MS&E 213 / CS 269O
Appendix A - Norms*

By Aaron Sidford (sidford@stanford.edu)

November 18, 2019

Here we review and prove some facts about norms we use throughout the course

1 Norms

We begin by recalling some basic facts about norms.

Definition 1. $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is called a *norm* if the following hold

- *(absolute homogeneity)* $\|\alpha x\| = |\alpha| \cdot \|x\|$ for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$
- *(triangle inequality)* $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$
- *(definiteness)* if $\|x\| = 0$ then $x = \vec{0}$

If the last condition does not hold we say that $\|\cdot\|$ is a *semi-norm*.

We note some immediate consequence of this definition.

Lemma 2. *If $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a norm then $\|\vec{0}\| = 0$ and $\|x\| \geq 0$ for all $x \in \mathbb{R}^n$.*

Proof. The first fact follows from the fact that absolute homogeneity implies

$$\|\vec{0}\| = \|0 \cdot \vec{0}\| = |0| \cdot \|\vec{0}\| = 0$$

and the second follows from the fact that triangle inequality and absolute homogeneity imply

$$\|0\| = \|x - x\| \leq \|x\| + \|-x\| = 2 \cdot \|x\| \,.$$

□

Now for every norm there is a natural induced dual norm given as follows.

Definition 3. For a norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ we define the dual norm $\|\cdot\|_* : \mathbb{R}^n \to \mathbb{R}$ for all $x \in \mathbb{R}^n$ by

$$\|x\|_* = \max_{\|y\| \leq 1} y^\top x \,.$$

*These notes are a work in progress. They are not necessarily a subset or superset of the in-class material and there may also be occasional *TODO* comments which demarcate material I am thinking of adding in the future. These notes will converge to a superset of the class material that is TODO-free. Your feedback is welcome and highly encouraged. If anything is unclear, you find a bug or typo, or if you would find it particularly helpful for anything to be expanded upon, please do not hesitate to post a question on the discussion board or contact me directly at sidford@stanford.edu.

It is not too hard to show that the dual norm is always a norm.

Lemma 4. *For a norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ it's dual norm $\|\cdot\|_*$ is a norm.*

Proof. Note that for all $x \in \mathbb{R}^n$ since for $\widehat{x} = \frac{1}{\|x\|}x$ it is the case $\|\widehat{x}\| = 1$ we have

$$\|x\|_* = \max_{\|z\|\leq 1} z^\top x \geq \widehat{x}^\top x = \frac{\|x\|^2}{\|x\|} .$$

Consequently $\|x\|_* \geq 0$ for all $x \neq 0$ and clearly $\|0\|_* = 0$. Using this we see that

$$\|\alpha \cdot x\|_* = \max_{\|z\|\leq 1} \alpha \cdot z^\top x = \max_{\|z\|\leq 1} |\alpha| \cdot z^\top x = |\alpha| \cdot \|x\|_*$$

where we used that $\max_{\|z\|\leq} z^\top x = \max_{\|z\|\leq 1} z^\top(-x)$ since if $\|z\| \leq 1$ then $\|-z\| \leq 1$. The last property follows from

$$\|x+y\|_* = \max_{\|z\|\leq 1} z^\top(x+y) \leq \max_{\|z\|\leq 1} z^\top x + \max_{\|w\|\leq 1} w^\top y = \|x\|_* + \|y\|_* .$$

□

1.1 Norm Properties

Here we provide a variety of properties of norms. First, we show that a generalization of Cauchy Schwarz holds in any norm.

Lemma 5 (Cauchy Schwarz)**.** *For a norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$, it's dual norm $\|\cdot\|_*$, and all $x, y \in \mathbb{R}^n$ we have*

$$\left|x^\top y\right| \leq \|x\| \cdot \|y\|_* .$$

Proof. If either $x = 0$ or $y = 0$ then the claim is trivial. Otherwise we have

$$x^\top y = \|x\| \left(\frac{x}{\|x\|}\right)^\top y \leq \|x\| \cdot \max_{\|z\|\leq 1} z^\top y = \|x\| \cdot \|y\|_* .$$

Applying the same to $(-x)^\top y$ we have the desired result. □

The dual norm occurs shows up naturally in solving unconstrained minimization problems involving the norm. For example in Chapter 5, we will consider computing steps which are the result of solving the following optimization problem:

$$\min_x f(x_k) + \bigtriangledown f(x_k)^\top (x - x_k) + \frac{L}{2}\| \bigtriangledown f(x_k)\|^2$$

When the norm is ℓ_2 we saw the the value of this minimizer is $f(x_k) - \frac{1}{2L}\| \bigtriangledown f(x_k)\|_2^2$. The following lemma allows us to immediately conclude that an analogous bound holds for arbitrary norms, i.e. that its value is $f(x_k) - \frac{1}{2L}\| \bigtriangledown f(x_k)\|_*^2$.

Lemma 6. *For any norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$, and $\alpha > 0$ we have that*

$$\min_{y\in\mathbb{R}^n} x^\top y + \frac{\alpha}{2}\|y\|^2 = -\frac{1}{2\alpha}\|x\|_*^2$$

Proof. Let z be such that $\|z\| \leq 1$ and $x^\top z = \|x\|_*$. Let $y = -\frac{\|x\|_*}{\alpha} z$. We have that

$$x^\top y + \frac{\alpha}{2}\|y\|^2 = -\frac{\|x\|_*^2}{\alpha} + \frac{1}{2\alpha}\|x\|_2^2 = -\frac{1}{2\alpha}\|y\|_*^2$$

and consequently the left-hand side is at least the right-hand side. However we have that by Cauchy Schwarz

$$x^\top y + \frac{\alpha}{2}\|y\|^2 \geq -\|x\|_*\|y\| + \frac{\alpha}{2}\|y\|^2 \geq \min_\beta -\beta\|x\|_* + \frac{\alpha}{2}\beta^2 = -\frac{1}{2\alpha}\|y\|_*^2$$

□

An alternative way to derive the proof of Lemma 6 is as follows. Let $M = \|y_*\|$ where y_* is a of $x^\top y + \frac{\alpha}{2}\|y\|^2$. Note that

$$\text{argmin}_y x^\top y + \frac{\alpha}{2}\|y\|^2 \in \text{argmin}_{\|y\|\leq M} x^\top y \in M \cdot \text{argmax}_{\|y\|\leq 1}(-x)^\top y\,.$$

Note that this last formula is the formula for the dual norm of $(-x)$. Consequently, we see that there is a minimizer of $x^\top y + \frac{\alpha}{2}\|y\|^2$ is of the form $-\beta \cdot z$ where z is such that $\|z\| = 1$ and $x^\top z = \|x\|_*$ (ignoring the trivial case where $x = 0$). However, since

$$\text{argmin}_\beta x^\top(-\beta z) + \frac{\alpha}{2}\|\beta z\|_*^2 = \text{argmin}_\beta - \beta\|x\|_* + \frac{\alpha\beta^2}{2} = \frac{\|x\|_*}{\alpha}$$

we see that $-(\|x\|_*/\alpha)z$ is a minimizer of $x^\top y + \frac{\alpha}{2}\|y\|^2$ and computing its value gives the proof. Essentially the stated proof just directly shows that this same vector is optimal by Cauchy Schwarz, rather than computing it as done here.

1.2 Examples

Here we provide some examples of norms and their dual.

1.3 Norms induced by PSD Matrices

First we consider norms induced by positive definite matrices. Recall that symmetric $\mathbf{A} \in \mathbb{R}^{n\times n}$ is positive semidefinite (PSD) if an only if $x^\top \mathbf{A} x \geq 0$ for all $x \in \mathbb{R}^n$ and positive definite (PD) if and only if $x^\top \mathbf{A} x > 0$ for all $x \neq 0$. With this definition in hand we define the following norm and prove it is a norm.

Definition 7. For any symmetric PSD matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ let $\|x\|_{\mathbf{A}} \stackrel{\text{def}}{=} \sqrt{x^\top \mathbf{A} x}$ where $\mathbf{A}$ is PSD.

Lemma 8. *If symmetric matrix* $\mathbf{A}$ *is PSD then* $\|\cdot\|_{\mathbf{A}}$ *is a semi-norm if* $\mathbf{A}$ *and if it is PD then* $\|\cdot\|_{\mathbf{A}}$ *is a norm.*

Proof. Note that clearly $\|\cdot\|_{\mathbf{A}}$ obeys absolute homogeneity in each case. Further, note that since $\mathbf{A}$ is PSD it has a square-root, i.e. there is a unique matrix $\mathbf{A}^{1/2}$that is symmetric, PSD, and obeys that $(\mathbf{A}^{1/2})(\mathbf{A}^{1/2}) = \mathbf{A}$. Note that this is the matrix with the same eigenvalues and eigenspaces as $\mathbf{A}$ where every eigenvalue is replaced with its square-root, i.e. if $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top$ for orthonormal $\mathbf{U}$ and diagonal $\mathbf{\Sigma}$ with positive entries on the diagonal then $\mathbf{A}^{1/2} = \mathbf{U}\mathbf{\Sigma}^{1/2}\mathbf{U}^\top$ where $\mathbf{\Sigma}^{1/2}$ is the diagonal matrix with $\mathbf{\Sigma}_{ii}^{1/2} = \sqrt{\mathbf{\Sigma}_{ii}}$. Now note that for all $x \in \mathbb{R}^n$ we have $\|x\|_{\mathbf{A}} = \|\mathbf{A}^{1/2}x\|_2$. Consequently, for all $x, y \in \mathbb{R}^n$ we have

$$\|x + y\|_{\mathbf{A}} = \|\mathbf{A}^{1/2}x + \mathbf{A}^{1/2}y\|_2 \leq \|\mathbf{A}^{1/2}x\|_2 + \|\mathbf{A}^{1/2}y\|_2 = \|x\|_{\mathbf{A}} + \|y\|_{\mathbf{A}}\,.$$

Where we used that $\|\cdot\|_2$ is a norm. Consequently, triangle inequality holds.

Lastly, note that if $\mathbf{A}$ is positive definite then $\|x\|_{\mathbf{A}} = \sqrt{x^\top \mathbf{A} x} > 0$ for all $x \neq 0$ and therefore $\|x\|_{\mathbf{A}} = 0$ implies $x = 0$. □

1.4 p-Norms

Here, we consider the p-norms define as follows.

Definition 9. For $p \geq 1$ and $x \in \mathbb{R}^n$ we let $\|x\|_p \stackrel{\text{def}}{=} (\sum_{i \in [n]} x_i^p)^{1/p}$ and let $\|x\|_\infty \stackrel{\text{def}}{=} \max_{i \in [n]} |x_i|$

Below we prove that the p-norms are in fact norms and compute their duals. First, we we provide the following technical lemma to help the proof.

Lemma 10 (Young's Inequality)**.** *For all $a, b \geq 0$ and $p, q \in \mathbb{R}_{>0}$ with $\frac{1}{p} + \frac{1}{q} = 1$ we have that*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \,.$$

Proof. Suppose without loss of generality that $a, b > 0$ as the claim is trivial if $a = 0$ or $b = 0$. Now, note that $f(x) = -\log(x)$ is convex for $x > 0$ as $\frac{d}{dx} f(x) = -\frac{1}{x}$ and therefore $\frac{d^2}{d^2 x} f(x) = \frac{1}{x^2} \geq 0$ for such x. Consequently, letting $t = \frac{1}{p}$ we have that by convexity

$$f(t \cdot a^p + (1-t) \cdot b^q) \leq t \cdot f(a^p) + (1-t) \cdot f(b^q)$$

Using the definition of f, t, and negating each side yields that

$$\frac{1}{p} \cdot \log(a^p) + \frac{1}{q} \log b^q \leq \log\left(\frac{1}{p} a^p + \frac{1}{q} b^q\right)$$

Since the left hand side of the above inequality is $\log(ab)$, exponentiating both sides yields the claim. □

Using the above lemma we prove Holder's inequality, which we will see is simply Cauchy Schwarz for the p-norm.

Lemma 11 (Holder's Inequality)**.** *For all $x, y \in \mathbb{R}^n$ and $p, q \in \mathbb{R}_{>0}$ with $\frac{1}{p} + \frac{1}{q} = 1$ we have that*

$$x^\top y \leq \|x\|_p \cdot \|y\|_q \,.$$

Proof. Note that if either $\|x\| = 0$ or $\|y\| = 0$ then $x = 0$ or $y = 0$ and the claim is trivial. Otherwise, let $a = x / \|x\|_p$ and $b = y / \|y\|_q$. By Young's inequality 10, we have that

$$\begin{aligned} a^\top b &\leq \sum_{i \in [n]} |a_i| \cdot |b_i| \leq \sum_{i \in [n]} \frac{|a_i|^p}{p} + \sum_{j \in [n]} \frac{|b_i|^q}{q} = \frac{1}{p} \|a\|^p + \frac{1}{q} \|b\|^q \\ &= \frac{1}{p} \cdot \frac{\|a\|^p}{\|a\|^p} + \frac{1}{q} \cdot \frac{\|b\|^q}{\|b\|^q} = \frac{1}{p} + \frac{1}{q} = 1 \,. \end{aligned}$$

Consequently, the result follows from the fact that

$$a^\top b = \frac{x^\top y}{\|x\|_p \|y\|_q} \,.$$

□

Using Holder's inequality we can show that all the p-norms are in fact norms.

Lemma 12 (Holder's Inequality (Cauchy Schwarz for ℓ_p-norm))**.** *For all $p \geq 1$ then $\|x\|_p \stackrel{\text{def}}{=} \left(\sum_{i \in [n]} x_i^p\right)^{1/p}$ is a norm.*

Proof. Note that clearly $\|\cdot\|_p$ obeys homogeneity and definiteness and consequently it only remains to prove triangle-inequality. For this, let $x, y \in \mathbb{R}^n$ be arbitrary and note that

$$\|x+y\|_p^p = \sum_{i\in[n]} |x_i + y_i|^p \leq \sum_{i\in[n]} (|x_i| + |y_i|) \cdot |x_i + y_i|^{p-1} .$$

Now, let q be such that $\frac{1}{p} + \frac{1}{q} = 1$. Note that $pq = p + q$ and therefore $(p-1)q = p$ and therefore by Cauchy Schwarz

$$\begin{aligned}\sum_{i\in[n]} (|x_i| + |y_i|) \cdot |x_i + y_i|^{p-1} &= \sum_{i\in[n]} |x_i| \cdot |x_i + y_i|^{p-1} + \sum_{i\in[n]} |x_i| \cdot |x_i + y_i|^{p-1} \\ &\leq \|x\|_p \|x+y\|_p^{p/q} + \|y\|_p \|x+y\|_p^{p/q} .\end{aligned}$$

Combining, and considering the case when $\|x+y\|_p = 0$ separately, yields that

$$\|x+y\|_p = \|x+y\|_p^{p-(p/q)} \leq \|x\|_p + \|y\|_p .$$

□

Leveraging this we can also formally show that the p-norm and q-norm are dual when $\frac{1}{p} + \frac{1}{q} = 1$.

Lemma 13. *For $p, q \geq 0$ if $\frac{1}{p} + \frac{1}{q} = 1$ then $\|\cdot\|_p$ and $\|\cdot\|_q$ are the dual norms of each other.*

Proof. Let $x \in \mathbb{R}^n$ be arbitrary. By symmetry between p and q, it suffices to show that

$$\max_{y\in\mathbb{R}^n \,|\, \|y\|_p \leq 1} x^\top y = \|x\|_q .$$

Note that this is trivial in the case that $x = 0$ and clearly by Cauchy Schwarz

$$\max_{y\in\mathbb{R}^n \,|\, \|y\|_p \leq 1} x^\top y \leq \max_{y\in\mathbb{R}^n \,|\, \|y\|_p \leq 1} \|y\|_p \|x\|_q \leq \|x\|_q .$$

Consequently, it remains to show that for $x \neq 0$ there is a y with $\|y\|_p \leq 1$ such that $x^\top y = \|x\|_q$. Let $z \in \mathbb{R}^n$ be defined with $z_i = \text{sign}(x_i) \cdot |x_i|^{q-1}$ where $\text{sign}(x_i) = 1$ if $x_i > 1$, $\text{sign}(x_i) = -1$ if $x_i < 0$, and $\text{sign}(x_i) = 0$ otherwise and let $y \stackrel{\text{def}}{=} z/\|z\|_p$. Note that since $x \neq 0$ we have $z \neq 0$ and therefore $\|z\|_p > 0$ and this is well defined. However,$\|y\|_p = 1$ by construction and then

$$x^\top y = \frac{\sum_{i\in[n]} |x_i|^q}{\left(\sum_{i\in[n]} |x_i|^{(q-1)p}\right)^{1/p}} = \frac{\|x\|_q^q}{\|x\|_q^{q/p}} = \|x\|_q$$

where we used that $(q-1)p = q$ and $q - \frac{q}{p} = 1$. Consequently y has the desired properties. □