Recitation 4-3-08: Convex Programs + Duality

Dual Functions

<u>Def:</u> Dual of $F(x)$ is $\boxed{F^*(y) = \sup_x (x \cdot y - F(x))}$

e.g. $F(x) = x^2$

$$F^*(y) = \sup_x (xy - x^2) = \frac{y^2}{4}$$

(using $\frac{\partial}{\partial x}(xy - x^2) = y - 2x$)

e.g. $F(x) = \|x\|$ where $x \in \mathbb{R}^n$, $\|\cdot\|$ is a norm on $\mathbb{R}^n$

Recall: dual norm is $\|z\|_* = \sup_{\|x\| \le 1} (z \cdot x)$

Consider $F^*(y) = \sup_x (x \cdot y - \|x\|)$. How do we know how $\|\cdot\|$ behaves? Look at 2 cases:

1) If $\|y\|_* > 1$

then $\sup_{\|x\| \le 1} (y \cdot x) > 1$, so $y \cdot x > 1$ for some x where $\|x\| \le 1$.

Let $w = tx$, so

$$w \cdot y - \|w\| = t\underbrace{(x \cdot y - \|x\|)}$$

As $t \to \infty$, this $(\;) \to \infty$, so $F^*(y) = \infty$

2) If $\|y\|_* \le 1$

then $y \cdot x \le \|x\| \|y\|_*$ for all x where $\|x\| \le 1$,

so $y \cdot x \le \|x\| \Rightarrow y \cdot x - \|x\| \le 0$

$x = 0$ maximizes this, so $F^*(y) = 0$

$$F^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \le 1 \\ \infty & \text{otherwise} \end{cases}$$

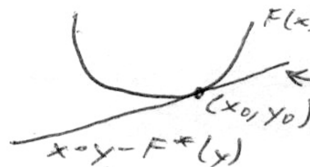
(cont'd)

Recall how we talked about duality + subgradients in lecture:

Fenchel's inequality: $F^*(y) \geq x \cdot y - F(x)$ for all x

$\Rightarrow F(x) \geq x \cdot y - F^*(y)$ $\Big\}$ linear lower bound on $F(x)$ which is tight for some x

e.g.



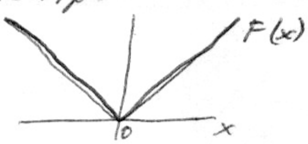
← lower bound determined by choice of y

y is slope: $y_0 \in \partial F(x_0)$

y_0 is a subgradient of F at x_0

So dual function lets us find subgradients of functions! Also, ranges + slopes have nice relationship:

$F(x) = \|x\|_2$



range: $-\infty$ to 0 to ∞

slope: $\downarrow -1$ $\quad \downarrow +1$

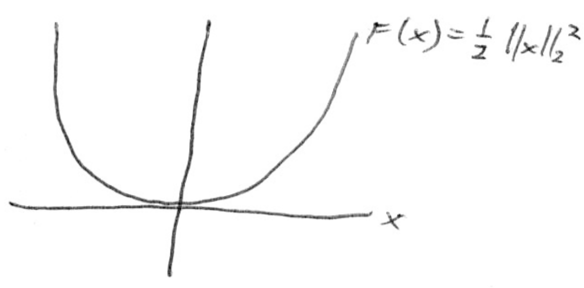
$F^*(x)$

range: $-\infty$ to -1 to 1 to ∞

slope: $\downarrow -\infty$ $\quad \downarrow 0$ $\quad \downarrow \infty$

e.g. $F(x) = \frac{1}{2}\|x\|_*^2$

What does $F^*(x)$ look like, just based on what $F(x)$ looks like?



Answer: pretty similar!

Now do the math:

$y \cdot x \leq \|y\|_* \|x\|$ (as before)

$\Rightarrow y \cdot x - \frac{1}{2}\|x\|^2 \leq \|y\|_* \|x\| - \frac{1}{2}\|x\|^2$ $\Big\}$ Right side max at $\|x\| = \|y\|_*$

$\Rightarrow F^*(y) \leq \frac{1}{2}\|y\|_*^2$

Let x be vector such that $y \cdot x = \|y\|_* \|x\|$ and $\|x\| = \|y\|_*$

For this x, $y \cdot x - \frac{1}{2}\|x\|^2 = \frac{1}{2}\|y\|_*^2 \Rightarrow F^*(y) \geq \frac{1}{2}\|y\|_*^2$

The duals of similar functions (like $\|x\|$ versus $\frac{1}{2}\|x\|^2$) can be very different, and this can affect how ~~transformations~~ you may choose to phrase ~~do on~~ optimization problems.

For example: 2 norm approximation problems

e.g. 1 Approximate $Ax \approx b$ by solving:

$$\left\{ \min_x \|Ax - b\| \right\}$$

Reformulate it as:

$$\left\{ \begin{array}{l} \min_{x,y} \|y\| \\ \text{s.t. } Ax - b = y \end{array} \right\}$$

Derive dual:

$$L(x,y,\lambda) = \|y\| + \lambda^T(Ax - b - y)$$

where $\lambda \in \mathbb{R}^n$

Minimize $L(x,y,\lambda)$ w.r.t. x,y:

$$\inf_{x,y} L(x,y,\lambda) = \left(\inf_x \lambda^T Ax\right) - \lambda^T b - \sup_y \left(\lambda^T y - \|y\|\right)$$

If $\lambda^T A \neq 0$, then this is $-\infty$.

If $\lambda^T A = 0$, then we get

$$\max_\lambda -\lambda^T b - \sup_y (\lambda^T y - \|y\|)$$

$\Downarrow$ ← using our example from before with $F(x) = \|x\|$

$$\max_\lambda -\lambda^T b - I_{\{\|y\|_* \leq 1\}}(\lambda)$$

$\Downarrow$ ← since we will never choose to make objective $-\infty$

$$\left\{ \begin{array}{l} \max_\lambda -\lambda^T b \\ \text{s.t. } \|\lambda\|_* \leq 1 \\ \quad A^T\lambda = 0 \end{array} \right\}$$

Duals of indicators

$$I_S(x) = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

From before, we know for $F(x) = \|x\|$, we have $F^*(x) = \begin{cases} 0 & \text{if } \|y\|_* \leq 1 \\ \infty & \text{otherwise} \end{cases}$

i.e. $F^*(x) = I_{\{\|y\|_* \leq 1\}}(x)$

(2) Now suppose we change the objective:

$$\left\{\begin{array}{l}\min_{x,y} \ \|y\| \\ \text{s.t. } Ax-b=y\end{array}\right\} \longrightarrow \left\{\begin{array}{l}\min_{x,y} \ \frac{1}{2}\|y\|^2 \\ \text{s.t. } Ax-b=y\end{array}\right\}$$

These are equivalent in terms of optimal values, but they are not the same + lead to different duals.

$$L(x,y,\lambda) = \frac{1}{2}\|y\|^2 + \lambda^T(Ax-b-y)$$

$$= \underbrace{\frac{1}{2}\|y\|^2 - \lambda^T y}_{} + \lambda^T Ax - \lambda^T b$$

$$-\sup_y\left(\lambda^T y - \frac{1}{2}\|y\|^2\right) = -\frac{1}{2}\|\lambda\|_*^2 \quad \text{(from before)}$$

$$\left\{\begin{array}{l}\max_\lambda \ -\frac{1}{2}\|\lambda\|_*^2 - \lambda^T b \\ \text{s.t. } A^T\lambda = 0\end{array}\right\}$$

More on dual functions

Infimal convolution: $(F_1 \square F_2)(x) = \inf_{a+b=x}\left(F_1(a) + F_2(b)\right)$

From class: $G(x) = F_1(x) \square F_2(x) \Longrightarrow G^*(y) = F_1^*(y) + F_2^*(y)$

$G(x) = F_1(x) + F_2(x) \Longrightarrow G^*(y) = F_1^*(y) \square F_2^*(y)$

e.g. $G(x) = x^2 + e^x$ What is $G^*(y)$?

Write $F_1(x) = x^2$ $F_2(x) = e^x$

(From class,) $F_1^*(a) = \frac{a^2}{2}$ $F_2^*(b) = b\ln b - b$

$$G^*(y) = (F_1^* \square F_2^*)(y)$$

$$= \inf_{a+b=y}\left(F_1^*(a) + F_2^*(b)\right)$$

$$= \inf_{a+b=y}\left(\frac{a^2}{2} + b\cdot\ln(b) - b\right)$$

Differentiable functions

Suppose $G(x)$ is convex + differentiable with domain $G = \mathbb{R}^n$. What is $G^*(y) = \sup_x (y \cdot x - G(x))$?

If x^* maximizes $y \cdot x - G(x)$,

then $y = \nabla G(x^*)$. (take gradient + set $=0$)

Also, $y = \nabla G(x^*) \Rightarrow x^*$ maximizes $y \cdot x - G(x)$.

So if $y = \nabla G(x^*)$, then

$$G^*(y) = x^{*T} \nabla G(x^*) - G(x^*)$$

1) If we can solve $y = \nabla G(z)$ for z, we ~~know~~ can find $G^*(y)$.

2) Given any $z \in \mathbb{R}^n$, we can find $G^*(y) = z^T \nabla G(z) - G(z)$ for some y where $y = \nabla G(z)$.

e.g. (from before)

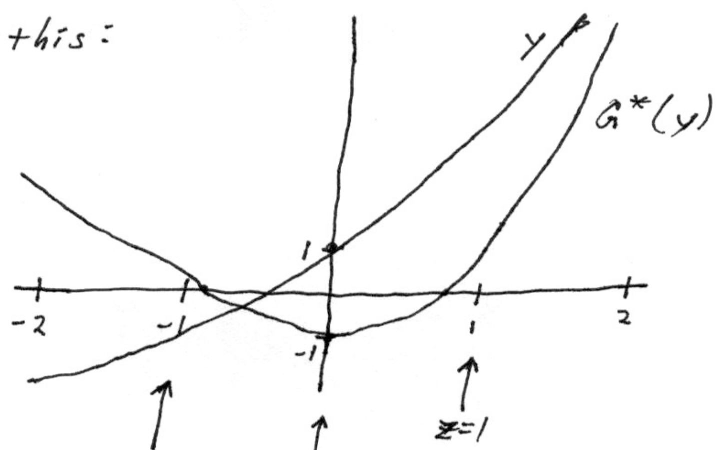
$G(x) = x^2 + e^x \qquad \nabla G(z) = 2z + e^z$

Choose $z \in \mathbb{R}$. Let $y = 2z + e^z$.

Then $G^*(y) = z(2z + e^z) - z^2 - e^z$

$= z^2 + ze^z - e^z$

Plot this:



$z=-1$, $y = -2 + e^{-1}$, $G^*(y) = 1 - 2e^{-1}$

$z=0$, $y=1$, $G^*(y) = -1$

$z=1$, $y = 2+e$, $G^*(y) = 1$

Note: It's hard to solve $y = \nabla G(z)$ for z sometimes, but this still lets us plot $G^*(y)$.

Dual of a Convex Program

e.g. Motivation: LP of form $\begin{cases} \min_x & c^T x \\ \text{s.t.} & a_i^T x \le b_i, \quad i=1,\dots,m \end{cases}$

But not sure about a_i,

so say $a_i \in \mathcal{E}_i = \{ \bar{a}_i + P_i u \mid \|u\|_2 \le 1 \}$

(a_i is in some ellipsoid).

Then we can rewrite this (Boyd + V., p. 157)

as an SOCP: $\begin{cases} \min_x & c^T x \\ \text{s.t.} & \bar{a}_i^T x + \|P_i^T x\|_2 \le b_i, \quad i=1,\dots,m \end{cases}$

Note: $\|P_i^T x\|_2$ terms prevent x from being large in directions ~~where~~ with large uncertainty in a_i (regularization).

The constraints are 2nd order cone constraints.

i.e. $\|P_i^T x\|_2 \le -\bar{a}_i^T x + b_i$

means

$(P_i^T x, -\bar{a}_i^T x + b_i)$ lies in the 2nd order cone (in $\mathbb{R}^{k+1}$ where P_i is $k \times n$)

Say we have $\min_{x,y} \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$

$$\text{s.t.} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \left\| \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\|_2 \le 1$$

i.e. $c = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, $\bar{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $P = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$, $b = 1$

Then the constraint says $\left\| \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\|_2 \le 1 - x - y$

i.e. $\left(\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, 1-x-y \right) \in K \leftarrow$ SOC in $\mathbb{R}^3$

i.e. $\underbrace{\left(x, y, 1 - \tfrac{x}{3} - y \right)}$ $\in K$

This is a hyperplane which cuts out a conic section of K; here, it is an ellipse.

Our (x,y) must lie within this ellipse.

Now find the dual of this problem.

From class: primal $\begin{cases} \min f(x) \\ \text{s.t. } Ax + b \in K \end{cases}$

$\downarrow$

dual $\begin{cases} \max\limits_{z} \; -b^T z - f^*(A^T z) \\ \text{s.t. } z \in K^* \end{cases}$

Rewrite our constraint:

$$\left(\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, 1-x-y \right) \in K$$

$$\Updownarrow$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \in K$$

So dual is:

$$\begin{cases} \max\limits_{z} \; -\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^T z - f^*\left(\begin{pmatrix} 3 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}^T z \right) \\ \text{s.t. } z \in K^* = K \quad (2^{nd} \text{ order cone is self-dual}) \end{cases}$$

$$f^*(t) = \sup_{v} \left(t^T v - c^T v \right) = \begin{cases} 0 & \text{if } t = c \\ \infty & \text{if } t \neq c \end{cases}$$

So rewrite dual as:

$$\begin{cases} \max\limits_{z} \; -\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^T z \\ \text{s.t. } z \in K \quad \text{and} \quad \begin{pmatrix} 3 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}^T z = c \end{cases}$$

Other Stuff

$$\begin{aligned} \min\ & f_0(x) \\ \text{s.t. } & f_i(x) \le 0,\ i=1,\dots,m \\ & Ax=b \end{aligned}$$

Slater's condition

Strong duality holds if there is $x \in \operatorname{relint}(\underset{\text{domain}}{\mathcal{D}})$ with $\underbrace{f_i(x) \le 0,\ i=1,\dots,k}_{f_i \text{ affine}}$ $\underbrace{f_i(x) < 0,\ i=k+1,\dots,m}_{f_i \text{ convex}}$, $Ax=b$

(and f_0 convex)

e.g. LPs + QPs

All f_i are affine, so strong duality holds if problem is feasible.

e.g. $\begin{cases} \min\limits_{X} \ \log|X^{-1}| \\ \text{s.t. } a_i^T X a_i \le 1,\ i=1,\dots,m \end{cases}$ (min volume covering ellipsoid)

(Note implicit constraint $X \in S^n_{++}$)

Slater's condition holds since there is always some $X \in S^n_{++}$ such that $a_i^T X a_i \le 1$, $\forall i$.