Mathematical Tools in Computer Science, Winter 2006

Lecture 2

Lecturer: Nati Linial *Scribe: Shai Shalev-Shwartz*

The structure of this lecture is as follows. We start with a short review of the equivalence between norms and convex bodies which are symmetric relative to the origin. We then define the notion of dual norms from several perspectives.

1 Norms and Dual Norms

1.1 Norms and unit balls

In the previous lecture we defined a vector norm, $\|\mathbf{x}\|$, as a function that satisfies:

1. Linearity: $\forall\ \lambda \in \mathbb{R}, \quad \|\lambda \mathbf{x}\| = |\lambda|\, \|\mathbf{x}\|$
2. Triangle inequality: $\forall\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^n, \quad \|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$
3. Positivity: $\forall\ \mathbf{v} \in \mathbb{R}^n, \quad \|\mathbf{v}\| \ge 0$ and equality holds iff $\mathbf{v} = \mathbf{0}$.

A norm $\|\cdot\|$ defines a unit ball, $B = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \le 1\}$. B is a convex set, which means that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are in B then all the line defined by $\{\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 : \lambda \in [0,1]\}$ is in B as well. This property follows from the triangle inequality and from the linearity of the norm,

$$\|\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\| \;\le\; \|\lambda \mathbf{x}_1\| + \|(1-\lambda)\mathbf{x}_2\| \;=\; \lambda\|\mathbf{x}_1\| + (1-\lambda)\|\mathbf{x}_2\| \;\le\; \lambda + (1-\lambda) \;=\; 1 \quad .$$

In addition, B is symmetric with respect to the origin. That is, if $\mathbf{x} \in B$ then $-\mathbf{x} \in B$ (this follows directly from the linearity of the norm and the definition of B). Finally, from the positivity condition we get that B is full dimensional (i.e $B \neq \{0\}$).

The other side is also true. That is, if we have a convex set B, which is symmetric around the origin, and which is full dimensional, it defines a norm as follows. We set $\|\mathbf{0}\|_B = 0$ and for each $\mathbf{x} \neq \mathbf{0}$, we define $\|\mathbf{x}\|_B = \frac{1}{\max\{\lambda > 0 : \lambda \mathbf{x} \in B\}}$.

1.2 Dual norms and dual sets

Given a norm $\|\cdot\|$ we define its dual norm as,

$$\|\mathbf{x}\|^\star \;=\; \max_{\mathbf{v} \neq 0} \frac{|\langle \mathbf{x}, \mathbf{v}\rangle|}{\|\mathbf{v}\|} \quad . \tag{1}$$

From the linearity of the norm and of the inner product, we can rewrite the above as,

$$\|\mathbf{x}\|^\star \;=\; \max_{\mathbf{v}\,:\,\|\mathbf{v}\|=1} |\langle \mathbf{x}, \mathbf{v}\rangle| \quad . \tag{2}$$

First, let us prove that this indeed defines a norm. The linearity property follows from the linearity of the inner product and the positivity property follows directly from the definition. To prove the triangle inequality, we use the triangle inequality on scalars as follows,

$$\begin{aligned}
\|\mathbf{x}_1 + \mathbf{x}_2\|^\star &= \max_{\mathbf{v}\,:\,\|\mathbf{v}\|=1} |\langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{v}\rangle| \\
&= \max_{\mathbf{v}\,:\,\|\mathbf{v}\|=1} |\langle \mathbf{x}_1, \mathbf{v}\rangle + \langle \mathbf{x}_2, \mathbf{v}\rangle| \\
&\leq \max_{\mathbf{v}\,:\,\|\mathbf{v}\|=1} (|\langle \mathbf{x}_1, \mathbf{v}\rangle| + |\langle \mathbf{x}_2, \mathbf{v}\rangle|) \\
&\leq \max_{\mathbf{v}\,:\,\|\mathbf{v}\|=1} |\langle \mathbf{x}_1, \mathbf{v}\rangle| + \max_{\mathbf{v}\,:\,\|\mathbf{v}\|=1} |\langle \mathbf{x}_2, \mathbf{v}\rangle| \\
&= \|\mathbf{x}_1\|^\star + \|\mathbf{x}_2\|^\star \quad .
\end{aligned}$$

We can interpret the dual norm as a special case of matrix operator norms. To see this, we can think on a vector $\mathbf{x} \in \mathbb{R}^d$ as a $1 \times d$ matrix (a linear functional) that defines the linear transformation $f_x(\mathbf{v}) = \langle \mathbf{x}, \mathbf{v}\rangle$. Using this notation, the dual norm of the vector $\mathbf{x}$ from Eq. (1) is exactly the matrix-norm of the $1 \times d$ matrix $\mathbf{x}$.

Recall that a norm $\|\cdot\|$ is completely defined by its unit ball B and that B is a convex set. In general, for a convex set B we define its dual convex set as,

$$B^\star \;=\; \{\mathbf{x} \;:\; \forall \mathbf{v} \in B,\ \langle \mathbf{x}, \mathbf{v}\rangle \leq 1\} \quad .$$

The following lemma connects the definitions of dual sets and norms.

Lemma 1 *Let $\|\cdot\|$ be a norm and let B be its unit ball. Let $B^\star$ be the dual set of B and let $\|\cdot\|^\star$ be the dual norm of $\|\cdot\|$. Then, $B^\star$ is the unit ball of $\|\cdot\|^\star$.*

Proof Let $\mathbf{x} \in B^\star$. Then, for all $\mathbf{v} \in B$ we have that $\langle \mathbf{x}, \mathbf{v}\rangle \leq 1$. In particular, for all $\mathbf{v}$ such that $\|\mathbf{v}\| = 1$ we have that $\langle \mathbf{x}, \mathbf{v}\rangle \leq 1$. We therefore get from Eq. (2) that $\|\mathbf{x}\|^\star \leq 1$. Now, assume that $\mathbf{x} \notin B^\star$. Then, there exists a vector $\mathbf{v} \in B$ and a scalar $a > 1$ such that $\langle \mathbf{x}, \mathbf{v}\rangle = a$. From the linearity of the norm we can assume that $\|\mathbf{v}\| = 1$. Therefore, from Eq. (2) we get that $\|\mathbf{x}\|^\star \geq a > 1$. In summary, a vector $\mathbf{x}$ is in $B^\star$ iff $\|\mathbf{x}\|^\star \leq 1$, which concludes our proof. ■

The above lemma leads to a generalization of Hölder inequality.

Lemma 2 *Let $\|\cdot\|$ be a norm and let $\|\cdot\|^\star$ be its dual norm. Then, for any two vectors $\mathbf{x}$ and $\mathbf{v}$ we have that,*

$$|\langle \mathbf{x}, \mathbf{v}\rangle| \;\leq\; \|\mathbf{x}\|\,\|\mathbf{v}\|^\star \quad .$$

Proof Let B and $B^\star$ be the unit balls of $\|\cdot\|$ and $\|\cdot\|^\star$. First note that from the definition of B we have that $\mathbf{x}/\|\mathbf{x}\| \in B$ and $\mathbf{v}/\|\mathbf{v}\|^\star \in B^\star$. The definition of $B^\star$ now implies that

$$|\langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{v}}{\|\mathbf{v}\|^\star}\rangle| \;\leq\; 1 \quad .$$

The claim in the lemma now follows from the linearity of the inner product. ■

1.3 The p-norm

For each $p \geq 1$ we define the p-norm of a vector $\mathbf{v} \in \mathbb{R}^d$ to be,

$$\|\mathbf{v}\|_p \;=\; \left(\sum_{k=1}^{d} |v_i|^p\right)^{\frac{1}{p}} \quad .$$

We now show that $\|\cdot\|_p^\star = \|\cdot\|_q$, where

$$\frac{1}{p} + \frac{1}{q} \;=\; 1 \quad .$$

To show this,[1] we denote by B_p the unit ball of $\|\cdot\|_p$, and similarly, B_q denotes the unit ball of $\|\cdot\|_q$. It is sufficient to show that $B_q = B_p^\star$.

($\subseteq$) : Let $\mathbf{x} \in B_q$. Then, from Hölder inequality (defined and proved in the next subsection) we get that $\langle \mathbf{x}, \mathbf{v} \rangle \leq \|\mathbf{x}\|_q \|\mathbf{v}\|_p \leq \|\mathbf{v}\|_p$. Thus, for each $\mathbf{v} \in B_p$ we get that, $\langle \mathbf{x}, \mathbf{v} \rangle \leq 1$ which implies that $\mathbf{x} \in B_p^\star$.

($\supseteq$) : Let $\mathbf{x} \in B_p^\star$. Then, for each $\mathbf{v}$ such that $\|\mathbf{v}\|_p \leq 1$ we have that $\langle \mathbf{x}, \mathbf{v} \rangle \leq 1$. Define $\mathbf{v} \in \mathbb{R}^d$ to be the vector for which

$$v_k = \frac{\mathrm{sign}(x_k)\,(x_k)^{q-1}}{\|\mathbf{x}\|_q^{q/p}} \quad .$$

Note that $p\,q = p + q$ and thus,

$$\|\mathbf{v}\|_p^p \;=\; \frac{1}{\|\mathbf{x}\|_q^q} \sum_{k=1}^{d} |x_k|^{(q-1)p} \;=\; \frac{1}{\|\mathbf{x}\|_q^q} \sum_{k=1}^{d} |x_k|^q \;=\; 1 \quad .$$

Therefore,

$$1 \;\geq\; \langle \mathbf{x}, \mathbf{v} \rangle \;=\; \sum_{k=1}^{d} \mathrm{sign}(x_k)\, x_k^{1+q-1} \;=\; \sum_{k=1}^{d} |x_k|^q \;=\; \|\mathbf{x}\|_q^q \quad ,$$

which implies that $\|\mathbf{x}\|_q \leq 1$ and thus $\mathbf{x} \in B_q$.

1.4 Young, Hölder and Minkovsky inequalities

Lemma 3 (Young inequality) *Let $p, q \in \mathbb{R}$ be two positive scalars such that $\frac{1}{p} + \frac{1}{q} = 1$. Then,*

$$ab \;\leq\; \frac{1}{p} a^p + \frac{1}{q} b^q \quad .$$

[1] If $p = 1$ then $q = \infty$ and we define $\|\mathbf{x}\|_\infty = \max_i |x_i|$. In this case, one can directly show that $\|\cdot\|_\infty$ is the dual norm of $\|\cdot\|_1$.

Proof The log function is a concave function. That is, for each $\lambda \in [0,1]$ and scalars a, b we have that, $\log(\lambda a + (1-\lambda)b) \geq \lambda \log(a) + (1-\lambda)\log(b)$. Therefore,

$$\log(a\,b) = \frac{1}{p}\log(a^p) + \frac{1}{q}\log(b^q) \leq \log\left(\frac{1}{p}a^p + \frac{1}{q}b^q\right) \quad .$$

Taking exponent of both sides gives the desired inequality. ■

Lemma 4 (Hölder inequality) *Let $p, q \in \mathbb{R}$ be two positive scalars such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$|\langle \mathbf{x}, \mathbf{v} \rangle| \leq \|\mathbf{x}\|_p \, \|\mathbf{v}\|_q \quad .$$

Proof The proof uses the triangle inequality $|a + b| \leq |a| + |b|$ and Young inequality as follows,

$$\begin{aligned}
\frac{|\langle \mathbf{x}, \mathbf{v} \rangle|}{\|\mathbf{x}\|_p \, \|\mathbf{v}\|_q} &= \frac{\left|\sum_{k=1}^d x_k v_k\right|}{\|\mathbf{x}\|_p \, \|\mathbf{v}\|_q} \leq \frac{\sum_{k=1}^d |x_k| \, |v_k|}{\|\mathbf{x}\|_p \, \|\mathbf{v}\|_q} = \sum_{k=1}^d \left(\frac{|x_k|}{\|\mathbf{x}\|_p} \frac{|v_k|}{\|\mathbf{v}\|_q}\right) \\
&\leq \sum_{k=1}^d \left(\frac{1}{p}\left(\frac{|x_k|}{\|\mathbf{x}\|_p}\right)^p + \frac{1}{q}\left(\frac{|v_k|}{\|\mathbf{v}\|_q}\right)^q\right) = \frac{1}{p}\frac{\sum_{k=1}^d |x_k|^p}{\|\mathbf{x}\|_p^p} + \frac{1}{q}\frac{\sum_{k=1}^d |v_k|^q}{\|\mathbf{v}\|_q^q} \\
&= \frac{1}{p} + \frac{1}{q} = 1 \quad .
\end{aligned}$$

■

Finally, the **Minkowski** inequality states that if $p \geq 1$, then, $\|\mathbf{x} + \mathbf{v}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{v}\|_p$. This inequality follows directly from the triangle inequality property of the p-norm.