Recitation notes

April 29, 2008

1 Convex conjugates

1.1 If $f\left(x_1, x_2, \ldots, x_n\right) = \sum_{i=1}^{m} g_i\left(x_i\right)$, then $f^*\left(y_1, y_2, \ldots, y_n\right) = \sum_{i=1}^{m} g_i^*\left(y_i\right)$

By the definition of a convex conjugate of a function $g_i : \mathbb{R}^{n_k} \rightarrow \mathbb{R}$:

$$g_i^*\left(y\right) = \sup_x \left(y^T x - g_i\left(x\right)\right)$$

While:

$$\begin{aligned} f^*\left(y_1, y_2, \ldots, y_n\right) &= \sup_{x_1, x_2, \ldots, x_n} \left(\sum_{i=1}^{n} y_i^T x_i - \sum_{i=1}^{m} g_i\left(x_i\right)\right) \\ &= \sup_{x_1, x_2, \ldots, x_n} \left(\sum_{i=1}^{n} \left(y_i^T x_i - g_i\left(x_i\right)\right)\right) \end{aligned}$$

Since, in the supremum, the x_is are "independent":

$$f^*\left(y_1, y_2, \ldots, y_n\right) = \sum_{i=1}^{n} \sup_{x_i} \left(y_i^T x_i - g_i\left(x_i\right)\right)$$

1.2 For convex $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f^{**} = f$

Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex. Its convex conjugate f^* is then given by:

$$f^*\left(y\right) = \sup_x \left(y^T x - f\left(x\right)\right)$$

And its "double conjugate" f^{**} by:

$$\begin{aligned} f^{**}\left(z\right) &= \sup_y \left(z^T y - f^*\left(y\right)\right) \\ &= \sup_y \left(z^T y - \sup_x \left(y^T x - f\left(x\right)\right)\right) \\ &= \sup_y \inf_x \left(f\left(x\right) - y^T\left(x - z\right)\right) \end{aligned}$$

Since we may always choose $x = z$, we see that $f(x)$ is an element of the infimum, so $\inf_x \left(f(x) - y^T(x-z)\right) \le f(z)$ for all z, which implies that $f^{**}(z) \le f(z)$.

Next, note that since f is convex, we'll have that it is bounded below by any supporting hyperplane. In particular, $f(x) \ge f(z) + w^T(x-z)$ for some w. Choosing $y = w$ gives that:

$$\begin{aligned}\inf_x \left(f(x) - y^T(x-z)\right) &\ge \inf_x \left(f(z) + w^T(x-z) - y^T(x-z)\right) \\ &\ge f(z)\end{aligned}$$

Hence, $\sup_y \inf_x \left(f(x) - y^T(x-z)\right) \ge f(z)$ for all z, which implies that $f^{**}(z) \ge f(z)$. *Important*: this argument hinges on the existence of a supporting hyperplane of f which intersects $(z, f(z))$ for all $z \in \mathbb{R}^n$. Any convex function $f : \mathbb{R}^n \to \mathbb{R}$ must be continuous, so such a hyperplane exists. However, a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ (where $\overline{\mathbb{R}}$ is the extended reals: the real numbers, along with $\pm\infty$) need not be continuous (why?), so if we make even slightly different assumptions about f, the argument may stop working.

Combining these results gives that $f^{**}(z) = f(z)$.

1.3 Convex conjugates and norms

Recall from an earlier recitation that norms are convex.

1.3.1 Dual norms

Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ be a norm. Then the dual norm $\|\cdot\|^* : \mathbb{R}^n \to \mathbb{R}$ is defined as:

$$\|x\|^* = \sup_{\|y\| \le 1} \left|y^T x\right|$$

Note that the above supremum must be achieved, since it is a supremum of a continuous function over a compact set (the closed unit ball), and furthermore, that any y which maximizes this supremum must have $\|y\| = 1$. Further note that a general version of "Hölder's inequality" follows from this definition, since $\left|y^T x\right| \le \|x\|^*$ when $\|y\| \le 1$, so that $\left|\left(\frac{y}{\|y\|}\right)^T x\right| \le \|x\|^* \Rightarrow \left|y^T x\right| \le \|y\| \, \|x\|^*$ for all $y \in \mathbb{R}^n$.

1.3.2 Example: l^p norms

For the l^p norms for $1 \le p < \infty$, we'll have that $\|\cdot\|_p^* = \|\cdot\|_q$ (where $\frac{1}{p} + \frac{1}{q} = 1$), since by Hölder's inequality, $\left|y^T x\right| \le \|y\|_p \|x\|_q$, so that $\sup_{\|y\|_p \le 1} \left|y^T x\right| \le \|x\|_q$. This supremum is in fact achieved, since if we let $z_i = |x_i|^{q-1}$:

$$\begin{aligned}\left|z^T x\right| &= \sum_{i=1}^n z_i |x_i| \\ &= \sum_{i=1}^n |x_i|^q \\ &= \|x\|_q^q\end{aligned}$$

While:

$$\begin{aligned}\|z\|_p &= \left(\sum_{i=1}^n z_i^p\right)^{\frac{1}{p}} \\ &= \left(\sum_{i=1}^n |x_i|^{pq-p}\right)^{\frac{1}{p}} \\ &= \left(\sum_{i=1}^n |x_i|^q\right)^{\frac{1}{p}}\end{aligned}$$

The above because $\frac{1}{p} + \frac{1}{q} = 1 \Rightarrow pq = p + q$. Hence:

$$\begin{aligned}\|z\|_p \|x\|_q &= \left(\sum_{i=1}^n |x_i|^q\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |x_i|^q\right)^{\frac{1}{q}} \\ &= \sum_{i=1}^n |x_i|^q \\ &= \|x\|_q^q\end{aligned}$$

Showing that $z^T x = \|z\|_p \|x\|_q$. Letting $y = \frac{z}{\|z\|_p}$ gives that $\|y\|_p = 1$, and $y^T x = \|y\|_p \|x\|_q$, so that the supremum in Hölder's inequality is achieved, showing that $\|\cdot\|_p^* = \|\cdot\|_q$.

1.3.3 Boyd and Vandenberghe, example 3.26

Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ be a norm, and $f(x) = \|\cdot\|$. We will find the convex conjugate $f^*(y) = \sup_x \left(y^T x - f(x)\right)$.

Suppose that $\|y\|^* > 1$. Then, by the fact that the supremum defining the dual norm is achieved, there must exist a z with $\|z\| \leq 1$ such that $\left|z^T y\right| > 1$. Choosing $x = tz$ gives that:

$$y^T x - f(x) = t y^T z - |t| \|z\|$$

Since $m = \left|z^T y\right| > 1$, we may make $t y^T z$ greater than $|t|$ by choosing the sign of t appropriately, showing that there exists a $t > 1$ such that:

$$y^T x - f(x) \geq t(m - \|z\|)$$

Since $\|z\| \leq 1$ and $m > 1$, we will have that $m - \|z\| > 0$, so that the above may be made arbitrarily large, by choosing t large. Hence, if $\|y\|^* > 1$:

$$f^*(y) = \infty$$

Now suppose that $\|y\|^* \leq 1$. Then, by the analogue to Hölder's inequality, we'll have that $\left|y^T x\right| \leq \|y\|^* \|x\| \leq \|x\|$ for all x. Hence:

$$\begin{aligned}y^T x - f(x) &\leq \|x\| - \|x\| \\ &\leq 0\end{aligned}$$

Choosing $x = \vec{0}$ gives that $y^T x - f(x) = 0 - \left\|\vec{0}\right\| = 0$, so this shows that, if $\|y\|^* \le 1$:

$$f^*(y) = 0$$

Combining these results:

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

1.3.4 Boyd and Vandenberghe, example 3.27

Now let $f(x) = \frac{1}{2}\|x\|^2$. We will find the convex conjugate $f^*(y) = \sup_x \left(y^T x - f(x)\right)$.

By the analogue to Hölder's inequality, we'll have that $y^T x \le \left|y^T x\right| \le \|y\|^* \|x\|$ for all x. Hence:

$$y^T x - f(x) \le \|y\|^* \|x\| - \frac{1}{2}\|x\|^2$$

This is a quadratic function in $\|x\|$, which is maximized when $\|x\| = \|y\|^*$. Of course, an x satisfying this condition must exist, so:

$$\begin{aligned} f^*(y) &= \sup_x \left(y^T x - f(x)\right) \\ &\le \frac{1}{2}\left(\|y\|^*\right)^2 \end{aligned}$$

For the other direction, let y be fixed, and choose z with $\|z\| = 1$ which maximizes the supremum $\|y\|^* = \sup_{\|z\|=1}\left|z^T y\right|$ (recall from section 1.3.1 that such a z must exist). Let $x = \operatorname{sign}\left(z^T y\right)\|y\|^* z$, so that:

$$\begin{aligned} y^T x - f(x) &= \operatorname{sign}\left(z^T y\right)\|y\|^* y^T z - \frac{1}{2}\|x\|^2 \\ &= \|y\|^* \left|y^T z\right| - \frac{1}{2}\left(\|y\|^*\right)^2 \\ &= \left(\|y\|^*\right)^2 - \frac{1}{2}\left(\|y\|^*\right)^2 \\ &= \frac{1}{2}\left(\|y\|^*\right)^2 \end{aligned}$$

Since such an x exists for every y, this shows that:

$$\begin{aligned} f^*(y) &= \sup_x \left(y^T x - f(x)\right) \\ &\ge \frac{1}{2}\left(\|y\|^*\right)^2 \end{aligned}$$

Combining these results:

$$f^*(y) = \frac{1}{2}\left(\|y\|^*\right)^2$$

2 Duality

Suppose that we wish to optimize the following problem:

$$\begin{aligned} \text{minimize} &: f_0(x) \\ \text{subject to} &: f_i(x) \le 0 \\ & \quad h_i(x) = 0 \end{aligned}$$

With $\mathrm{dom} f_i = \mathrm{dom} h_i = \mathbb{R}^n$. The Lagrangian is:

$$L(x, \lambda, \nu) = f_0(x) + \sum_i \lambda_i f_i(x) + \sum_i \nu_i h_i(x)$$

When $\lambda \succeq 0$ and x is feasible (so that $f_i(x) \le 0$ for $i > 0$ and $h_i(x) = 0$):

$$\begin{aligned} L(x, \lambda, \nu) &= f_0(x) + \sum_i \lambda_i f_i(x) \\ &\le f_0(x) \end{aligned}$$

The Lagrange dual function as:

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

And since for any $\lambda \succeq 0$, we have that $L(x, \lambda, \nu) \le f_0(x)$ for all feasible x, it follows that:

$$g(\lambda, \nu) \le p^*$$

When $\lambda \not\succeq 0$ and $g(\lambda, \nu) > -\infty$, we call the point (λ, ν) dual feasible.

2.1 Relation to convex conjugates (Boyd and Vandenberghe, section 5.1.6)

Consider the following problem:

$$\begin{aligned} \text{minimize} &: f_0(x) \\ \text{subject to} &: Ax \preceq b \\ & \quad Cx = d \end{aligned}$$

The Lagrangian will be:

$$\begin{aligned} L(x, \lambda, \nu) &= f_0(x) + \sum_i \lambda_i \left(A_{i,:}^T x - b_i\right) + \sum_i \nu_i \left(C_{i,:}^T x - d_i\right) \\ &= f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \end{aligned}$$

So that the Lagrange dual will be:

$$\begin{aligned} g(\lambda,\nu) &= \inf_x \left(f_0(x) + \lambda^T A x + \nu^T C x\right) - \lambda^T b - \nu^T d \\ &= -\sup_x \left(\left(-\lambda^T A - \nu^T C\right) x - f_0(x)\right) - \lambda^T b - \nu^T d \\ &= -f^* \left(-\lambda^T A - \nu^T C\right) - \lambda^T b - \nu^T d \end{aligned}$$

Where f^* is the convex conjugate of f.

2.1.1 Equality constrained norm minimization (Boyd and Vandenberghe, page 221)

Consider the following problem:

$$\begin{aligned} \text{minimize} &: \; f_0(x) = \|x\| \\ \text{subject to} &: \; Ax = b \end{aligned}$$

From the above:

$$g(\lambda,\nu) = -f_0^* \left(-\nu^T A\right) - \nu^T b$$

While from section 1.3.3:

$$f_0^*(y) = \begin{cases} 0 & \text{if } \|y\|^* \leq 1 \\ \infty & \text{if } \|y\|^* > 1 \end{cases}$$

Hence:

$$g(\lambda,\nu) = \begin{cases} -\nu^T b & \text{if } \left\|\nu^T A\right\|^* \leq 1 \\ -\infty & \text{if } \left\|\nu^T A\right\|^* > 1 \end{cases}$$

2.1.2 Entropy maximization (Boyd and Vandenberghe, page 222)

Consider the following problem:

$$\begin{aligned} \text{minimize} &: \; f_0(x) = \sum_{i=1}^{n} x_i \log_2 x_i \\ \text{subject to} &: \; Ax \preceq b \\ &: \; \vec{1}^T x = 1 \end{aligned}$$

Note that there is an implicit constraint that $x \succeq 0$, if we adopt the convention that $0 \log 0 = 0$. This problem can be interpreted as that of finding a multinomial distribution (the constraints $x \succ 0$ and $\vec{1}^T x = 1$ force x to be a probability distribution), subject to a set of linear conditions on each of the probabilities, which maximizes the entropy (we minimize the negative entropy).

From the above:

$$g(\lambda,\nu) = -f_0^* \left(-\lambda^T A - \nu\vec{1}\right) - \lambda^T b - \nu$$

By section 1.1, the convex conjugate of $\sum_{i=1}^{n} x_i \log_2 x_i$ satisfies $f_0^*(y) = \sum_{i=1}^{n} h^*(x_i)$, where h^* is the convex conjugate of $h(x) = x \log_2 x$. The convex conjugate of $h(x)$ satisfies $h^*(y) = \sup_y (yx - x \log_2 x)$. For fixed y, note that $yx - x \log_2 x$ is a continuous function of x. Also, $\lim_{x \to \infty} (yx - x \log_2 x) = -\infty$, and for $x = 0$ we have that $yx - x \log_2 x = 0$. Hence, this function is maximized where $\frac{d}{dx}(yx - x \log_2 x) = 0$, provided that the function value at such a point (if it exists) is at least 0. Since $\frac{d}{dx}(yx - x \log_2 x) = y - \log_2 x - 1$, we have that $x = 2^{y-1} \geq 0$, so that $h^*(y) = y2^{y-1} - 2^{y-1} \log_2 2^{y-1} = 2^{y-1}$. Therefore, $f_0^*(y) = \sum_{i=1}^{n} 2^{y_i - 1}$:

$$
\begin{aligned}
g(\lambda, \nu) &= -\sum_{i=1}^{n} 2^{-\left(\lambda^T A\right)_i - \nu - 1} - \lambda^T b - \nu \\
&= -2^{-\nu-1} \sum_{i=1}^{n} 2^{-\left(\lambda^T A\right)_i} - \lambda^T b - \nu
\end{aligned}
$$