CPSC 540 Notes on Norms

Mark Schmidt

January 18, 2019

1 General Norms

In class we've used the ℓ_2-norm and the ℓ_1-norm as a measure of the length of a vector, and the concept of a norm generalizes this idea. In particular, we say that a function f is a *norm* if it satisfies the following three properties:

1	$f(x) = 0$ implies $x = 0$	(separate points)
2	$f(\alpha x) = \lvert\alpha\rvert f(x)$, for scalar α	(absolute homogeneity)
3	$f(x + y) \leq f(x) + f(y)$	(triangle inequality)

An important implication of these properties are that norms are non-negative,

$$f(x) \geq 0,$$

and norms are convex. All possible norms are said to be *equivalent*, in that sense that if f and g are norms then there exist constants β_1 and β_2 such that

$$\beta_1 f(x) \leq g(x) \leq \beta_2 f(x).$$

2 General ℓ_p-norms

The most important class of norms on vectors are the ℓ_p-norms, defined by

$$\|x\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p},$$

for some $p \geq 1$. The three most important special cases are the ℓ_1-norm, ℓ_2-norm, and ℓ_∞ norm,

$$\|x\|_1 = \sum_{i=1}^d |x_i|, \quad \|x\|_2 = \sqrt{\sum_{i=1}^d x_i^2}, \quad \|x\|_\infty = \max_i\{|x_i|\}.$$

We recognize the ℓ_2-norm as the standard straight-line distance in Euclidean-space, and it is often simply denoted $\|x\|$. The constants that relate these norms are

$$\|x\|_\infty \leq \|x\|_1 \leq d\|x\|_\infty$$
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{d}\|x\|_\infty$$
$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{d}\|x\|_2.$$

Some works use $p < 1$ (e.g., as a regularizer that encourages a higher degree of sparsity than the ℓ_1-norm), but this does not satisfy the properties of a norm (e.g., $\|x\|_p$ for $p < 1$ is not convex). Some works also define the zero 'norm' to be the number of non-zero elements, $\|x\|_0 = \sum_{i=1}^d I(x_i \neq 0)$, and again this is not an actual norm.

Another generalization of the Euclidean norm is the set of quadratic norms. Given any positive-definite matrix H, quadratic norms are defined by

$$\|x\|_H = \sqrt{x^T H x}.$$

3 Cauchy-Schwartz and Hölder Inequalities, Dual Norm

The Cauchy-Schwartz inequality bounds inner products by the product of Euclidean norms,

$$\sum_{i=1}^d |x_i| \cdot |y_i| \leq \|x\| \cdot \|y\|,$$

which implies

$$x^T y \leq \|x\|\|y\|,$$

and (according to Wikipedia) is one of the most important inequalities in mathematics. A generalization is Hölder's inequality,

$$\sum_{i=1}^d |x_i| \cdot |y_i| \leq \|x\|_p \cdot \|y\|_q,$$

where

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

Cauchy-Schwartz is the special case where $p = q = 2$, but we could also have $p = 1$ and $q = \infty$. The norm $\|\cdot\|_q$ is said to be the *dual norm* of $\|\cdot\|_p$. The general definition of the dual norm $\|\cdot\|_*$ for a general norm $\|\cdot\|$ is

$$\|y\|_* = \sup_x \{y^T x \mid \|x\| \leq 1\}.$$

4 Matrix Norms

Above we define norms on vectors, but we can also define norms on matrices. In this case we say a function f on matrices is a *norm* if it satifies the same three properties:

1	$f(X) = 0$ implies $X = 0$	(separate points)
2	$f(\alpha X) = \lvert\alpha\rvert f(X)$, for scalar α	(absolute homogeneity)
3	$f(X + Y) \leq f(X) + f(Y)$	(triangle inequality)

Two important classes of norms are "entry-wise" matrix norms and "induced" matrix norms. Entry-wise matrix norms treat all of the elements of the matrix as one big vector, and apply a vector norm. For example, if we use the L2-norm of elements of the elements in a matrix we obtain

$$\|X\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^d x_{ij}^2},$$

which is called the Frobenius norm. The induced matrix norm, with respect to a paritcular vector norm, measures how multiplying by the matrix can change a vector's norm. For p-norms we define the induced matrix norm by

$$\|X\|_p = \sup \left\{ \frac{\|Xw\|_p}{\|w\|_p} \right\}.$$

In the case of the vector L2-norm, the induced matrix norm is the maximum singular value of X,

$$\|X\|_2 = \sigma_1(X).$$