ANDREW TULLOCH

CONVEX OPTIMIZATION EXAMPLES

TRINITY COLLEGE
THE UNIVERSITY OF CAMBRIDGE

Contents

1

Example Sheet 1

Ex 1. (i) Theorem 3.5 is proven in the lecture notes.

(ii) Proposition 3.10 is proven as follows

(i) Recall that a function f is convex if and only if epi f is convex. Then we have

$$\operatorname{epi} \sup_{i \in I} f_i = \cap_{i \in I} \operatorname{epi} f_i \tag{1.1}$$

which is the intersection of convex sets, hence convex. Thus, $\sup_{i \in I} f_i$ is convex.

(ii) The case for $|I| = 1$ is trivial. For $|I| = 2$, let f, g be strictly convex and $h = \sup\{f, g\}$. Let $x, y \in X, \lambda \in (0,1), z = \lambda x + (1-\lambda)y$. Then

$$\begin{aligned}
h(z) &= \sup\{f(z), g(z)\} && (1.2)\\
&< \sup\{\lambda f(x) + (1-\lambda) f(y), \lambda g(x) + (1-\lambda) g(y)\} && (1.3)\\
&\leq \lambda \sup\{f(x), g(x)\} + (1-\lambda) \sup\{f(y), g(y)\} && (1.4)\\
&= \lambda h(x) + (1-\lambda) h(y) && (1.5)
\end{aligned}$$

(iii) Let $C_i, i \in I$ be convex, and let $C' = \cap_{i \in I} C_i$. Let $x, y \in C', \lambda \in (0,1), z = \lambda x + (1-\lambda)y$. Then $z \in C_i$ for all $i \in I$ (as C_i are convex), and so $z \in C'$. Thus C' is convex.

(iv) Let $f^k = \sup_{k \geq n} f_i$. f^k is convex as a pointwise supremum of convex functions. Let $x, y \in X, \lambda \in (0,1), z = \lambda x + (1-\lambda)y$.

Then

$$f^k(z) \leq \lambda f^k(x) + (1-\lambda) f^k(y) \tag{1.6}$$
$$\tag{1.7}$$

and taking $k \to \infty$ on both sides, we have

$$\begin{aligned}\limsup f_i(z) &= \lim_{n\to\infty} f^k(z) &(1.8)\\ &\leq \lim_{n\to\infty} f^k(x) + (1-\lambda) f^k(y) &(1.9)\\ &= \lambda \limsup f_i(x) + (1-\lambda)\limsup f_i(y) &(1.10)\end{aligned}$$

and so $\limsup f_i$ is convex.

(iii) Proposition 3.15 proceeds as follows.

(i) It is sufficient to prove for $m = 2$ and use induction. Let A, B be convex sets, let $u, v = (a_1, b_1), (a_2, b_2) \in A \times B, \lambda \in (0,1), z = \lambda u + (1-\lambda) v$. Then

$$z = (\lambda a_1 + (1-\lambda) a_2, \lambda b_1 + (1-\lambda) b_2) \in A \times B \tag{1.11}$$

as A, B are convex.

(ii) Let $y_1, y_2 \in L(C)$. Then $y_i = A x_i + b$ for some $x_i \in C$. Let $\lambda \in (0,1), z = \lambda y_1 + (1-\lambda) y_2$. Then

$$\begin{aligned}z &= \lambda(Ax_1 + b) + (1-\lambda)(Ax_2 + b) &(1.12)\\ &= A(\underbrace{\lambda x_1 + (1-\lambda) x_2}_{\in C}) + b &(1.13)\\ &\in L(C) &(1.14)\end{aligned}$$

as C is convex.

(iii) Let $x_1, x_2 \in L^{-1}(C)$. Let $y_i = A x_i + b$. Let $\lambda \in (0,1), z = \lambda x_1 + (1-\lambda) x_2$. Note that $L(z) = \lambda y_1 + (1-\lambda) y_2 \in C$, and so $z \in L^{-1}(C)$ as required.

(iv) This is the image of the function $f(x_1, x_2) = x_1 + x_2$ on the convex set $C_1 \times C_2$, and is thus convex.

(v) This is the image of the function $f(x) = \lambda x$ on the convex set C, and is thus convex.

Ex 2. (i) This is the intersection of the half planes formed by perpendicular bisectors between points, thus intersection of convex sets, and hence convex.

(ii) Let $(x_1, t_1), (x_2, t_2) \in K, \lambda \in (0, 1)$. Then by properties of the norm,

$$\|\lambda x_1 + (1-\lambda)x_2\| \leq \lambda\|x_1\| + (1-\lambda)\|x_2\| \tag{1.15}$$

$$\leq \lambda t_1 + (1-\lambda)t_2 \tag{1.16}$$

Thus $(\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2) \in K$, and so K is convex.

(iii) Y_1 is convex as the unit sphere is convex. Y_2 is the intersection of the half planes $x_1 \leq 2, x_1 \geq 0, x_2 \leq 1, x_2 \geq -1$, and thus the intersection of convex sets. Thus $Y_1 + Y_2$ is the sum of convex sets, and hence convex.

Ex 3. Let $x \in \Delta_n$. For purposes of contradiction, assume x can be written in two different forms $x = \sum_{i=0}^m \lambda_i v_i = \sum_{i=0}^m \gamma_i v_i, \lambda_i, \gamma_i \geq 0, \sum_{i=0}^m \gamma_i = \sum_{i=0}^m \lambda_i = 0$. Then consider

$$0 = x - x = \sum_{i=0}^{m} (\lambda_i - \gamma_i) v_i \tag{1.17}$$

Then by affine independence of v_i, we have $\lambda_i = \gamma_i$ for all i as required.

Ex 4. If f is an improper convex function, then $f(x) = -\infty$ for every $x \in \operatorname{rint} \operatorname{dom} f$. To show this, let $f(u) = -\infty$, and let $x \in \operatorname{rint} \operatorname{dom} f$. Then there exists $\mu > 1$ such that $y \in \operatorname{dom} f$, where $y = (1-\mu)u + \mu x$. Then $x = (1-\lambda)u + \lambda y$. Then

$$f(x) \leq (1-\lambda)f(u) + \lambda f(y) < (1-\lambda)\alpha + \lambda\beta \tag{1.18}$$

for any $\alpha > f(u)$ and $\beta > f(y)$. As $f(u) = -\infty$ and $f(y) < \infty$, we must have $f(x) = -\infty$.

If f is an improper lower semicontinuous convex function, then the set of points for $f(x) = -\infty$ includes $\operatorname{cl} \operatorname{rint} \operatorname{dom} f$ by lower semi-continuity, and

$$\operatorname{cl} \operatorname{rint} \operatorname{dom} f = \operatorname{cl} \operatorname{dom} f \subset \operatorname{dom} f \tag{1.19}$$

and so an improper lower semicontinuous convex function can have no finite values.

Ex 5. (i) (i) $f''(x) = \frac{2}{x^3} > 0$, thus convex.

(ii) $f''(x) = \exp x > 0$, thus convex.

(iii) $f''(x) = \frac{1}{x^2} > 0$, thus convex.

(iv)

$$H(x,y) = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} \tag{1.20}$$

$$= \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix}^T \begin{bmatrix} y \\ -x \end{bmatrix} \tag{1.21}$$

and so H is positive semidefinite as required.

(v) $\|X\|_\sigma$ is a norm on the s, and all norms are convex. This follows as $x, y \in X, \lambda \in (0,1)$ gives

$$\|\lambda x + (1-\lambda)y\| \leq \lambda\|x\| + (1-\lambda)\|y\| \tag{1.22}$$

by triangle inequality and homogeneity.

(vi)

$$\lambda_{\max}(X) = \sup_{\|v\|=1} \langle Xv, v\rangle \tag{1.23}$$

which is the supremum of convex functions $v \mapsto \langle Xv, v\rangle$, and is hence convex.

(ii) If f is convex, then g is the composition of f with an affine mapping.

(iii) The forward direction is trivial, as it is the composition of a convex function with an affine function, and so is convex.

If g is convex for all t and $u, v \in \mathbb{R}^n$, then for any $\lambda \in (0,1), u, v \in \mathbb{R}^n$ and $t_1, t_2 \in \mathbb{R}$, we must have

$$f(u + [\lambda t_1 + (1-\lambda)t_2]v) = g(\lambda t_1 + (1-\lambda)t_2) \tag{1.24}$$

$$\leq \lambda g(t_1) + (1-\lambda)g(t_2) \tag{1.25}$$

$$= \lambda f(u + t_1 v) + (1-\lambda) f(u + t_2 v) \tag{1.26}$$

Now, we show f is necessarily convex. Let $x, y \in \mathbb{R}^n, \lambda \in (0,1)$. Then, we must show

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \tag{1.27}$$

we just choose u, v, t_1, t_2 such that

$$\lambda x + (1-\lambda)y = u + [\lambda t_1 + (1-\lambda)t_2]v \tag{1.28}$$

$$x = u + t_1 v \tag{1.29}$$

$$y = u + t_2 v \tag{1.30}$$

Such u, v, t_1, t_2 can always be found, and thus f is convex.

Ex 6. As $x \mapsto -log(x)$ is convex on $\mathbb{R}^+$, we have

$$-\log \sum_{i=1}^{k} \lambda_i x_i \leq -\sum_{i=1}^{k} \lambda_i \log x_i \tag{1.31}$$

$$\sum_{i=1}^{k} \lambda_i x_i \geq e^{\sum_{i=1}^{k} \lambda_i \log x_i} \tag{1.32}$$

$$\prod_{i=1}^{k} x_i^{\lambda_i} \leq \sum_{i=1}^{k} \lambda_i x_i \tag{1.33}$$

and letting $\lambda_i = \frac{1}{k}$, we obtain

$$\prod_{i=1}^{n} x_i^{\frac{1}{n}} = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1.34}$$

as required.

Ex 7. We first prove for $n = 1$.

(i) ((1) $\Rightarrow$ (3)) Let f be convex. Then for $x, y \in C, t \in (0,1)$, we have

$$f(x + t(y-x)) \leq (1-t)f(x) + tf(y) \tag{1.35}$$

$$f(y) \leq f(x) + \frac{f(x + t(y-x)) - f(x)}{t} \tag{1.36}$$

and letting $t \to 0$, we obtain

$$f(y) \geq f(x) + f'(x)(y-x) \tag{1.37}$$

(ii) ((3) $\Rightarrow$ (2)) Adding the identities for (x, y) and (y, x) gives

$$f(x) + f(y) \leq f'(x)(y - x) + f'(y)(x - y) + f(x) + f(y) \tag{1.38}$$

which when re-arranged yields

$$(x - y)(f'(x) - f'(y)) \geq 0 \tag{1.39}$$

as required.

(iii) ((2) $\Rightarrow$ (1)) Let $y = x + \epsilon$ for $\epsilon > 0, x, y \in X$. Then

$$(x - y)(f'(x) - f'(y)) \geq 0 \Rightarrow f(x + \epsilon) \geq f(x) \tag{1.40}$$

or alternatively, f' is an increasing function.

Let $x < z < y \in X$.

$$\frac{f(z) - f(x)}{z - x} = f'(\nu)\frac{f(y) - f(z)}{y - z} = f'(\mu) \tag{1.41}$$

for $\nu \in (x, z), \mu \in (z, y)$. Note that $f'(\nu) \leq f'(\mu)$ as f' is increasing. Thus,

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(z)}{y - z} \tag{1.42}$$

and thus f is convex.

(iv) ((1) $\Rightarrow$ (4)) Fill this in

(v) ((4) $\Rightarrow$ (1)) Fill this in

Ex 8. Let $x \in \operatorname{con} X$, so $x = \sum_{i=1}^{p} \lambda_i x_i$ with $\lambda_i \geq 0, \sum_{i=1}^{p} \lambda_i = 1$. If $p \leq n + 1$, there is nothing to prove. Thus, assume $p > n + 1$

Consider the elements $x_j - x_1, 2 \leq j \leq p$. These are $p - 1 > n$ elements of $\mathbb{R}^n$, and thus are linearly dependent. Let $\sum_{i=2}^{p} \gamma_i (x_i - x_1) = 0$ with not all γ_i zero. Let $\gamma_1 = -\sum_{i=2}^{p} \gamma_i$, and then we have

$$\sum_{i=1}^{p} \gamma_i x_i = 0 \tag{1.43}$$

with $\sum_{i=1}^{p} \gamma_i = 0$.

Let $\alpha = \min\{\frac{\lambda_i}{\gamma_i} | \gamma_i > 0\}$. Then $\lambda_i - \alpha\gamma_i$ is non-negative and zero for at least on i. Then we have

$$x = x - 0 = \sum_{i=1}^{p} x_i(\lambda_i - \alpha\gamma_i) = \sum_{i=1}^{p} \theta_i x_i \tag{1.44}$$

with at least one θ_i zero. Thus, we can write x as a convex combination of $p-1$ coefficients. Induction on p shows that every element $x \in \operatorname{con} X$ can be written as a convex combination of at most $n+1$ elements of X as required.

Ex 9. Let $\{v_j\} \in \operatorname{con} C$ be an infinite sequence. By Caratheordory's theorem, there exist $\lambda_{ij} \geq 0$ and $x_{ij} \in X$ such that for every j,

$$v_j = \sum_{i=1}^{n+1} \lambda_{ij} x_{ij} \tag{1.45}$$

and $\sum_{i=1}^{n+1} \lambda_{ij} = 1$.

Note that the simplex $K = \{(\lambda_1, \ldots, \lambda_{n+1}) | \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1\}$ is closed and bounded in $\mathbb{R}^{n+1}$, and is thus compact. Then, we can take an infinite subsequence j' of the λ_{ij} and x_{ij} such that $x_{ij'} \to x_i \in C, \lambda_{ij'} \to \lambda_i \in K$. The subsequence $\{v_{j'}\}$ converges to $\sum_{i=1}^{n+1} \lambda_i x_i \in \operatorname{con} X$ as required. Thus, every sequence has a convergent subsequence, and so $\operatorname{con} X$ is compact.

Ex 10. (i) $K = K_n^{SDP}$ is a cone as $0 \in K$, $A \in K \Rightarrow \lambda A \in K$ for $\lambda \geq 0$ ($x^T A x \geq 0 \Rightarrow x^T \lambda A x \geq 0$). K is a convex cone as $K + K \subseteq K$ (the sum of positive semidefinite matrices is positive semidefinite).

Show K is closed.

(ii) Note that $f(X) = -\log\det X^{-1} = \log\det X$ by properties of the determinant. Consider the function $g(t)$ defined by $g(t) = \log\det(Z + tV)$ for $Z, V \in K$. Then

Isn't this question incorrect?

$$\begin{aligned} g(t) &= \log\det(Z + tV) && (1.46)\\ &= \log\det(Z^{\frac{1}{2}}(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})Z^{\frac{1}{2}}) && (1.47)\\ &= \sum_{i=1}^{n} \log(1 + t\lambda_i) + \log\det Z && (1.48) \end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $Z^{-\frac{1}{2}}VZ^{-\frac{1}{2}}$. Then we have

$$g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1+t\lambda_i)^2} < 0 \tag{1.49}$$

and thus $g''(t) \leq 0$, and so f is concave.

2

Example Sheet 2

Ex. 1 The first direction is trivial. Assume $0 \in \int(C - D)$ and a separating hyperplane (b, β) exists. Then there exists $\epsilon > 0$ such that $B_\epsilon(0) \subseteq \int(C - D)$. Let b_i be some non-zero element of b. Thus, there exists $(x^1, y^1), (x^2, y^2), (x^3, y^3) \in C \times D$ such that $\langle b, x^1 - y^1 \rangle = 0$, $\langle b, x^2 - y^2 \rangle = \epsilon b_i$ and $\langle b, x^3 - y^3 \rangle = -\epsilon b_i$.

Note however, that the condition of the separating hyperplane is such that $\langle b, x - y \rangle \leq 0$ for all $(x, y) \in C \times D$. By contradiction, we have that no such hyperplane exists.

Opposite direction

Ex. 2 At points of continuity of f, the subgradient is simply the singleton set $\{\nabla f\}$. Thus, for $x \neq 0$, $\partial f(x) = \{\frac{x}{\|x\|}$. At $x = 0$, we seek the set of $v \in \mathbb{R}^n$ such that

$$V = \|x\| \geq 0 + \langle v, x \rangle = \langle v, x \rangle \tag{2.1}$$

for all $x \in \mathbb{R}^n$.

We claim that $V = B_1$. First, let $v \in B_1$. Then by definition of the norm as

$$\|x\| = \sup_{v \leq 1} \langle v, x \rangle, \tag{2.2}$$

we have $v \in V$.

Now, let $v \in V$. Then taking $x = v$ in (2.2), we have $\|v\| \geq \|v\|^2$, and so $\|v\| \leq 1$. Thus $v \in B_1$.

Hence, $V = B_1$.

Ex. 3 The problem is convex (sum of composition of convex function $f : x \mapsto x^2$ with affine transform $g : x \mapsto x - a^i$). The function is convex, continuous, level bounded, and proper. Thus, by Theorem 2.14, $\inf f$ is nonempty.

Optimality conditions at $x \in \mathbb{R}^n$ are equivalent to requiring that $0 \in \partial f(x)$. Taking derivatives, this is equivalent to

$$0 = \begin{cases} \sum_{i=1}^m \frac{w_i(x-a_i)}{\|x-a_i\|} & x \neq a_i \forall i \\ \sum_{i=1}^m v_i & \|v_i\| \leq 1, a_i = x \end{cases} \tag{2.3}$$

with obvious interpolation between the two solutions.

In the case where $n = 1$, then the L^1 and L^2 norms are equal, and this is just computing the weighted medians of the a_i. Can just compute

What are the applications for this technique in image processing

Ex. 4 Note that $g(x)$ is affine sum of convex functions, and so is convex. Let x minimize g. Then $0 \in \partial g(x)$, and we have

$$\partial_i g(x) = \begin{cases} 1 + \mu \nabla_i f(x) & x_i > 0 \\ [-1 + \mu \nabla_i f(x), 1 + \mu \nabla_i f(x)] & x_i = 0 \\ -1 + \mu \nabla_i f(x) & x_i < 0 \end{cases} \tag{2.4}$$

and thus if $0 \in \partial g(x)$, we must have

$$\begin{cases} x_i = x_i - \nabla_i f(x) - \frac{1}{\mu} & x_i > 0 \\ |\nabla_i f(x)| \leq \frac{1}{\mu} & x_i = 0 \\ x_i = x_i - \nabla_i f(x) + \frac{1}{\mu} & x_i > 0 \end{cases} \tag{2.5}$$

which is equivalent to the shrinkage operation.

Ex. 5 Note that K is a closed convex cone. As such, we have that $K^{\star\star} =$

$\operatorname{cl} K = K$. We have

$$\begin{aligned} K^{\star\star} &= \{w \in \mathbb{R}^n | \langle w, x\rangle \leq 0 \forall x \in K^\star\} && (2.6)\\ &= \{w \in \mathbb{R}^n | \langle w, Ax\rangle \leq 0 \forall x \geq 0\} && (2.7)\\ &= \{w \in \mathbb{R}^n | \left\langle A^T w, x\right\rangle \leq 0 \forall x \geq 0\} && (2.8)\\ &= \{w \in \mathbb{R}^n | A^T w \leq 0\} && (2.9)\end{aligned}$$

By uniqueness, we have our result.

Now, consider Farkas's lemma. Consider the cone K as above, and consider the two cases, $b \in K^\star$ and $b \notin K^\star$. In the first case, we have that there exists $x \geq 0$ such that $Ax = b$. In the second case, we have $b \notin K^{\star\star\star} = K^\star = \{w | \langle w, x\rangle \leq 0 \forall x \in K\}$, and so there must exist $x \in K$ such that $\langle b, x\rangle > 0$, which is equivalent to requiring that $A^T x \leq 0$ and $\langle b, x\rangle > 0$, and so letting $y = -x$, we have our alternative.

Is this correct? It seems like we must be taking a shortcut since the separating hyperplane theorem is so deep and this seems to require only elementary manipulation.

Ex. 6 (i) $\operatorname{aff} C$ is a closed set, as

(ii) $\operatorname{cl} C$ is the smallest closed set containing C. Then $\operatorname{cl} D$ is a closed set containing D. Thus, $\operatorname{cl} D$ contains C, and so $\operatorname{cl} C \subseteq \operatorname{cl} D$.

(iii) $\int D$ is the largest open set contained in D. Then $\int C$ is an open set contained in D. Thus, $\int C$

(iv)

proof

Ex. 7 (i) Recall that the affine

Ex. 8

Ex. 9

Ex. 10

Ex. 11

Ex. 12

Ex. 13

Ex. 14

Ex. 15

Ex. 16 We can show that $f(u)$ is convex, lower semicontinuous, and proper. Since $f(u) \geq \frac{1}{2}\|u-g\|_2^2$, we have level boundedness. Thus we are guaranteed the existence of a solution.

Since $\|v\| = \sup_{x \neq 0} \frac{\langle x, v \rangle}{\|x\|}_2$.

We can then find a product of scaled unit balls D such that $f(u) = \|u-g\|_2^2 + (\delta_D)^\star(Lu)$.

If we form the perturbed function f', we have

$$f'(u,w) = k(u) + h(Lu + w) \tag{2.10}$$
$$(f')^\star(v,y) = k^\star(-L^T y + v) + \underbrace{h^\star}_{\delta_D(y)}(y) \tag{2.11}$$

with

$$k^\star(z) = \sup_u \langle z, u \rangle - \frac{1}{2}\|u-g\|_2^2 \tag{2.12}$$
$$= \frac{1}{2}\|z-g\|_2^2 - \frac{1}{2}\|g\|_2^2 \tag{2.13}$$

which is a special case of the dual of $\frac{1}{2}\|x\|^2$ is $\frac{1}{2}\|u\|^2$.

Then

$$f^\star(v,y) = \| - L^T y + v - g\|_2^2 + \frac{1}{2}\|g\|_2^2 + \delta_D(y) \tag{2.14}$$
$$\psi(y) = -f^\star(0,y) = -\frac{1}{2}\| - L^T y - g\|_2^2 + \frac{1}{2}\|g\|_2^2 - \delta_D(y) \tag{2.15}$$

and so we have transformed our problem into a quadratic.

$$p(w) = \inf_u f'(u,w) \tag{2.16}$$
$$q(v) = \inf_y f'^\star(v,y) \tag{2.17}$$

Then (u,y) is a primal-dual solution if and only if $(0,y) \in \partial f^{-1}(u,0) \iff (0,y) \in (u - g + L^T \partial(\delta_D)^\star(Lu), \partial(\delta_D)^\star())$

3

Example Sheet 3

Ex. 1

Ex. 2 We first show $\operatorname{con}\{\nabla f_i | i \in I(x)\} \subseteq \partial f(x)$.

First, note that we have for all x, z and $k \in I(x)$,

$$f(z) \geq f_k(z) \geq f_k(x) + \langle \nabla f_k(x), z - x \rangle = f(x) + \langle \nabla f_k(x), z - x \rangle \tag{3.1}$$

and so $\nabla f_k(x) \in \partial f(x)$.

Now, let g be a convex combination of $\nabla f_k(x), k \in I(x)$. Then we have

$$\begin{aligned}
f(x) + \langle g, z - x \rangle &= f(x) + \left\langle \sum_k \lambda_k \nabla f_k, z - x \right\rangle && (3.2)\\
&= f(x) + \sum_k \langle \lambda_k \nabla f_k, z - x \rangle && (3.3)\\
&\leq f(x) + \sum_k \lambda_k (f(z) - f(x)) && (3.4)\\
&= f(x) = f(z) - f(x) && (3.5)\\
&= f(z) && (3.6)
\end{aligned}$$

as required.

We must now show $\partial f(x) \subseteq \operatorname{con}\{\nabla f_i | i \in I(x)\}$.

Recall that $\partial f(x) = \{v | (v, -1) \in N_{\operatorname{epi} f}(x, f(x))\}$.

Then we claim

$$N_{\text{epi}\{\max_i f_i\}}(x, f(x)) = \sum_{i=1}^{n} N_{\text{epi} f_i}(x, f_i(x)). \tag{3.7}$$

We show

Fill in

Ex. 3 We have

$$\begin{aligned}
& y \in B_{\tau f^\star}(x^\star) && (3.8)\\
\iff & y \in (I + \tau \partial f^\star)^{-1}(x^\star) && (3.9)\\
\iff & y \in (I + \tau(\partial f^{-1})^{-1})^{-1}(x^\star) && (3.10)\\
\iff & x^\star \in (I + \tau(\partial f)^{-1})(y) && (3.11)\\
\iff & 0 \in y - x^\star + \tau(\partial f)^{-1}(y) && (3.12)\\
\iff & \frac{x^\star - y}{\tau} \in (\partial f)^{-1}(y) && (3.13)\\
\iff & y \in \partial f(\frac{x^\star - y}{\tau}) && (3.14)\\
\iff & 0 iny - \partial f(\frac{x^\star - y}{\tau}) && (3.15)\\
\iff & 0 \in y + \frac{1}{\tau}\partial f(y - x^{star}) && (3.16)\\
\iff & 0 \in (I + \frac{1}{\tau}\partial f(\cdot - x^\star))(y) && (3.17)\\
\iff & y \in (I + \frac{1}{\tau}\partial f(\cdot - x^\star))^{-1}(0) && (3.18)\\
\iff & y \in (I + \frac{1}{\tau}\partial f)^{-1}(-x^\star) && (3.19)
\end{aligned}$$

Ex. 4 Consider $f_z(x, u) = k(x) + h(z + u - x)$. Then

$$\begin{aligned}
p(u) &= \inf_x f_z(x, u) && (3.20)\\
&= \inf_y k(y) + h(z + u - y) && (3.21)\\
&= F(z + u) && (3.22)
\end{aligned}$$

as required.

Thus $p(0) = F(z)$. By properness of h, z, $F(z) = p(0) \in \mathbb{R}$, and by lsc of h, z, $F(z) = p(0)$ is lsc. Thus strong duality holds.

Consider the dual objective. First, we compute $f^\star(v, y)$. We have

$$f^\star(v, y) = \langle -z, y\rangle + k^\star(y + v) + h^\star(y). \qquad (3.23)$$

Then $\psi(y) = -f^\star(0, y) = \langle z, y\rangle - k^\star(y) - h^\star(y)$.

Thus, we have $\sup_y \psi(y) = \sup_y \langle z, y\rangle - h^\star(y) - k^\star(y) = (h^\star + k^\star)^\star(z)$.

Ex. 5 Given an LP of the form $\max \langle c, x\rangle$ s.t $Ax \leq b$, an $SOCP$ of the form $\max c, x$ s.t $\|A_i x + b_i\|_2 \leq \langle c_i, x\rangle + d_i$, $Fx = g$, and an SDP of the form $\inf \langle c, x\rangle$ s.t. $Ax - b$ is positive semidefinite.

Note that by setting $A_i, b_i = 0$, we obtain that $LP \subseteq SOCP$. Now, note by setting

Ex. 6

Ex. 7

Ex. 8

Ex. 9

Ex. 10 Let our pre-Hilbert space $\mathcal{G}$ be given as the span of κ_x, and let $f, g \in G$. Thus $f = \sum_{i=1}^n a_i \kappa_{x_i}, g = \sum_{j=1}^m b_j \kappa_{x'_j}$. Then let our inner product on G be given as

$$\langle f, g\rangle_{\mathcal{G}} = \sum_{i=1}^n \sum_{j=1}^m a_i \overline{b_j} \kappa(x_i, x'_j) \qquad (3.24)$$

This trivially satisfies the properties of the norm - linearity, conjugate symmetric, and positive definite.

Now, let $\mathcal{H}$ be the metric space completion of $\mathcal{G}$. By Hilbert space theory, $\mathcal{G}$ is dense in $\mathcal{H}$, and we can write every element of $\mathcal{H}$ in the form

$$\sum_{i=1}^\infty a_i \kappa_{x_i}. \qquad (3.25)$$

with appropriate L^2 condition on a_i.

Let $f = \sum_{i=1}^{\infty} a_i \kappa_{x_i}$. Then

$$\langle k_x, f \rangle = \sum_{i=1}^{\infty} a_i \kappa(x_i, x) = f(x). \tag{3.26}$$

as required.

Let κ be a Mercel kernel, and let $\mathcal{H}$ be the Hilbert space constructed before. Then

$$\nu : \mathcal{F} \to \mathcal{H} \tag{3.27}$$

$$\nu(x) \mapsto \kappa_x \tag{3.28}$$

satisfies this requirement, with

$$\left\langle \nu(x), \nu(x') \right\rangle = \langle \kappa_x, \kappa_{x'} \rangle = \kappa(x, x') \tag{3.29}$$

4

Bibliography