Unit 3: Fenchel and Lagrange Duality

Outline

1. Statement of Lagrange Duality
2. Examples
3. Subgradients (properties of nonsmooth, convex functions)
4. Fenchel Duality
5. Fenchel Biconjugates (more properties of nonsmooth convex functions)
6. Proof of Lagrange Duality

1. Statement of Lagrange Duality

We previously showed Lagrange multipliers are necessory for local minimizers. Here we will greatly expand this idea to describe a dual optimization problem over all multipliers providing lower bounds on optimality (like we had with LPs).

Consider the functionally constrained problem (our primal problem)

$$p^* = \begin{cases} \min & f(x) \\ \text{s.t.} & g_i(x) \le 0 \quad \forall i = 1 \dots m \end{cases}$$

for $f, g_1, \dots g_m : E \to \mathbb{R} \cup \{\pm\infty\}$.

Recall we defined the associated Lagrangian as

$$L(x;\lambda) = f(x) + \lambda^T g(x)$$

$$\hookrightarrow \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}$$

for $\lambda \in \mathbb{R}^m$.

Claim: $\displaystyle \sup_{\lambda \geq 0} L(x;\lambda) = \begin{cases} f(x) & \text{if } g(x) \leq 0 \ \forall i \\ +\infty & \text{otherwise} \end{cases}$ for all $x \in E$.

Proof. If x is feasible, then $g(x) \leq 0$. ← elementwise

$\Rightarrow \lambda^T g(x) \leq 0 \quad \forall \lambda \geq 0$ since $\sum \lambda_i g_i(x)$ has all nonpositive summands.

$\Rightarrow L(x;\lambda) \leq f(x) \quad \forall \lambda \geq 0$

This is attained at $L(x;0) = f(x)$,

so $\displaystyle \sup_{\lambda \geq 0} L(x;\lambda) = f(x)$.

If x is infeasible, some $g_i(x) > 0$.

$\Rightarrow$ Taking $\lambda_i \to \infty$ gives $L(x;\lambda) \to \infty$.

$\Rightarrow \displaystyle \sup_{\lambda \geq 0} L(x;\lambda) = +\infty$. □

As a result, we immediately can describe our primal problem as

Lagrange Primal Problem

$$\boxed{p^* = \inf_{x \in E} \sup_{\lambda \geq 0} L(x;\lambda).}$$

Then the dual optimization problem is given by

Lagrange Dual Problem

$$\boxed{d^* = \sup_{\lambda \geq 0} \inf_{x \in E} L(x;\lambda).}$$

This inner function of λ, $\Phi(\lambda) = \inf_{x\in E} L(x;\lambda)$, is the dual function.

Theorem (Weak Duality)

For any $f, g_1, \dots g_m : E \to \mathbb{R} \cup \{\pm\infty\}$, $p^* \geq d^*$.

[Note none of the inf and sup defing p^* and d^* need be attained.]

Proof. Let $\lambda^{(i)} \geq 0$ attain d^* in the limit: $\lim_{i\to\infty} \Phi(\lambda^{(i)}) = d^*$.

Consider any feasible $\bar{x}$. Then

$$\begin{aligned} f(\bar{x}) &= \sup_{\lambda\geq 0} L(\bar{x};\lambda) && \text{by previous claim} \\ &\geq \limsup_{i\to\infty} L(\bar{x};\lambda^{(i)}) && \text{restricting to a smaller sup} \\ &\geq \limsup_{i\to\infty} \inf_{x\in E} L(x;\lambda^{(i)}) && \text{taking worst case over } x \\ &= d^*. \end{aligned}$$

□

To derive a strong duality guarantee ($p^* = d^*$), we will need to assume additional structure. Following the idea of HW1Q4, consider

$$v(\Delta) = \begin{cases} \min f(x) \\ \text{s.t. } g_i(x) \leq \Delta_i \quad \forall i=1\dots m. \end{cases}$$

(for $\Delta \in \mathbb{R}^m$) ↑ called the value function

Theorem (Lagrange Strong Duality)

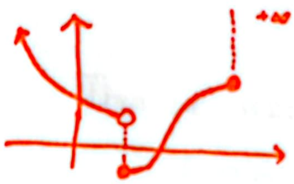
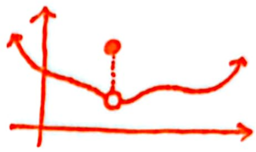
Suppose $v : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is convex with $v(0) = p^* \in \mathbb{R}$.

Then $p^* = d^*$ if and only if $v(\cdot)$ is lower semicontinuous at $\Delta = 0$.

[A function f is <u>lower semicontinuous at x</u> if

$$\liminf_{\tilde{x} \to x} f(\tilde{x}) \geq f(x).$$

Continuity would require the limit equal f, rather than liminf, $\geq$.

For example,  . For non-example,  .]

If all of $f, g_1, \dots g_m$ are convex, then v is convex.
[Proof is left as an exercise.]

In this case, if there exists a Slater point (i.e. $g_i(x_s) < 0 \ \forall i$), then v will be lower semicontinuous at 0.
[Proof will come at the end of this unit, once we have built theory for $v(\cdot)$.]

In these cases, we can strengthen our result with existence:

Theorem (Lagrange Strong Duality with Existence)

If $f, g_1, \dots g_m$ are convex and some x_s has $g_i(x_s) < 0 \ \forall i$, then $p^* = d^*$ and d^* is attained by some λ^* whenever finite.

2. Examples

Linear Programming Consider the problem $\begin{cases} \min -b^T y \\ \text{s.t. } A^T y - c \le 0 \end{cases}$

(so $f(y) = -b^T y$, $g_i(y) = A_i^T y - c_i$).

Then the dual function is

$$\begin{aligned}\Phi(\lambda) &= \min_y \ L(y;\lambda) \\ &= \min_y \ -b^T y + \lambda^T(A^T y - c) \\ &= \begin{cases} -c^T\lambda & \text{if } A\lambda = b \\ -\infty & \text{otherwise.} \end{cases}\end{aligned}$$

in this case, y disappears.

in this case we can pick $y = \eta(A\lambda - b)$ as $\eta \to +\infty$.

Thus we recover our LP duality as

$$\begin{cases} \min -b^T y \\ \text{s.t. } A^T y \le c \end{cases} = \begin{cases} \max -c^T\lambda \\ \text{s.t. } A\lambda = b \\ \qquad \lambda \ge 0. \end{cases}$$

For LPs, the value function is piecewise linear on its domain, which one can show implies it must be lower semicontinuous. Hence LP duality holds as long as one of the problems has a nonempty feasible region (just like we proven in week 2 via simplex).

[So once we prove Lagrange Duality, you can replace the combinatorial, simplex proof of LP Duality with an analysis style proof, never mentioning things like degeneracy or pivoting.]

Quadratic Programming

Consider the quadratic program $\begin{cases} \min & c^T x + \frac{1}{2} x^T H x \\ \text{s.t.} & Ax \le b \end{cases}$

Then the Lagrangian is given by

$$L(x;\lambda) = c^T x + \tfrac{1}{2} x^T H x + \lambda^T (Ax - b)$$

The dual function is given by minimizing this over x.

For ease, lets assume $H \succ 0$ (i.e. H is positive definite and so invertible).

Then the dual function has a closed form ...

$$\begin{aligned} \Phi(\lambda) &= \inf_x L(x;\lambda) \\ &= \inf_x \tfrac{1}{2} x^T H x + (c + A^T\lambda)^T x - \lambda^T b \\ &= -\tfrac{1}{2}(c + A^T\lambda)^T H^{-1} (c + A^T\lambda) - \lambda^T b \end{aligned}$$

you could compute this by completing the square, or by finding the unique stationary pt of the given quadratic.

Hence the dual of a (strictly positive definite) quadratic program is the quadratic program ...

$$\sup_{\lambda \ge 0} \Phi(\lambda) = \begin{cases} \max & -\frac{1}{2}(c + A^T\lambda)^T H^{-1} (c + A^T\lambda) - b^T\lambda \\ \text{s.t.} & \lambda \ge 0. \end{cases}$$

[HW2 had you play with a proposed SDP dual. You can similarly compute that dual problem via minimizing the Lagrangian.]

Norm Optimization

This duality is the same as the classic duality you may be familiar with from linear algebra. Recall for a given norm $\|\cdot\|: E \to \mathbb{R}$, the dual norm is

$$\|v\|_* = \sup_{\|x\| \le 1} \langle v, x \rangle .$$

[For $E = \mathbb{R}^n$, the most classic examples are p-norms $\|x\|_p = (\sum |x_i|^p)^{1/p}$. (The dual of a p-norm is the q-norm with $\frac{1}{p} + \frac{1}{q} = 1$.)

For $E = \mathbb{R}^{n \times m}$, a classic example is the Schatten p-norm, which just takes the p-norm of the singulor values, still dual to the Schatten q-norm with $\frac{1}{p} + \frac{1}{q} = 1$.

3D printed versions of these dualities are on canvas and in my office.]

Consider the norm minimization (for a generic norm $\|\cdot\|$)

$$\begin{cases} \min & \|x\| \\ \text{s.t.} & Ax \le b. \end{cases}$$

Then the Lagrangian is $L(x;\lambda) = \|x\| + \langle \lambda, (Ax - b) \rangle$

and the dual function is

$$\Phi(\lambda) = \inf_x \|x\| + \langle (A^*\lambda), x \rangle - \langle b, \lambda \rangle$$

$$= \begin{cases} -\langle b, \lambda \rangle & \text{if } \|A^*\lambda\|_* \le 1 \\ -\infty & \text{otherwise.} \end{cases}$$

So the dual problem is

$$\begin{cases} \max & -\langle b, \lambda \rangle \\ \text{s.t.} & \|A^*\lambda\|_* \le 1 \\ & \lambda \ge 0. \end{cases}$$

3. Subgradients (and properties of nonsmooth functions)

Recall HW1Q4 had an example where the value function was non differentiable. So we cannot rely on gradients.

Our focus here will be on functions $f: E \to (-\infty, \infty]$ that are

<u>convex</u> : $\forall_{\substack{x,y \in E \\ \lambda \in [0,1]}} \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$

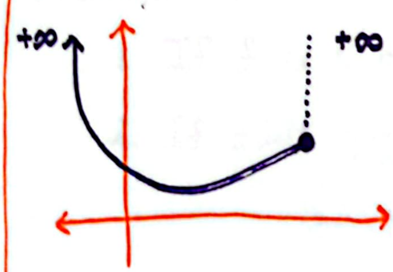
(ruling out local minima)

<u>proper</u> : $\mathrm{dom}\, f = \{x \mid f(x) \neq \infty\} \neq \emptyset$. (feasible somewhere)

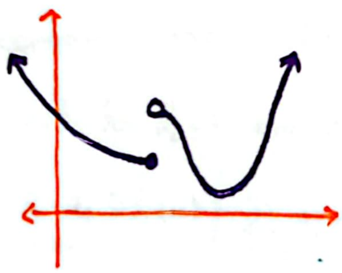
<u>closed</u> : f is lower semicontinuous (regularity condition)

(i.e. $\forall x \in E \quad \liminf_{x' \to x} f(x') \geq f(x)$).

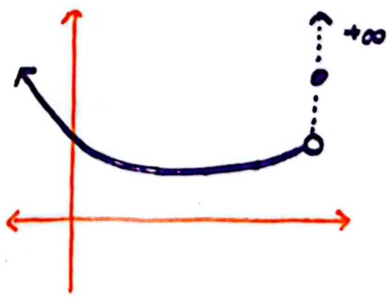
A proper, closed, convex example



A proper, closed, but not convex example



A proper, convex, but not closed example



(As a closed, convex, but not proper example, consider $f(x) = +\infty$ everywhere.)

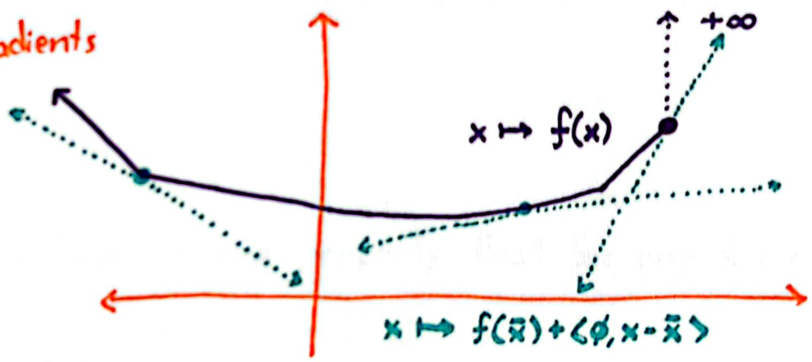
Definition We say $\phi \in E$ is a <u>subgradient</u> of f at $\bar{x} \in E$ if

$$\forall x \in E \quad f(x) \geq f(\bar{x}) + \langle \phi, x - \bar{x} \rangle .$$

(one can view this as providing a one-sided linearization, flipping the inequality above gives sup-gradients.)

The set of all subgradients, called the subdifferential is denoted by $\partial f(\bar{x}) = \{ \phi \mid \forall x \; f(x) \geq f(\bar{x}) + \langle \phi, x - \bar{x} \rangle \}$.

Some example subgradients



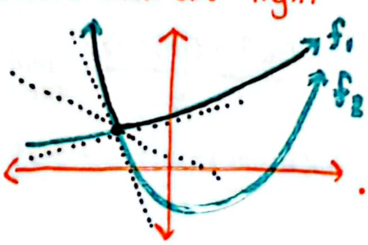
A few examples computing subgradients

1. If f is C^1 and convex, then $\partial f(x) = \{ \nabla f(x) \}$.
2. If $f(x) = \max_{i=1 \ldots m} \{ f_i(x) \}$ for $f_i \in C^1$ and convex, then

$$\partial f(x) = \Big\{ \sum_{i \in I(x)} \lambda_i \nabla f_i(x) \;\Big|\; \sum_{i \in I(x)} \lambda_i = 1,\; \lambda \geq 0 \Big\}$$

where $I(x) = \{ i \mid f_i(x) = f(x) \}$.

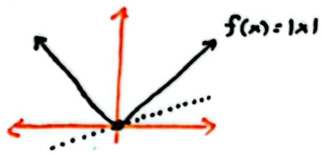
(that is, any combination of gradients of components that are "tight" gives a subgradient of the max. Three subgradients shown dotted.



.)

As a result, the absolute value $|x| = \max\{x, -x\}$ has subgradient

$$\partial|\cdot|(x) = \begin{cases} +1 & \text{if } x>0 \\ [-1,1] & \text{if } x=0 \\ -1 & \text{if } x<0. \end{cases}$$



3. If $f(x) = \|x\|$ for a generic norm $\|\cdot\|$, (note norms are never differentiable at 0.)

$$\phi \in \partial\|\cdot\|(0) \Leftrightarrow \forall_{x\in E} \; \|x\| \geq 0 + \langle \phi, x-0 \rangle$$

$$\Leftrightarrow \|\phi\|_* \leq 1. \quad \text{(using } \|\cdot\|_* \text{ as the dual norm.)}$$

Hence $\partial\|\cdot\|(0) = \{\phi \mid \|\phi\|_* \leq 1\}$, so the subgradients of a norm give the dual norm ball.

Recall convex functions had the nice property that for any $x, v \in E$

$$t \longmapsto \frac{f(x+tv) - f(x)}{t}$$

is non-decreasing on $t \in \mathbb{R}$. This gives the following characterization of subgradients in relation to directional derivatives...

<u>Lemma</u> $\phi \in \partial f(x)$ if and only if $\langle \phi, v \rangle \leq f'(x;v) \;\; \forall_{v\in E}$.

[Recall $f'(x;v) = \lim_{t\searrow 0} \dfrac{f(x+tv)-f(x)}{t}$ is the directional derivative in direction v.]

Proof ($\Rightarrow$). Let $\phi \in \partial f(x)$. Then $f'(x,v) = \lim_{t\searrow 0} \dfrac{f(x+tv) - f(x)}{t}$ (by definition of f')

$$\geq \frac{f(x) + \langle \phi, x+tv-x \rangle - f(x)}{t} \quad \text{(by definition of } \partial f\text{)}$$

$$= \frac{\langle \phi, tv \rangle}{t} \quad \text{(Simplifying)}$$

$$= \langle \phi, v \rangle . \quad \text{(Simplifying)}$$

Hence $f'(x; v) \geq \langle \phi, v \rangle$ for any $v \in E$.

($\Leftarrow$) Consider any $x' \in E$ and let $v = x' - x$ and $t = 1$.

Then
$$f(x') - f(x) = \frac{f(x+tv) - f(x)}{t} \quad \text{(by our choice of } v \text{ and } t\text{)}$$
$$\geq f'(x; v) \quad \text{(since convexity makes this nondecreasing in } t.)$$
$$\geq \langle \phi, v \rangle \quad \text{(by assumption)}$$
$$\Rightarrow f(x') \geq f(x) + \langle \phi, x' - x \rangle \quad \forall x' \in E$$
$$\Rightarrow \phi \in \partial f(x).$$

$\square$

Subgradients also provide an alternative to gradients when giving optimality conditions. Almost by definition having...

Lemma $\bar{x} \in E$ is a global (unconstrained) minimizer of f if and only if $0 \in \partial f(\bar{x})$.

[Note this result does not need convexity to hold.]

Proof. $0 \in \partial f(\bar{x}) \Leftrightarrow \forall x \in E \quad f(x) \geq f(\bar{x}) + \langle 0, x - \bar{x} \rangle$
$$= f(\bar{x}).$$

$\square$

[A useful extension of this lemma is the following...

$$\phi \in \partial f(\bar{x}) \Leftrightarrow 0 \in \partial (f - \langle \phi, \cdot \rangle)(\bar{x})$$
$$\Leftrightarrow \bar{x} \text{ globally minimizes } f(x) - \langle \phi, x \rangle.$$
]

Theorem (Max Formula) (An Existence Theorem for Subgradients)

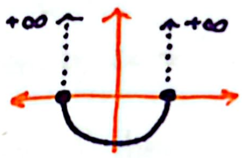
For any closed, convex, proper f with $\bar{x} \in \text{int dom} f$, $v \in E$, the following are equal and finite

$$f'(\bar{x}; v) = \begin{cases} \max & \langle \phi, v \rangle \\ \text{s.t.} & \phi \in \partial f(x). \end{cases}$$

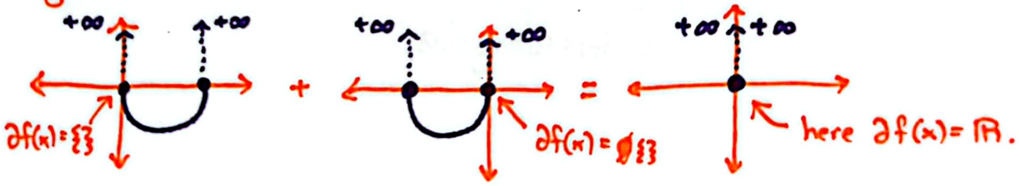
In particular, $\partial f(\bar{x})$ is nonempty.

[As an aside, on the boundary of the domain of f, there may not be any subgradients, so the interior condition is critical.

For example, $f(x) = \begin{cases} -\sqrt{1-x^2} & \text{if } x \in [-1,1] \\ +\infty & \text{otherwise} \end{cases}$ is closed, convex and proper but has no subgradients at ± 1.



Similarly, we will not have a sum rule for subgrad at the boundary of domains. For example, adding two functions with no subgradients at 0 can yield subgradients at 0...



]

Proof. First we show $f'(\bar{x}; v)$ is finite/exists.

For small enough $s > 0$, $\bar{x} \pm sv \in \text{dom} f$ (since $\bar{x} \in \text{int dom} f$).

Then our nondecreasing property from convexity ensures

$$\forall t \in [-s, s] \setminus \{0\} \quad \frac{f(\bar{x} - sv) - f(\bar{x})}{s} \leq \frac{f(\bar{x} + tv) - f(\bar{x})}{t} \leq \frac{f(\bar{x} + sv) - f(\bar{x})}{s}.$$

$\Rightarrow \lim_{t \searrow 0} \frac{f(\bar{x}+tv) - f(\bar{x})}{t}$ exists since it is monotone decreasing and bounded below.

Noting every $\phi \in \partial f(\bar{x})$ has $f'(\bar{x}; v) \geq \langle \phi, v \rangle$, (by our previous lemma's characterization)

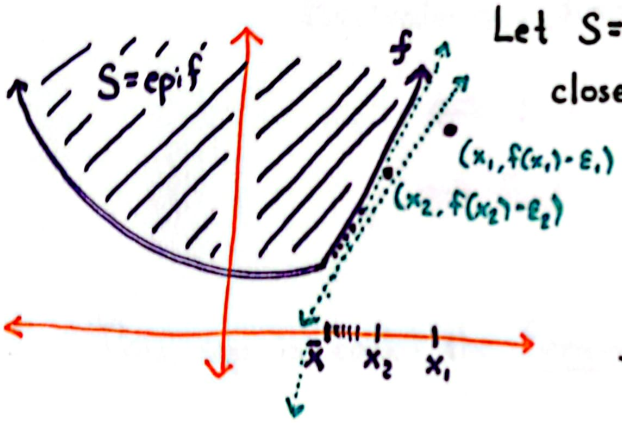
we immediately get the inequality

$$f'(\bar{x}; v) \geq \begin{cases} \max \langle v, \phi \rangle \\ \text{s.t. } \phi \in \partial f(\bar{x}). \end{cases}$$

It then suffices to show some ϕ exists with $f'(x; v) = \langle \phi, v \rangle$.

The seperating hyperplane theorem will let us construct this...

<u>Proof by picture</u>. Pick $t_n \searrow 0$ and set $x_n = \bar{x} + t_n v \to \bar{x}$.



Let $S = \text{epi } f = \{(x, h) \mid f(x) \leq h\}$ be the closed, convex set of everything above f's graph.

Then for any $\varepsilon_n > 0$, $\varepsilon_n \searrow 0$, the points $(x_n, f(x_n) - \varepsilon_n) \notin S$, so we can seperate them by some tangent plane.

Taking a subsequence of these planes will give a sequence converging to a plane touching the graph of f. This is our subgradient!! $\square$

Using this theorem, we can verify our previous claim that $f \in C^1$, convex must have $\partial f(x) = \{\nabla f(x)\}$. This follows since $f \in C^1$ implies

$$f'(x, v) \underbrace{=}_{\text{by } f \in C^1} \langle \nabla f(x), v \rangle \underbrace{=}_{\text{by our theorem}} \max \{\langle \phi, v \rangle \mid \phi \in \partial f(x)\}$$

$$\Rightarrow 0 = \max \{\langle \phi - \nabla f(x), v \rangle \mid \phi \in \partial f(x)\}$$

$$\Rightarrow \partial f(x) = \{\nabla f(x)\}.$$

4. Fenchel Duality

We have just seen that given $x \in \operatorname{int} \operatorname{dom} f$, we can always find some $\phi \in \partial f(\bar{x})$.

Fenchel Duality asks for the reverse of this:

Given $\phi \in E$, find $\bar{x}$ with $\phi \in \partial f(\bar{x})$.

Equivalenty, find $0 \in \partial(f - \langle \phi, \cdot \rangle)(\bar{x})$

$\iff$ find $\bar{x}$ attaining $\inf f - \langle \phi, \cdot \rangle$

$\iff$ find $\bar{x}$ attaining $\sup \langle \phi, \cdot \rangle - f$.

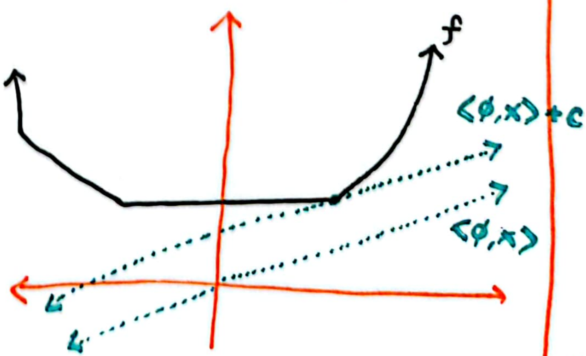
This sup is called the Fenchel Conjugate

$$f^*(\phi) = \sup_{x \in E} \underbrace{\langle \phi, x \rangle - f(x)}.$$

Note each of these terms is linear in ϕ. So the conjugate is just a maximum of (infinitely many) linear functions.

As was our motivation, this sup is attained whereever $\phi \in \partial f(x)$, so computing f^* can be geometrically seen as raising $\langle \phi, x \rangle + c$ until it is tangent to graph f.

(then $f^*(\phi) = -c$)



Properties of Fenchel Conjugates

These are essentially by definition

(1) If $\bar{x}$ attains the max value of $\sup \langle \phi, x \rangle - f(x)$,

then $\phi \in \partial f(\bar{x})$.

(2) If $\phi \in \partial f(\bar{x})$, then $\bar{x}$ attains $\sup \langle \phi, x \rangle - f(x)$.

(3) $\forall \bar{x}, \phi \in E \quad f^*(\phi) + f(x) \geq \langle \phi, \bar{x} \rangle$

Proof. $f^*(\phi) = \sup \langle \phi, x \rangle - f(x) \geq \langle \phi, \bar{x} \rangle - f(\bar{x})$. $\square$

Hence if f is proper, f^* is never equal to $-\infty$.

(1)-(3) are summarized by the following standard proposition...

Proposition (Fenchel-Young Inequality)

For any $\bar{x}, \phi \in E$, $f^*(\phi) + f(x) \geq \langle \phi, x \rangle$

and equality holds if and only if $\phi \in \partial f(\bar{x})$.

Proof delayed until after some examples

(4) f^* is convex (regardless of whether f is).

(5) f^* is closed (regardless of whether f is).

Proof delayed to next section.

(6) $f^{**} = f$ if and only if f is closed and convex.

Examples of Fenchel Conjugates

(1) Linear functions, $f(x) = \langle c, x \rangle$ for some $c \in E$.

$$f^*(\phi) = \sup_x \langle \phi, x \rangle - f(x)$$
$$= \sup_x \langle \phi - c, x \rangle$$
$$= \begin{cases} 0 & \text{if } \phi = c \\ +\infty & \text{if } \phi \neq c. \end{cases}$$

Indicator function for whether $\phi = c$.

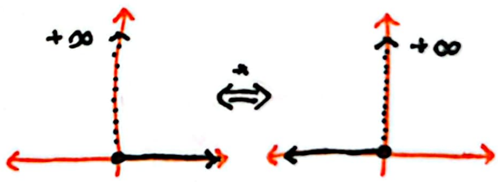
(2) Indicator functions, $f(x) = \begin{cases} 0 & \text{if } x = c \\ +\infty & \text{if } x \neq c \end{cases}$ for some $c \in E$.

$$f^*(\phi) = \sup_x \langle \phi, x \rangle - f(x)$$
$$= \sup_{x = c} \langle \phi, x \rangle$$
$$= \langle c, \phi \rangle.$$

A linear function. This verifies $f = f^{**}$ for linear functions.

(4) Nonnegativity Indicator functions, $f(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ +\infty & \text{if } x < 0. \end{cases}$

$$f^*(\phi) = \sup_x \langle \phi, x \rangle - f(x)$$
$$= \sup_{x \geq 0} \langle \phi, x \rangle$$
$$= \begin{cases} +\infty & \text{if } \phi > 0 \\ 0 & \text{if } \phi \leq 0. \end{cases}$$



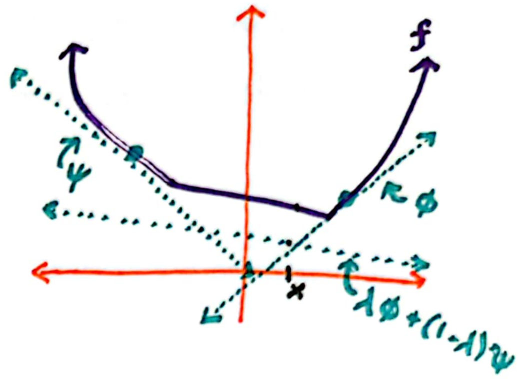
(5) A few interesting nonlinear functions (calculations verifying omitted)

$$f(x) = \frac{1}{p}|x|^p \Longleftrightarrow f^*(\phi) = \frac{1}{q}|\phi|^q \quad \text{for } \frac{1}{p} + \frac{1}{q} = 1.$$

$$f(x) = \sqrt{1 + x^2} \Longleftrightarrow f^*(\phi) = \begin{cases} -\sqrt{1 - \phi^2} & \text{if } \phi \in [-1, 1] \\ +\infty & \text{otherwise.} \end{cases}$$

Proof of Property (4): f^* is always convex.

Consider any $\phi, \psi \in E$, $\lambda \in [0,1]$.



For any $x \in E$, Fenchel-Young ensures

$$\begin{aligned} &\lambda\ \big(f^*(\phi) \geq \langle \phi, x\rangle - f(x)\big) \\ + &(1-\lambda)\ \big(f^*(\psi) \geq \langle \psi, x\rangle - f(x)\big) \\ \hline &\lambda f^*(\phi) + (1-\lambda) f^*(\psi) \geq \langle \lambda\phi + (1-\lambda)\psi, x\rangle - f(x) \end{aligned}$$

Taking the sup over all $x \in E$, gives the claimed convexity result:

$$\lambda f^*(\phi) + (1-\lambda) f^*(\psi) \geq f^*(\lambda\phi + (1-\lambda)\psi). \qquad \square$$

Proof of Property (5): f^* is always closed.

(i.e. lower semicontinuous, meaning $\liminf_{\phi' \to \phi} f^*(\phi') \geq f^*(\phi)$.)

Consider any $\phi_i \to \phi \in E$ and $x \in E$.

Then $f^*(\phi_i) \geq \langle \phi_i, x\rangle - f(x)$

$\Rightarrow \liminf_{i\to\infty} f^*(\phi_i) \geq \langle \phi, x\rangle - f(x)$ ← RHS is continuous in ϕ.

Taking the sup over all $x \in E$, gives the claimed closedness result:

$$\liminf f^*(\phi_i) \geq \sup_x \langle \phi, x\rangle - f(x) = f^*(\phi).$$

Theorem (Fenchel Duality)

For any functions $f: E \to (-\infty, \infty]$ and $g: Y \to (-\infty, \infty]$ and linear map $A: E \to Y$. Denote primal, dual problems as

$$p^* = \inf_{x \in E} f(x) + g(Ax)$$

$$d^* = \sup_{\phi \in Y} -f^*(A^*\phi) - g^*(-\phi).$$

In general, weak duality holds: $p^* \geq d^*$.

Furthermore, if f and g are both convex with

$$0 \in \operatorname{int}(\operatorname{dom} g - A \operatorname{dom} f),$$

then strong duality holds, $p^* = d^*$, and the dual is attained if finite.

Proof. Weak duality follows from Fenchel-Young as all $x, \phi \in E \times Y$ have

$$\begin{array}{rl} & f(x) + f^*(A^*\phi) \geq \langle x, A^*\phi \rangle \\ + & g(Ax) + g^*(-\phi) \geq \langle Ax, -\phi \rangle \\ \hline & f(x) + g(Ax) + f^*(A^*\phi) + g^*(-\phi) \geq 0. \end{array}$$

$\Rightarrow$ Primal objective at x $\geq$ Dual objective at ϕ $\quad \forall x \in E, \phi \in Y$.

For strong duality, we look at the perturbed value function

$$h(u) = \inf_{x \in E} f(x) + g(Ax + u), \text{ for } u \in Y$$

Note $h(0) = p^*$.

Claim: h is convex with $\operatorname{dom} h = \operatorname{dom} g - A \operatorname{dom} f$
$= \{ y - Ax \mid y \in \operatorname{dom} g, x \in \operatorname{dom} f \}$

Proof. HW3. $\square$

By assumption, $0 \in \text{int dom } h$.

Then by the max formula, $-\phi \in \partial h(0)$ exists.

Consider any $u \in Y$, $x \in E$, then we have

$$h(0) \leq h(u) + \langle \phi, u \rangle \quad \leftarrow \text{by definition of } \partial h(0)$$

$$\leq f(x) + g(Ax+u) + \langle \phi, u \rangle \quad \leftarrow \text{by definition of } h$$

$$= f(x) - \langle A^*\phi, x \rangle + g(\underbrace{Ax+u}) - \langle -\phi, \underbrace{Ax+u} \rangle$$

Note this can be any $y \in Y$.

Taking the infimum over all $u \in Y$ gives

$$h(0) \leq f(x) - \langle A^*\phi, x \rangle - g^*(-\phi).$$

Taking the infimum over all $x \in E$ gives

$$h(0) \leq -f^*(A^*\phi) - g^*(-\phi) \leq d^* \leq p^* = h(0).$$

Hence it must be equality all the way through!

So $d^* = p^*$, giving strong duality

and $-f^*(A^*\phi) - g^*(-\phi) = d^*$, so dual optimal is attained. □

As an example, lets derive Semidefinite Programming duality.

The primal $\begin{cases} \min \langle C, X \rangle \\ \text{s.t. } AX = b \\ \quad X \succeq 0 \end{cases}$ can be written as $f(X) = \begin{cases} \langle C, X \rangle & \text{if } X \succeq 0 \\ +\infty & \text{otherwise} \end{cases}$

$$g(y) = \begin{cases} 0 & \text{if } y = b \\ +\infty & \text{otherwise.} \end{cases}$$

The conjugates of these are $f^*(S) = \begin{cases} 0 & \text{if } S - C \preceq 0 \\ +\infty & \text{otherwise} \end{cases}$ and $g^*(y) = \langle b, y \rangle$.

Then Fenchel Duality (given a slater point) ensures

$$\begin{cases} \min \langle C, X \rangle \\ \text{s.t. } AX = b \\ \quad X \succeq 0 \end{cases} = p^* = d^* = \begin{cases} \max \langle b, y \rangle \\ \text{s.t. } A^* y \preceq C. \end{cases}$$

Corollary (Convex Subgradient Calculus)

If f, g satisfy the above conditions for strong duality,

then $\partial(f + g \circ A)(x) = \partial f(x) + A^* \partial g(Ax)$.

this gives both the sum rule and the chain rule for linear maps.

Proof. HW3. $\square$

An aside Duality can be viewed as switching the roles of solutions and gradients in our problem.

To illustrate this, suppose $f = f^{**}$ (which we will show often holds in the next section).

Then Fenchel-Young on f ensures

$$f^*(\phi) + f(x) = \langle \phi, x \rangle \text{ iff } \phi \in \partial f(x).$$

However, Fenchel-Young on f^* ensures

$$f(x) + f^*(\phi) = \langle x, \phi \rangle \text{ iff } x \in \partial f^*(\phi).$$

Thus $\phi \in \partial f(x) \Leftrightarrow x \in \partial f^*(\phi)$.

They are (set-valued) inverses!

If f and f^* are both C^1, then the nonlinear map

$$\nabla f : E \to E \quad \text{has inverse } \nabla f^* = \nabla f^{-1}.$$

5. Fenchel Biconjugates

Previously we claimed (as Property (6)) that $h = h^{**}$ if and only if h is closed and convex.

Lemma For any $h : E \to (-\infty, \infty]$, $h^{**} \leq h$. ← meaning all $x \in E$ have $h^{**}(x) \leq h(x)$.

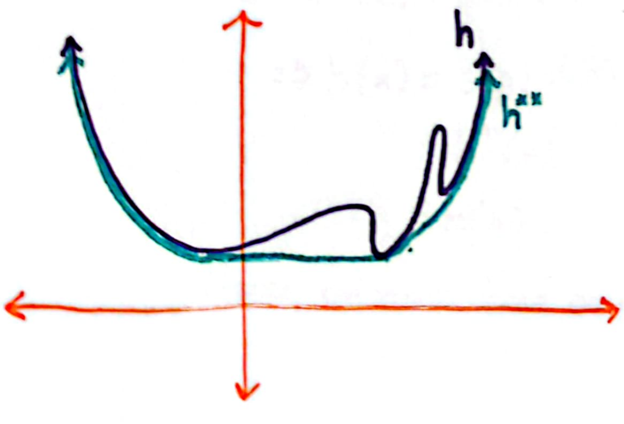
Proof. For any $x, \phi \in E$, Fenchel-Young ensures $\langle x, \phi \rangle - h^*(\phi) \leq h(x)$.

Consider a sequence ϕ_i attaining $\sup \langle x, \phi \rangle - h^*(\phi)$.

Then
$$h^{**}(x) = \lim_{i\to\infty} \langle x, \phi_i \rangle - h^*(\phi_i)$$
$$\leq \lim_{i\to\infty} h(x)$$
$$= h(x) \text{ for any } x \in E. \qquad \square$$

[In fact, one can show the biconjugate h^{**} is the largest closed, convex function minorizing h (that is, $h^{**} \leq h$).

(Proof of this claim. Suppose g is closed, convex with $g \leq h$. Then $g^* \geq h^*$. Then $g = g^{**} \leq h^{**}$. $\square$)



(Interestingly, in this example sketch, $h \in C^2$ but h^{**} is only C^1.)]

We say $\alpha(x) = \langle \phi, x \rangle + c$ is an <u>affine minorant</u> of h if $\alpha \leq h$ where $\phi \in E$, $c \in \mathbb{R}$.

<u>Theorem (Fenchel Biconjugation)</u>

The following are equivalent for any proper $h : E \to (-\infty, \infty]$.

(i) h is closed and convex,

(ii) $h = h^{**}$,

(iii) For all $x \in E$, $h(x) = \sup\{\alpha(x) \mid \alpha$ is an affine minorant of $h\}$.

Proof. $\boxed{(ii) \Rightarrow (i)}$

h^{**} is closed and convex since all conjugates are.

$\Rightarrow h = h^{**}$ is also closed and convex.

$\boxed{(iii) \Rightarrow (ii)}$

Every affine function $\alpha(x) = \langle \phi, x \rangle + c$ has $\alpha = \alpha^{**}$ (we computed this as an example previously).

Since each affine minorant is closed and convex, $\alpha = \alpha^{**} \leq h^{**}$.

$$\begin{aligned} \Rightarrow h(x) &= \sup\{\alpha(x) \mid \alpha \text{ is an affine minorant}\} && \text{by assumption} \\ &\leq h^{**}(x) && \text{by above reasoning} \\ &\leq h(x) && \text{by previous lemma} \end{aligned}$$

Hence we must have equality all the way through: $h = h^{**}$.

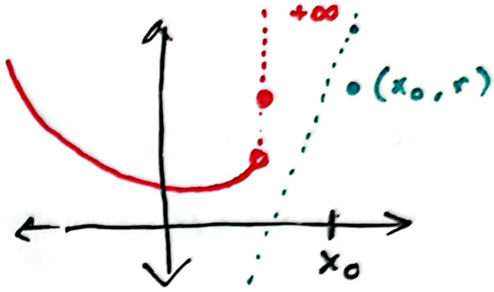
$\boxed{(i) \Rightarrow (iii)}$ Suppose h is closed and convex

Let $x_0 \in E$ be generic.

If $x_0 \notin \operatorname{cl} \operatorname{dom} h$ (i.e. $h(x_0) = +\infty$ and bounded away from points that do.)

Seperating hyperplane can give $\alpha(x_0) \geq r$ $\forall r$.

Take $r \to \infty$



Suppose $x_0 \in \operatorname{cl} \operatorname{dom} h$.

Fix any $r < h(x_0)$.

Lower semicontinuity ensures $\{x \mid h(x) > r\}$ is open.

$\Rightarrow$ There exists an open, convex neighborhood of x_0, U, within $\{x \mid h(x_0) > r\}$.

Let $\delta_U(x) = \begin{cases} 0 & \text{if } x \in U \\ +\infty & \text{otherwise} \end{cases}$

Fenchel Duality Theorem ensures $\phi \in E$ attaining

$$r \leq \underbrace{\inf_x \{h(x) + \delta_U(x)\}}_{\text{primal}} = \underbrace{-h^*(\phi) - \delta_U^*(-\phi)}_{\text{dual}}.$$

Consider $\alpha(x) = \langle \phi, x \rangle + \delta_U^*(-\phi) + r$

This minorizes h: $\alpha(x) \leq h(x) \underbrace{+ h^*(\phi) + \delta_U^*(-\phi) + r}_{\geq 0}$
$\leq h(x)$ (by above)

Further, $\alpha(x_0) \geq r$ as $\alpha(x_0) = -\langle -\phi, x_0 \rangle + \delta_U^*(-\phi) + r$

$\geq -(\delta_U(x_0) + \delta_U^*(-\phi)) + \delta_U^*(-\phi) + r$

$= r$.

Taking $r \to h(x_0) \Rightarrow (iii)$ $\square$

We could similarly prove the following local version of this result (proof omitted for time)...

<u>Theorem</u> For any convex, proper $h: E \to [-\infty, \infty]$ and $x \in \operatorname{dom} h$,

$$h(x) = h^{**}(x) \text{ if and only if } \underbrace{\liminf_{x' \to x} h(x') \geq h(x)}_{\text{that is, } h \text{ is lower semicontinuous at } x.}.$$

6. Lagrange Duality

Recall our primal and dual Lagrange problems are

$$p^* = \inf_{x \in E} \sup_{\lambda \geq 0} L(x;\lambda) \quad \left(= \begin{cases} \min f(x) \\ \text{s.t. } g_i(x) \leq 0 \end{cases} \right)$$

$$d^* = \sup_{\lambda \geq 0} \inf_{x \in E} L(x;\lambda) \quad \left(= \max \Phi(\lambda) \right)$$

where $L(x;\lambda) = f(x) + \lambda^T g(x)$ is the Lagrangian.

As motivated by HW2 for LPs, we will derive our strong duality between these via the value function

$$v(\Delta b) = \begin{cases} \min f(x) \\ \text{s.t. } g_i(x) \leq \Delta b_i \quad \forall i = 1 \dots m \end{cases}$$

for any perturbation of the constraints $\Delta b \in \mathbb{R}^m$.

Theorem For any $f, g_i : E \to (-\infty, \infty]$ (not necessarily convex), the following are true...

(1) Primal Optimal is given by $p^* = v(0)$.

(2) The Dual Function is given by $v^*(-\lambda) = \begin{cases} -\Phi(\lambda) & \text{if } \lambda \geq 0 \\ +\infty & \text{otherwise.} \end{cases}$

(3) Dual Optimal is given by $d^* = v^{**}(0)$.

[Note this implies all of our original claims about Lagrange Duality from the start of this unit:

- Weak duality holds as $v^{**} \le v$, so $p^* = v(0) \ge v^{**}(0) \ge d^*$.
- Strong Duality for convex f, g_i amounts to needing $v^{**}(0) = v(0)$, which holds iff v is lower semicontinuous at 0.
- The set of dual optimal solutions is given by $0 \in \partial v^*(-\lambda) \Leftrightarrow -\lambda \in \partial v^{**}(0)$.]

Proof. (1) is just by definition true.

(2) We can directly evaluate v^*, giving the claimed formula

$$v^*(-\lambda) = \sup_{\Delta b} \{ \langle -\lambda, \Delta b \rangle - v(\Delta b) \}$$

$$\begin{pmatrix}\text{Plugging in the}\\ \text{Definition of } v\end{pmatrix} = \sup \{ \langle -\lambda, \Delta b \rangle - f(x) \mid x \in \operatorname{dom} f,\ \Delta b \in \mathbb{R}^m,\ g(x) \le \Delta b \}$$

$$\begin{pmatrix}\text{Change of Variables}\\ g(x)+s=\Delta b\\ s \ge 0\end{pmatrix} = \sup \{ \langle -\lambda, g(x)+s \rangle - f(x) \mid x \in \operatorname{dom} f,\ s \ge 0 \}$$

$$\begin{pmatrix}\text{Computing the}\\ \text{sup over } s \ge 0\end{pmatrix} = \begin{cases} +\infty & \text{if } \lambda \not\ge 0 \\ \sup_x \underbrace{\langle -\lambda, g(x) \rangle - f(x)}_{= -L(x;\lambda)} & \text{if } \lambda \ge 0 \end{cases}$$

$$\begin{pmatrix}\text{By the Definition}\\ \text{of } \Phi(\cdot)\end{pmatrix} = \begin{cases} +\infty & \text{if } \lambda \not\ge 0 \\ -\Phi(\lambda) & \text{if } \lambda \ge 0. \end{cases}$$

(3) Also directly computed.

$$\begin{aligned} d^* &= \sup_{\lambda \ge 0} \Phi(\lambda) = \sup -v^*(-\lambda) \\ &= \sup \langle 0, \lambda \rangle - v^*(-\lambda) \\ (\text{replacing } \lambda \to -\lambda)\quad &= \sup \langle 0, \lambda \rangle - v^*(\lambda) \\ &= v^{**}(0). \end{aligned}$$

$\square$

A natural question is then, when is $v(\cdot)$ lower semicontinuous?
(at $\Delta b = 0$)

Constraint Qualification and Slater Points provide one answer.

Recall to apply our KKT optimality conditions we needed the "Constraint Qualification" that some $d \in E$ has

$$\langle d, \nabla g_i(\bar{x}) \rangle < 0 \quad \forall i \text{ where } g_i(\bar{x}) = 0.$$

For convex g_i, this is equivalent to the existence of a "Slater Point", namely $\tilde{x}_{SL} \in E$ with $g_i(\tilde{x}_{SL}) < 0 \ \forall i$.

These (equivalent) conditions suffice for convex programs to ensure strong duality (i.e. $v(\cdot)$ is lower semicontinuous at 0 and convex).

Proof Sketch. If some $\tilde{x}$ has $g(\tilde{x}) < 0$,
then $v(\cdot)$ is finite or $-\infty$ for all Δb near 0.

Claim: $v(\cdot)$ is convex if all f, g_i are convex.

$\Rightarrow$ v is either finite near 0 or always $-\infty$ near 0.

If finite, $0 \in \operatorname{int} \operatorname{dom} v$.

$\Rightarrow \exists \phi \in \partial v(0)$

$\Rightarrow$ v is lower semicontinuous.

If $-\infty$, the weak duality ensures $d^* = -\infty = p^*$. □