Chapter III
Fenchel Duality

March 11, 2023

1 Subgradients and Convex Functions

3.1.1

A function $f : \mathbb{E} \to (+\infty, +\infty]$ is sublinear if and only if it is positively homogeneous and subadditive. For a sublinear function f, the lineality space $\operatorname{lin} f$ is the largest subspace of $\mathbb{E}$ on which f is linear. Recall that

$$\operatorname{lin} f = \{x \in \mathbb{E} : -f(x) = f(-x)\}.$$

Proof: First suppose that f is sublinear. Then $f(\lambda x + \mu y) \leq \lambda f(x) + \mu f(y)$ for all $x, y \in \mathbb{E}$ and $\lambda, \mu \in \mathbb{R}_+$. Now let $x = y = 0$, then

$$f(0) \leq f(0) + f(0) \Rightarrow 0 \leq f(0).$$

Now let $\lambda = \mu = 0$ and so

$$f(0) \leq 0.$$

So, we have $f(0) = 0$. Now let $y = 0$ and conclude that

$$f(\lambda x) \leq \lambda f(x) \; \forall \lambda \in \mathbb{R}_+.$$

Let $\lambda > 0$ and thus

$$f(x) \leq \frac{1}{\lambda} f(\lambda x) \Rightarrow \lambda f(x) \leq f(\lambda x).$$

So $f(x) = \lambda f(x)$ for all $\lambda \in \mathbb{R}_+$ and thus f is positively homogeneous. Finally let $\lambda = \mu = 1$ and thus $f(x+y) \leq f(x) + f(y)$ and thus f is subadditive.

Conversely, suppose f is subadditive and positively homogeneous. Then $f(\lambda x + \mu y) \leq f(\lambda x) + f(\mu y) = \lambda f(x) + \mu f(y)$. Thus f is sublinear.

Now let $x, y \in \operatorname{lin} f$, then if $\lambda, \mu \in \mathbb{R}_-$, then

$$f(\lambda x + \mu y) = f(-\lambda(-x) - \mu(-y)) \leq -\lambda f(-x) - \mu f(-y) = \lambda f(x) + \mu f(y)$$

3.1.6.

If the function $f : \mathbb{E} \to (+\infty, +\infty]$ is convex and the point $\bar{x}$ lies in $\operatorname{dom} f$, then an element ϕ of $\mathbb{E}$ is a subgradient of f at $\bar{x}$ if and only if it satisfies $\langle \phi, . \rangle \leq f'(\bar{x}; .)$.

Proof:

" $\Rightarrow$ ".

For $d \in \mathbb{E}$ let $\epsilon > 0$ be small enough such that $x_t := \bar{x} + td \in \operatorname{dom} f$ for all $t \in [0, \epsilon]$. Then,

$$\langle \phi, x_t - \bar{x} \rangle \leq f(x_t) - f(\bar{x}) \Rightarrow \langle \phi, d \rangle \leq \frac{f(x + td) - f(\bar{x})}{t} \qquad \forall t \in (0, \epsilon].$$

Now, taking $t \downarrow 0$ completes the proof.

Conversely, suppose that $\langle \phi, . \rangle \leq f'(\bar{x}; .)$ and let $x \in \mathbb{E}$. Then since $t \mapsto \frac{f(\bar{x}+t(x-\bar{x}))-f(\bar{x})}{t}$ is nondecreasing we have $f'(x, x - \bar{x}) \leq f(x) - f(\bar{x})$

$$\langle \phi, x - \bar{x} \rangle \leq f'(x; x - \bar{x}) \leq f(x) - f(\bar{x}).$$

This completes the proof.

3.1.7.

Suppose that the function $p : \mathbb{E} \to (+\infty, +\infty]$ is sublinear and that the point $\bar{x}$ lies in core(dom p). Then the function $q(.) = p'(\bar{x}; .)$ satisfies the conditions

- $q(\lambda \bar{x}) = p(\lambda \bar{x})$.
- $q \leq p$.
- $\operatorname{lin} p + span\{\bar{x}\} \subseteq \operatorname{lin} q$.

Proof:

- $$q(\lambda\bar{x}) = \lim_{t \downarrow 0} \frac{p(\bar{x} + t\lambda\bar{x}) - p(\bar{x})}{t} = \lim_{t \downarrow 0} \frac{(1 + t\lambda)p(\bar{x}) - p(\bar{x})}{t} = \lambda p(\bar{x}).$$

  Note that for small enough t, $1 + t\lambda > 0$.

- For $d \in \mathbb{E}$,

  $$q(d) = \lim_{t \downarrow 0} \frac{p(\bar{x} + td) - p(\bar{x})}{t} \leq \lim_{t \downarrow 0} \frac{p(\bar{x}) + p(td) - p(\bar{x})}{t} = \lim_{t \downarrow 0} \frac{p(td)}{t} = p(d).$$

- Note that $\operatorname{lin} q$ is a linear subspace and thus due to part 1, it suffices to prove $\operatorname{lin} p \subseteq \operatorname{lin} q$. Suppose $d \in \operatorname{lin} p$ and hence $p(-d) + p(d) = 0$. Now we have

  $$q(d) \leq p(d) \Rightarrow -q(d) \geq -p(d) = p(-d) \geq q(-d) \Rightarrow q(-d) \leq -q(d) \leq q(-d) \Rightarrow q(d) + q(-d) = 0.$$

  Note that $-q(d) \leq q(-d)$ holds true since q is sublinear.

3.1.9. Subgradients of maximum eigenvalue

Prove

$$\partial\lambda_1(0) = \{Y \in \mathbb{S}^n_+ : \operatorname{Tr}(Y) = 1\}.$$

Proof: $Y \in \partial\lambda_1(0)$ if and only if $\operatorname{Tr}(XY) \leq \lambda_1(X)$ for all $X \in \mathbb{S}^n$. Let $X = I$ and $X = -I$ respectively to conclude that $\operatorname{Tr}(Y) = 1$. Now from Fan inequality we know

$$\operatorname{Tr}(XY) \leq \lambda(X)^T\lambda(Y) = \sum \lambda_i(X)\lambda_i(Y) \leq \lambda_1(X) \sum \lambda_i(Y) = \lambda_1(X).$$

3.1.6. (Bregman distances)

For a function $\phi : \mathbb{E} \to (\infty, +\infty]$ that is strictly convex and differentiable on $int(\operatorname{dom}\phi))$, define the *Bregman distance* $d_\phi : \operatorname{dom}\phi \times int(\operatorname{dom}\phi) \to \mathbb{R}$ by

$$d_\phi(x, y) = \phi(x) - \phi(y) - \phi'(y)(x - y).$$

- Prove $d_\phi(x, y) \geq 0$ with equality if and only if $x = y$.

- Compute d_ϕ when $\phi(t) = \frac{t^2}{2}$ and when ϕ is the function p defined in Exercise 27.
- Suppose ϕ is three times differentiable. Prove d_ϕ is convex if and only if $-1/\phi''$ is convex on $int(\mathrm{dom}\,\phi)$.

Proof:

- By definition of strictly convex.
- Let $\phi = t^2/2$, then $d_\phi(x,y) = \frac{(x-y)^2}{2}$. Also for the function p from Exercise 27, if $u, v > 0$, then

$$u \log u - u - v \log v + v - \log v(u - v) = u(\log u - \log v) - (u - v).$$

  Note that
$$u(\log u - \log v) - (u - v) \geq 0 \iff u/v \geq e^{1-\frac{v}{u}}.$$

  On the other hand, $e^x - xe$ is a convex function with minimum occurs at $x = 1$, and so $e^x - xe \geq 0$.

  However, if $u = 0$, then

$$d_\phi(0, v) = -v \log v + v - \log v(-v) = v.$$

- Note that the second derivative of d_ϕ can be calculated as the following:

$$\nabla^2 d_\phi(x,y) = \begin{bmatrix} \phi''(x) & -\phi''(y) \\ -\phi''(y) & \phi''(y) + \phi'''(y)(y-x) \end{bmatrix}$$

  Now, due to Schur complement criterion, $\nabla^2 d_\phi(x,y)$ is positive semi-definite if and only if
$$\phi''(x) > 0, \quad \phi''(y) + \phi'''(y)(y-x) - \phi''(y)^2/\phi''(x) > 0$$

3.1.20. Monotonicity of gradients

Suppose that $S \subseteq \mathbb{R}^n$ is open and convex and the function $f : S \subseteq \mathbb{R}$ is differentiable. Prove f is convex if and only if

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \qquad \text{for all } x, y \in S,$$

and f is strictly convex if and only if the above inequality holds strictly whenever $x \neq y$.

Proof: First suppose that f is convex and let $x, y \in S$, then

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle,$$
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle.$$

Summing the above two inequalities we obtain

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0,$$

as desired.

Conversely, suppose that for all $x, y \in S$, $\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq 0$ holds true. We wish to show that f is convex. It suffices to show that $g : t \in [0,1] \to f((1-t)x + ty)$ is convex. For that, we just need to show that $g'(t)$ is non-decreasing. But, $g'(t) = \langle \nabla f((1-t)x + ty), y - x\rangle$. Now let $0 \leq t_1 < t_2 \leq 1$ and notice

$$(t_2 - t_1)(g'(t_2) - g'(t_1)) = \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), y - x\rangle$$
$$= \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), ((1-t_2)x + t_2 y) - ((1-t_1)x + t_1 y)\rangle \geq 0,$$

thus $g'(t)$ is non-decreasing as desired.

Now suppose f is strictly convex and then

$$f(y) > f(x) + \langle \nabla f(x), y - x\rangle,$$
$$f(x) > f(y) + \langle \nabla f(y), x - y\rangle.$$

Summing the above two inequalities we obtain

$$\langle \nabla f(y) - \nabla f(x), y - x\rangle > 0,$$

as desired.

Conversely, suppose that for all $x, y \in S$, $\langle \nabla f(y) - \nabla f(x), y - x\rangle > 0$ holds true. We wish to show that f is strictly convex. It suffices to show that $g : t \in [0,1] \to f((1-t)x+ty)$ is strictly convex. For that, we just need to show that $g'(t)$ strictly increasing. But, $g'(t) = \langle \nabla f((1-t)x + ty), y - x\rangle$. Now let $0 \leq t_1 < t_2 \leq 1$ and notice

$$(t_2 - t_1)(g'(t_2) - g'(t_1)) = \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), y - x\rangle$$
$$= \langle \nabla f((1-t_2)x + t_2 y) - \nabla f((1-t_1)x + t_1 y), ((1-t_2)x + t_2 y) - ((1-t_1)x + t_1 y)\rangle > 0,$$

thus $g'(t)$ is strictly increasing as desired.

3.1.21. The log barrier

Use Exercise 20 (Monotonicity of gradients), Exercise 10 in Section 2.1. and Exercise 8 in Section 1.2 to prove that the function $f : \mathbb{S}^n_{++} \to \mathbb{R}$ defined by $f(X) = -\log\det(X)$ is strictly convex. Deduce the uniqueness of the minimum volume ellipsoid in Section 2.3, Exercise 8, and the matrix completion in Section 2.1, Exercise 12.

Proof: Recall that $\nabla f(X) = -X^{-1}$ and so we should prove for all $X, Y \in \mathbb{S}^n_{++}$ we have

$$\left\langle -X^{-1} + Y^{-1}, X - Y\right\rangle \geq 0 \iff \mathrm{Tr}(XY^{-1}) + \mathrm{Tr}(YX^{-1}) \geq 2n \iff \sum \lambda_i(A) + \sum \frac{1}{\lambda_i(A)} \geq 2n,$$

where $A = XY^{-1}$ and $\lambda_i(A)$ are eigenvalues of A. Note that eigenvalues of A and the positive definite matrix $Y^{-\frac{1}{2}}XY^{-\frac{1}{2}}$ are identical. So since for any $x > 0$ one has $x + \frac{1}{x} \geq 2$, the proof of convexity of f is complete.

Also, note that $X \mapsto ||Xy||^2 - 1$ is convex for any fixed $y \in \mathbb{R}^n$. In fact, we know that $\nabla g(X) = Xyy^T + yy^T X$. So we need to show that for $X, Y \in \mathbb{S}^n_{++}$ we have

$$\langle \nabla g(X) - \nabla g(Y), X - Y\rangle \geq 0 \iff \left\langle (X - Y)yy^T + yy^T(X - Y), X - Y\right\rangle \geq 0.$$

So, we need to prove for $Z \in \mathbb{S}^n$ one has for the positive semi-definite $A = yy^T$

$$\langle ZA + AZ, Z\rangle \geq 0 \iff 2\operatorname{Tr}(ZAZ) \geq 0 \iff 2\operatorname{Tr}((ZA^{\frac{1}{2}})(A^{\frac{1}{2}}Z)) \geq 0,$$

which is immediate as $(ZA^{\frac{1}{2}})(A^{\frac{1}{2}}Z)$ is positive semi-definite.

3.1.22.

Let $f : \mathbb{R}^n \to \mathbb{R}$ with

$$f(x) = \log(\sum_{i=1}^{m} \exp\langle a^i, x\rangle),$$

where $a^1, \cdots, a^m$ are vectors in $\mathbb{R}^n$. Compute the Hessian of f and prove it is positive semi-definite matrix.

Proof: Note that $e^{f(x)} = \sum_{i=1}^m \exp\langle a^i, x\rangle$ and thus

$$e^{f(x)}\nabla f(x) = \sum_{i=1}^{m} \exp(\langle a^i, x\rangle)a^i.$$

So,

$$e^{f(x)}\nabla^2 f(x) + e^{f(x)}\nabla f(x)\nabla^T f(x) = \sum_{i=1}^{m} \exp(\langle a^i, x\rangle)a^i {a^i}^T,$$

and hence,

$$e^{f(x)}\nabla^2 f(x) = \sum_{i=1}^{m} \exp(\langle a^i, x\rangle)a^i {a^i}^T - e^{f(x)}\nabla f(x)\nabla^T f(x).$$

Let $t_i = \exp\langle a^i, x\rangle$, then

$$e^{2f(x)}\nabla^2 f(x) = \left(\sum_{k=1}^{m} t_k\right)\left(\sum_{i=1}^{m} t_i a^i {a^i}^T\right) - \left(\sum_{i=1}^{m} t_i a^i\right)\left(\sum_{i=1}^{m} t_i a^i\right)^T.$$

Now let $\lambda_i = \frac{t_i}{\sum_{i=1}^m t_i}$, then $\lambda_i > 0$ and $\sum_{i=1}^m \lambda_i = 1$. Then

$$\nabla^2 f(x) = \left(\sum_{i=1}^{m} \lambda_i a^i {a^i}^T\right) - \left(\sum_{i=1}^{m} \lambda_i a^i\right)\left(\sum_{i=1}^{m} \lambda_i a^i\right)^T.$$

Now $\nabla^2 f(x) \succeq 0$ if and only if

$$\begin{bmatrix} 1 & \sum_{i=1}^m \lambda_i {a^i}^T \\ \sum_{i=1}^m \lambda_i a^i & \sum_{i=1}^m \lambda_i a^i {a^i}^T \end{bmatrix} \succeq 0.$$

But the above matrix equals to

$$\sum_{i=1}^{m} \lambda_i \begin{bmatrix} 1 & {a^i}^T \\ a^i & a^i {a^i}^T. \end{bmatrix}$$

which is clearly positive semi-definite.

3.1.23

Suppose $f : \mathbb{E} \to (\infty, +\infty]$ is essentially strictly convex, prove all distinct points x and y satisfy $\partial f(x) \cap \partial f(y) = \emptyset$. Deduce that f has at most one minimizer.

Proof: Let $s \in \partial f(x) \cap \partial f(y)$ for some $x, y \in \operatorname{dom} \partial f$. Then $g := f + \langle s, . \rangle$ is an essentially strictly convex that satisfies $0 \in \partial g(x) \cap \partial g(y)$. Thus, without loss of generality, suppose $s = 0$ and thus x and y are minimizer to f. However, since f is convex all the points lying on the line segment $[x, y]$ are also minimizers of f. Thus, $[x, y] \in \operatorname{dom} \partial f$ and this is a contradiction as f is essentially strictly convex on $\operatorname{dom} \partial f$.

3.1.25. Convex matrix functions

Consider a matrix C in $\mathbb{S}^n_+$.

- For matrices $X \in \mathbb{S}^n_{++}$ and D in $\mathbb{S}^n$, use a power series expansion to prove

$$\frac{d^2}{dt^2} \operatorname{Tr}(C(X + tD)^{-1})|_{t=0} \geq 0.$$

- Deduce $X \in \mathbb{S}^n_{++} \mapsto \operatorname{Tr}(CX^{-1})$ is convex.

- Prove similarly the function $X \in \mathbb{S}^n \mapsto \operatorname{Tr}(CX^2)$ and the function $X \in \mathbb{S}^n_+ \mapsto -\operatorname{Tr}(CX^{\frac{1}{2}})$ are convex.

- One version of *Hölder inequality* states, for real $p, q > 1$ satisfying $p^{-1} + q^{-1} = 1$ and functions $u, v : \mathbb{R}_+ \to \mathbb{R}$,

$$\int uv \leq \left(\int |u|^p \right)^{\frac{1}{p}} \left(\int |v|^q \right)^{\frac{1}{q}}$$

  when the right hand side is well-defined. Use this to prove the *gamma function* $\Gamma : \mathbb{R}_{++} \to \mathbb{R}$ given by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

  is log-convex.

- Note that

$$(X+tD)^{-1} = X^{-\frac{1}{2}}(I+tX^{-\frac{1}{2}}DX^{-\frac{1}{2}})^{-1}X^{-\frac{1}{2}} = X^{-\frac{1}{2}}(I-tX^{-\frac{1}{2}}DX^{-\frac{1}{2}}+t^2(X^{-\frac{1}{2}}DX^{-\frac{1}{2}})^2+O(t^3))X^{-\frac{1}{2}}.$$

  Thus,

$$\operatorname{Tr}(C(X+tD)^{-1}) = \operatorname{Tr}(CX^{-1}) - t\operatorname{Tr}(X^{-1}CX^{-1}D) + t^2 \operatorname{Tr}(CX^{-1}DX^{-1}DX^{-1}) + O(t^3).$$

  However,

$$\operatorname{Tr}(CX^{-1}DX^{-1}X^{-1}DX^{-1}) = \operatorname{Tr}(C^{\frac{1}{2}}X^{-1}DX^{-\frac{1}{2}}X^{-\frac{1}{2}}DX^{-1}C^{\frac{1}{2}}) = \operatorname{Tr}(AA^T) \geq 0,$$

  where $A := C^{\frac{1}{2}}X^{-1}DX^{-\frac{1}{2}}$. Note that AA^T is positive semidefinite and thus has non-negative trace.

- Let $Y \in \mathbb{S}^n_{++}$ and let $g : t \in [0,1] \to \mathrm{Tr}(C(X + t(Y - X))^{-1})$. Then due to part 1, g is convex and thus so is the function $X \in \mathbb{S}^n_{++} \mapsto \mathrm{Tr}(CX^{-1})$.

- Note that
$$\mathrm{Tr}(C(X + tD)^2) = \mathrm{Tr}(C(X^2 + t(DX + XD) + t^2 D^2)).$$
Thus,
$$\frac{d^2}{dt^2} \mathrm{Tr}(C(X + tD)^2)|_{t=0} = 2\,\mathrm{Tr}(CD^2) = 2\,\mathrm{Tr}(C^{\frac{1}{2}} DDC^{\frac{1}{2}}) \geq 0.$$
Also, for $X \in \mathbb{S}^n_{++}$,
$$\mathrm{Tr}(C(X + tY)^{\frac{1}{2}}) =$$

3.1.26. Log-convexity

Given a convex set $C \subseteq \mathbb{E}$, we say that a function $f : C \to \mathbb{R}_{++}$ is *log-convex* if $\log f(.)$ is convex.

- Prove any log-convex function is convex.

- If a polynomial $p : \mathbb{R} \to \mathbb{R}$ has all real roots, prove $1/p$ is log-convex on any interval on which p is strictly positive.

Proof:

- Suppose that $f : C \to \mathbb{R}_{++}$ is log-convex. Then let $x, y \in C$ and $\lambda \in (0,1)$, we need to show

  $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ or equivalently $\log(f(\lambda x + (1-\lambda)y)) \leq \log(\lambda f(x) + (1-\lambda) f(y))$.

  However, since f is log-convex, $\log(f(\lambda x + (1 - \lambda)y)) \leq \lambda \log f(x) + (1 - \lambda) \log f(y)$. Now notice that log is concave and so
$$\lambda \log f(x) + (1 - \lambda) \log f(y) \leq \log(\lambda f(x) + (1 - \lambda) f(y)).$$
  Thus the proof is complete.

- We show first that $p^{\frac{1}{n}}$ is concave. Let $a, a_1, \cdots, a_n$ be such that $p(x) = a \prod_{i=1}^n (x - a_i)$ and let $I = (\alpha, \beta)$ be any interval on which p is positive. Note that
$$\frac{d}{dt} p(t)^{\frac{1}{n}} = \frac{1}{n} p(t)^{\frac{1}{n}-1} p'(t),$$
  and so,
$$\begin{aligned}\frac{d^2}{dt^2} p(t)^{\frac{1}{n}} &= \frac{1}{n}(\frac{1}{n} - 1) p(t)^{\frac{1}{n}-2} p'(t)^2 + \frac{1}{n} p(t)^{\frac{1}{n}-1} p''(t) \\ &= \frac{1}{n^2} p(t)^{\frac{1}{n}-2} \left[(1 - n) p'(t)^2 + n p''(t) p(t)\right].\end{aligned}$$

On the other hand, for $t \neq a_i$,

$$p'(t) = a\left(\sum_{i=1}^{n} \frac{1}{t-a_i}\right) p(t),$$

and,

$$p''(t) = a\left(-\sum_{i=1}^{n} \frac{1}{(t-a_i)^2}\right) p(t) + a\left(\sum_{i=1}^{n} \frac{1}{t-a_i}\right)^2 p(t).$$

Let $r_i := (t-a_i)^{-1}$ and thus we have

$$\frac{n^2}{p(t)^{2-\frac{1}{n}}} \frac{d^2}{dt^2} p(t)^{\frac{1}{n}} = a^2 \left[(1-n)(\sum_{i=1}^{n} r_i)^2 + n(-\sum_{i=1}^{n} r_i^2 + (\sum_{i=1}^{n} r_i)^2)\right] p(t)^2.$$

However,

$$(1-n)(\sum_{i=1}^{n} r_i)^2 + n(-\sum_{i=1}^{n} r_i^2 + (\sum_{i=1}^{n} r_i)^2) = (\sum_{i=1}^{n} r_i)^2 - n\sum_{i=1}^{n} r_i^2 \leq 0,$$

due to Cauchy-Schwartz inequality. Thus $p^{\frac{1}{n}}$ is concave and so is $\log p^{\frac{1}{n}}$ since log is non-decreasing and concave. Thus $\log p$ is concave,

- Note that we have the following as a special case of *Hölder inequality*

$$\log \int_0^\infty u^\lambda v^{1-\lambda} dt \leq \lambda \log \int_0^\infty u dt + (1-\lambda) \log \int_0^\infty v dt,$$

where $u, v : \mathbb{R}_{++} \to \mathbb{R}_{++}$ and $\lambda \in (0,1)$. Now suppose $x, y \in \mathbb{R}_{++}$ and let

$$u(t) = t^x e^{-t}, v(t) = t^y e^{-t}.$$

Then the convexity of the Gamma function follows immediately.

---

2 The Value Function

3.2.1 Lagrangian sufficient conditions.

Prove the Lagrangian sufficient conditions:

Suppose $\bar{\lambda}$ is a Lagrangian multiplier for a feasible solution $\bar{x}$ such that $\bar{x}$ minimizes $L(., \bar{\lambda})$ over $\mathbb{E}$. Then $\bar{x}$ is an optimal solution.

Proof: Note that since $\bar{\lambda}$ is a Lagrangian multiplier we have $L(\bar{x}, \bar{\lambda}) = f(\bar{x})$. However,

$$f(\bar{x}) = L(\bar{x}, \bar{\lambda}) \leq L(y, \bar{\lambda}) \leq f(y) \quad \forall \text{ feasible solution } y.$$

Thus $\bar{x}$ is an optimal solution.

3.2.2.

Use the Lagrangian sufficient conditions to the following problem s.

- a)

$$\begin{aligned} &\inf x_1^2 + x_2^2 - 6x_1 - 2x_2 + 10 \\ &\text{subject to} \quad 2x_1 + x_2 - 2 \le 0 \\ &x_2 - 1 \le 0 \\ &x \in \mathbb{R}^2. \end{aligned}$$

Proof: Let $y_1 = x_1 - 3, y_2 = x_2 - 1$ then

$$\begin{aligned} &\inf y_1^2 + y_2^2 \\ &\text{subject to} \quad 2y_1 + y_2 + 5 \le 0 \\ &y_2 \le 0 \\ &y \in \mathbb{R}^2. \end{aligned}$$

If $y_2 = 0$, then the obvious minimum will be $25/4$. So suppose that $y_2 < 0$ and so $\bar{\lambda}_2 = 0$. Now let $\bar{\lambda}_1 = 2$, then

$$L(y, \bar{\lambda}) = y_1^2 + y_2^2 + 4y_1 + 2y_2 + 10 = (y_1 + 2)^2 + (y_2 + 1)^2 + 5.$$

So, $y = -(2, 1)$ minimizes $L(., \bar{\lambda})$. Since y is feasible and its objective value is $5 < 25/4$, the optimal value is 5 with the optimal solution $(1, 0)$.

-

$$\begin{aligned} &\inf -2x_1 + x_2 \\ &\text{subject to} \quad x_1^2 - x_2 \le 0 \\ &x_2 - 4 \le 0 \\ &x \in \mathbb{R}^2. \end{aligned}$$

Proof: If $x_2 = 4$ then the obvious minimum will be 0. So suppose that $x_2 < 4$ and then $\bar{\lambda}_2 = 0$,

$$L(x, \bar{\lambda}) = -2x_1 + x_2 + \bar{\lambda}_1(x_1^2 - x_2).$$

Now let $\bar{\lambda}_1 = 1$ and let $\bar{x} = (1, 1)$. Note that since

$$L(x, \bar{\lambda}) = x_1^2 - 2x_1 \ge -1,$$

and since $L(\bar{x}, \bar{\lambda}) = -1$. Thus $\bar{\lambda}$ is a Lagrangian multiplier $\bar{x}$ and since $\bar{x}$ minimizes $L(., \bar{\lambda})$, we conclude the optimum value is -1.

-

$$\begin{aligned} &\inf x_1 + \frac{2}{x_2} \\ &\text{subject to} - x_2 + \frac{1}{2} \le 0 \\ &-x_1 + x_2^2 \le 0. \end{aligned}$$

Proof: If $x_2 = \frac{1}{2}$ then the obvious inf will be $4 + \frac{1}{4}$. Now suppose that $x_2 > \frac{1}{2}$. Then $\bar{\lambda}_1 = 0$ and

$$L(x, \bar{\lambda}) = x_1 + \frac{2}{x_2} + \bar{\lambda}_2(-x_1 + x_2^2).$$

Now let $\bar{\lambda}_2 = 1$ then

$$L(x, \bar{\lambda}) = x_1 + \frac{2}{x_2} + \bar{\lambda}_2(-x_1 + x_2^2) = \frac{2}{x_2} + x_2^2 \geq 3.$$

Note that $\frac{2}{x_2} + x_2^2 = \frac{1}{x_2} + \frac{1}{x_2} + x_2^2 \geq 3$. Now let $\bar{x} = (1, 1)$, then the objective value equals to 3 and thus the optimum value is 3.

---

3 The Fenchel Conjugate

3.1.7. Quadratics

For all matrices A in $\mathbb{S}^n_{++}$, prove the function $x \in \mathbb{R}^n \to x^T Ax/2$ is convex and calculate its conjugate. Use the order preserving property to the conjugacy operation to prove

$$A \succeq B \iff B^{-1} \succeq A^{-1} \text{ for all } A \text{ and } B \text{ in } \mathbb{S}^n_{++}.$$

Proof: Note that for $f(x) = \frac{1}{2}x^T Ax$ we have $\nabla^2 f = A$ and thus f is convex. However, $\nabla f(x) = Ax$ and hence $\sup_x \langle x, y\rangle - f(x)$ is realized at $x = A^{-1}y$. Thus, $f^*(x) = \frac{1}{2}x^T A^{-1}x$. Now if $A \succeq B$ then $f_A \geq f_B$ and hence $f^*_B \geq f^*_A$.

3.1.3.
Verify the conjugates of the log barrier Ib and Id claimed in the text.
Proof: Let $f(x) = -\log x$, then

$$f^*(y) = \sup_{x \in \mathbb{E}} \langle x, y\rangle + \log x.$$

However, $(\langle x, y\rangle + \log x)'' = \frac{-1}{x^2} < 0$. Now since $y + \frac{1}{x}$. Hence, $f^*(x) = -1 + f(-x)$.

Now since $Id^*(X) = \sup_{Y \succ 0}\langle X, Y\rangle + \log\det Y$. Then if $X \not\prec 0$ then $\langle X, Y\rangle + \log\det Y$ is unbounded above as for $Xx = \lambda x$ with $\lambda \geq 0$ we have $\langle X, xx^T + I\rangle + \log\det(xx^T + I) = \lambda||x||^2 + 1 + ||x||^2 + \text{Tr}(X)$. Now let $||x|| \to +\infty$. Hence, $Id^*(X) = \infty$. So suppose that $X \prec 0$. Then since $\langle X, Y\rangle + \log\det(Y)$ is concave with gradient $Y^{-1} + X$ or $Y = -X^{-1}$. Now

$$\langle -X^{-1}, X\rangle + \log\det(-X^{-1}) = -n - \log\det(-X).$$

3.3.4 Self Conjugacy
Consider functions $f : \mathbb{E} \to (\infty, \infty]$.

- Prove $f^* = f$ if and only if $f(x) = ||x||^2/2$ for all points x in $\mathbb{E}$.
- Find two distinct functions f satisfying $f(-x) = f^*(x)$ for all x in $\mathbb{E}$.

Proof:

- Suppose that $f(x) = \frac{1}{2}||x||^2$. Then $f^*(x) = \sup\langle y, x\rangle - \frac{1}{2}||x||^2$. Now note that $\langle y, x\rangle - \frac{1}{2}||x||^2$ is strictly concave and $\nabla(\langle y, x\rangle - \frac{1}{2}||x||^2) = y - x$. Thus, $f^*(y) = \frac{1}{2}||y||^2$. Conversely, suppose that $f = f^*$. Then since f^* is convex, f is convex as well. If $\operatorname{dom} f = \emptyset$, then $f^* \equiv -\infty$ which is a contradiction as $f = f^*$ and f never takes the value $-\infty$. So $\operatorname{dom} f \neq \emptyset$. Let $x \in \operatorname{dom} f$, then $f(x) + f(y) \geq \langle x, y\rangle$ for all $y \in \operatorname{dom} f$. Also, if $y \in \partial f(x)$, then $f(x)+f(y) = \langle x, y\rangle$ and hence $y \in \partial f^*(x)$ and thus $y \in \partial f(x)$. Now since $f(x)+f(y) \ \geq \langle x, y\rangle$ we obtain $f(x) \geq \frac{1}{2}\langle x, x\rangle$. Now let $x \in \mathbb{E}$ and $y \in \partial f(x)$ then $\langle x, y\rangle \geq \frac{1}{2}(||x||^2 + ||y||^2)$. Hence, $f(x) = \frac{1}{2}||x||^2$. This completes the proof.

- Let $f(x) = -\log x$. Note that $f^*(y) = -1\log(-y) = -1 + f(-y)$. Thus, $f^*(x) = -1 + f(-x)$. Let $g(x) = f(x) - \frac{1}{2}$. Then $g^*(x) = g(-x)$.

Question 7. Maximum entropy example

- Prove the function g defined by

$$g(z) = \inf_{x\in\mathbb{R}^{m+1}} \left\{ \sum_i \exp^*(x_i) : \sum_i x_i = 1, \sum_i x_i a^i = z \right\}$$

  is convex.

- For any point $y \in \mathbb{R}^{m+1}$, prove

$$g^*(y) = \sup_{x\in\mathbb{R}^{m+1}} \left\{ \sum_i (x_i\langle a^i, y\rangle - \exp^*(x_i)) : \sum_i x_i = 1 \right\}.$$

- Apply Exercise 27 in Section 3.1 to deduce the conjugacy formula 3.3.2.

- Compute the conjugate of the function of $x \in \mathbb{R}^{m+1}$,

$$\begin{cases} \sum_i \exp^*(x_i) & \text{if} \quad \sum_i x_i = 1 \\ +\infty & \text{otherwise.} \end{cases}$$

Proof:

- Let $\epsilon > 0$ be arbitrary and fix $\lambda \in [0, 1]$. We show for all $z, z' \in \mathbb{R}^m$

$$g(\lambda z + (1 - \lambda)z') \leq \lambda g(z) + (1 - \lambda)g(z') + 2\epsilon.$$

  Let $\tilde{x}_i$ be such that

$$\sum \exp^*(\tilde{x}_i) \leq g(z) + \epsilon.$$

Similarly, let $\tilde{y}_i$ be such that

$$\sum \exp^*(\tilde{y}_i) \leq g(z') + \epsilon.$$

Let $\tilde{z}_i = \lambda\tilde{x}_i + (1-\lambda)\tilde{y}_i$, so

$$g(\tilde{z}_i) \leq \sum \exp^*(\tilde{z}_i) \leq \lambda g(z) + (1-\lambda)g(z') + 2\epsilon.$$

3.3.20. Pointed cones and bases

Consider a closed convex cone K in $\mathbb{E}$. A base for K is a convex set C with $0 \notin cl(C)$ and $K = \mathbb{R}_+C$. Prove the following properties are equivalent

- (a) K is pointed, i.e. $K \cap -K = \{0\}$.
- (b) $cl(K^\circ - K^\circ) = \mathbb{E}$.
- (c) $K^\circ - K^\circ = \mathbb{E}$.
- (d) K° has non-empty interior.
- (e) There exists a vector y in $\mathbb{E}$ and real $\epsilon > 0$ with $\langle y, x\rangle \geq \epsilon||x||$ for all points x in K.
- (f) K has a bounded base.

Proof:

- $(a) \Rightarrow (b)$. Suppose $cl(K^\circ - K^\circ) \neq \mathbb{E}$ and let $x \in \mathbb{E}\backslash cl(K^\circ - K^\circ)$. Then due to Hyperplane separation theorem, there exists $0 \neq \phi \in \mathbb{E}$ and $\alpha \in \mathbb{R}$ such that

  $$\langle x, \phi\rangle > \alpha \geq \langle z_1 - z_2, \phi\rangle \quad \forall z_1, z_2 \in K^\circ.$$

  Thus for all $z \in K^\circ$, $\langle z, \phi\rangle$ is bounded above and thus $\langle z, \phi\rangle \leq 0$ as K° is a cone. On the other hand, due to the above equation, $\langle -z, \phi\rangle$ is bounded as well for all $z \in K^\circ$ and so $\langle z, -\phi\rangle \leq 0$ for all $z \in K^\circ$. So, ϕ and $-\phi$ both belongs to $K^{\circ\circ}$ which equals to K. Thus, since K is pointed we conclude that $\phi = 0$. This contradiction completes the proof.

- $(b) \Rightarrow (c)$. Note that $K^\circ - K^\circ$ is a subspace. In fact, it is clearly a convex cone and since $-(K^\circ - K^\circ) = K^\circ - K^\circ$ and it contains 0, it is also a subspace. Thus, since every subspace is closed, $K^\circ - K^\circ = cl(K^\circ - K^\circ)$.

- $(c) \Rightarrow (d)$. We already now that every nonempty convex set in $\mathbb{E}$ has a nontrivial relative interior. Now note that

  $$\text{aff}(K^\circ) = \text{aff}(-K^\circ) = \text{aff}(K^\circ - K^\circ).$$

  Thus, since K° has a nonempty interior and its affine hull is $\mathbb{E}$, we conclude that K° has a nonempty interior.

- $(d) \Rightarrow (e)$. Let $y \in (K^\circ)$, and $\epsilon > 0$ such that $y + td \in K^\circ$ for all $t \in [-\epsilon, \epsilon]$ and any $d \in \mathbb{E}$ with $||d|| = 1$. So, we have

$$\langle y + td, x\rangle \leq 0 \; \forall t \in [-\epsilon, \epsilon], \forall ||d|| = 1, \forall x \in K,$$

  or equivalently for $0 \neq x \in K$,

$$|t\langle d, x\rangle| \leq \langle -y, x\rangle \Rightarrow |t\langle d, \frac{x}{||x||}\rangle| \leq \langle -y, \frac{x}{||x||}\rangle.$$

  Now let $d = \frac{x}{||x||}$ in above, and let $t = \epsilon$ we realize that

$$\epsilon \leq \langle -y, \frac{x}{||x||}\rangle.$$

  Thus, $\epsilon||x|| \leq \langle -y, x\rangle$ holds true for all $x \in K$.

- $(e) \Rightarrow (f)$. Now suppose $y \in \mathbb{E}$ and $\epsilon > 0$ are such that

$$\langle y, x\rangle \geq \epsilon||x|| \; \forall x \in K.$$

  Define

$$C = \{x \in K : \langle x, y\rangle = 1\}.$$

  First note that C is bounded, as if $x \in C$, then $\epsilon||x|| \leq 1$ and so $||x|| \leq \frac{1}{\epsilon}$. Also, note that clearly C is closed and also $0 \notin C$. Last, note that $\mathbb{R}_+ C = K$ as in fact for $0 \neq x \in K$, $\langle y, x\rangle \geq \epsilon||x|| > 0$ and thus $\langle y, x\rangle > 0$ and so there exists $\lambda > 0$ such that $\langle y, \lambda x\rangle = 1$. Thus, $\lambda x \in C$.

- $(f) \Rightarrow (a)$. Let C be a bounded base for K and suppose $a \in K \cap -K$. Then there exists $\lambda, \mu \in \mathbb{R}_+$ such that $a = \lambda c_1 = -\mu c_2$ for some $c_1, c_2 \in C$. Now, if $a \neq 0$ then λ and μ are both nonzero and thus $0 = \frac{\lambda}{\lambda+\mu} c_1 + \frac{\mu}{\mu+\lambda} c_2 \in C$ as C is convex. But this contradicts the fact that $0 \notin cl(C)$. So, K is pointed.