K is a nonempty convex subset of $\mathbb{R}^n$. Show that I_K is a convex function with domain K.

Solution:
The domain of I_K, defined as $\{x \in \mathbb{R}^n : -\infty < I_K(x) < \infty\}$ is K. For $x_1, x_2 \in \mathrm{dom}(I_K)$ and $\lambda \in [0,1]$ we have $I_K(\lambda x_1 + (1-\lambda)x_2) = 0 \leq \lambda I_K(x_1) + (1-\lambda)I_K(x_2)$, because $I_K(x_1) = I_K(x_2) = 0$ and $\lambda x_1 + (1-\lambda)x_2 \in K$. This establishes the convexity of I_K.

(b) Suppose K is a nonempty closed convex subset of $\mathbb{R}^n$. Let I_K^* denote the conjugate of the indicator function I_K. Show that I_K^* is a convex function, with $\mathrm{dom}(I_K^*)$ being nonempty.

Hint: In fact, the conjugate f^* of *any* function $f : \mathbb{R}^n \to \mathbb{R}$ (convex or not) with nonempty domain $\mathrm{dom}(f)$ (convex or not) is either a convex function or everywhere equal to ∞. You may find it easier to show this more general result.

Solution:
I_K^* is a convex fuction because the conjugate f^* of *any* function $f : \mathbb{R}^n \to \mathbb{R}$ with nonempty domain $\mathrm{dom}(f)$ is convex. To check this, let $z_1, z_2 \in \mathbb{R}^n$ and $\lambda \in [0,1]$. We have, for all $x \in \mathbb{R}^n$, the inequalities

$$f^*(z_1) \geq z_1^T x - f(x) \text{ and } f^*(z_2) \geq z_2^T x - f(x).$$

Adding λ times the first inequality and $(1-\lambda)$ times the second inequality gives

$$\lambda f^*(z_1) + (1-\lambda)f^*(z_2) \geq (\lambda z_1 + (1-\lambda)z_2)^T x - f(x).$$

Taking the supremum over x on the right hand side of this inequality gives

$$\lambda f^*(z_1) + (1-\lambda)f(^*(z_2) \geq f^*(\lambda z_1 + (1-\lambda)z_2),$$

which would establish that f^* is convex if $\mathrm{dom}(f^*)$ were nonempty, which can be shown in general. Here it suffices to show directly that $\mathrm{dom}(I_K^*)$ is nonempty, which is straightforward, because

$$I_K^*(0) := \sup_{x \in K} \left(0^T x - I_K(x)\right) = 0,$$

so $0 \in \mathrm{dom}(I_K^*)$.

(c) Let K be nonempty closed convex subset of $\mathbb{R}^n$. Because $\mathrm{dom}(I_K^*)$ is nonempty, as established in the preceding part of the problem, we can take the conjugate of I_K^*, which we denote by I_K^{**}. Show that $I_K^{**} = I_K$.

Remark: The claim in this part of the problem will not be true if K is a convex set that is not closed. In this case what will happen is that $I_K^{**} = I_{\bar{K}}$, where $\bar{K}$ denotes the closure of K. To get some intuition for this you can work out, for yourself, the case where K is the open interval $(0,1)$ in $\mathbb{R}$. In fact, you can try to prove for yourself that, more generally, if $K \subset \mathbb{R}^n$ is any nonempty set, then

$I_K^{**} = I_{\bar{\text{co}}(K)}$, where $\bar{\text{co}}(K)$ denotes the closed convex hull of K (i.e. the closure of the convex hull of K).

Remark: Let K be nonempty closed convex subset of $\mathbb{R}^n$. What we will have shown in this part of the problem is that

$$\begin{aligned} x \in K \quad &\Leftrightarrow \quad I_K(x) = 0 \\ &\Leftrightarrow \quad I_K(x) \leq 0 \\ &\Leftrightarrow \quad I_K^{**}(x) \leq 0 \\ &\Leftrightarrow \quad \sup_{z\in\mathbb{R}^n} \left(x^T z - I_K^*(z)\right) \leq 0 \\ &\Leftrightarrow \quad x^T z \leq I_K^*(z) \text{ for all } z \in \mathbb{R}^n. \end{aligned}$$

This way of expressing a containment constraint in terms of a family of linear constraints is what lies at the heart of duality in convex optimization, and we will explore this in more detail in the coming lectures.

Solution:

We will first show that $I_K^{**}(x) \leq I_K(x)$ for all $x \in \mathbb{R}^n$. In fact, if $f : \mathbb{R}^n \to \mathbb{R}$ is any function with $\text{dom}(f)$ nonempty and $\text{dom}(f^*)$ nonempty (in particular, if f is a convex function) then we have $f^{**}(x) \leq f(x)$ for all $x \in \mathbb{R}^n$. This can be seen by writing, from the definition of the conjugate, the formula

$$f^*(z) + f(x) \geq z^T x, \text{ for all } z \in \mathbb{R}^n \text{ and } x \in \mathbb{R}^n,$$

(this is called *Fenchel's inequality* or *Young's inequality* when f is differentiable). This gives, for all $x \in \mathbb{R}^n$, the inequality

$$f(x) \geq z^T x - f^*(z), \text{ for all } z \in \mathbb{R}^n.$$

Taking the supremum over $z \in \mathbb{R}^n$ gives

$$f(x) \geq f^{**}(x), \text{ for all } x \in \mathbb{R}^n.$$

So we have proved that $I_K^{**}(x) \leq I_K(x)$ for all $x \in \mathbb{R}^n$. Now, we have $I_K^*(0) = 0$, as proved in the preceding part of the problem, and so

$$I_K^{**}(x) := \sup_{z\in\mathbb{R}^n} \left(x^T z - I_K^*(z)\right) \geq 0, \text{ for all } x \in \mathbb{R}^n,$$

where the inequality comes from substituting the special case $z = 0$. Hence the only way we could have $I_K^{**}(x) < I_K(x)$ would be if $I_K^{**}(x) = s$ for some $0 \leq s < \infty$ while $I_K(x) = \infty$. Recall that K was assumed to be a closed convex set. If $I_K(x) = \infty$ this means $x \notin K$, so we can find a separating hyperplane (a, b), with $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, and $\epsilon > 0$ such that $a^T x + b > \epsilon$ and $a^T z + b < -\epsilon$ for all $z \in K$. But then we have, for all $\beta > 0$, the inequality

$$I_K^*(\beta a) = \sup_{z\in K} \beta a^T z \leq -\beta(b + \epsilon).$$

It follows that

$$\begin{aligned} I_K^{**}(x) \quad &= \quad \sup_{z\in\mathbb{R}^n} \left(x^T z - I_K^*(z)\right) \\ &\geq \quad \beta a^T z - I_K^*(\beta a) \\ &\geq \quad \beta a^T z + \beta(b + \epsilon) \\ &\geq \quad 2\beta\epsilon, \end{aligned}$$

which, by letting $\beta \to \infty$, contradicts the assumption that $I_K^{**}(x) = s$ for $0 \leq s < \infty$. This proves that I_K^{**} is the same as I_K. Recall again that we have assumed that K is a closed convex set to prove this – otherwise it will not be true in general.