Notes for 2015-02-02

Vector norms

A norm is a scalar-valued function from a vector space into the real numbers with the following properties:

1. *Positive-definiteness*: For any vector x, $\|x\| \geq 0$; and $\|x\| = 0$ iff $x = 0$
2. *Triangle inequality*: For any vectors x and y, $\|x + y\| \leq \|x\| + \|y\|$
3. *Homogeneity*: For any scalar α and vector x, $\|\alpha x\| = |\alpha|\|x\|$

We will pay particular attention to three norms on $\mathbb{R}^n$ and $\mathbb{C}^n$:

$$\|v\|_1 = \sum_i |v_i|$$
$$\|v\|_\infty = \max_i |v_i|$$
$$\|v\|_2 = \sqrt{\sum_i |v_i|^2}$$

Of the three properties, the triangle inequality is usually the one that takes the most work.

If $\|\cdot\|$ is a norm and M is any nonsingular square matrix, then $v \mapsto \|Mv\|$ is also a norm. The case where M is diagonal is particularly common in practice.

In finite-dimensional spaces, all norms are equivalent. That is, if $\|\cdot\|$ and $|||\cdot|||$ are two norms on the same space, then there are constants $C_1, C_2 > 0$ such that for all x,

$$C_1\|x\| \leq |||x||| \leq C_2\|x\|.$$

However, C_1 and C_2 are not necessarily small, so it makes sense to choose norms somewhat judiciously. In particular, if different elements of x have different units (furlongs and fortnights), it usually pays to nondimensionalize the problem before doing numerics; and this can be interpreted as choosing a diagonally scaled norm.

Norms and inner products

An *inner product* on a vector space is a function of two vectors with the following properties:

1. *Positive-definiteness*: $\langle x, x\rangle \geq 0$; and $\langle x, x\rangle = 0$ iff $x = 0$
2. *Linearity in the first argument*: $\langle \alpha x + y, z\rangle = \alpha\langle x, z\rangle + \langle y, z\rangle$.
3. *Sesquilinearity*: $\langle v, w\rangle = \overline{\langle w, v\rangle}$.

A vector space with an inner product automatically inherits the norm $\|v\| = \sqrt{\langle v, v\rangle}$. We are used to seeing this as the standard two-norm with the standard Euclidean inner product, but the same concept works with other inner products. Using an inner product and the associated norm, we can also define the notion of the angle between two vectors:

$$\cos(\theta_{v,w}) = \frac{\langle v, w\rangle}{\|v\|\|w\|}.$$

Two special matrix types

The symmetric (Hermitian) positive definite matrices are analogous to the positive real numbers. A matrix $H \in \mathbb{R}^{n\times n}$ is *symmetric positive definite* if $H = H^*$ is symmetric and the quadratic form

$$x \mapsto x^* H x$$

is positive definite (i.e. always non-negative, and zero iff $x = 0$). I have deliberately used x^* and H^* here, since the complex analogue of an SPD matrix is a *Hermitian* positive definite matrix. $H \in \mathbb{C}^{n\times n}$ is Hermitian if $H = H^*$. The SPD and HPD matrices have all positive real eigenvalues, among other nice properties.

In finite-dimensional vector spaces, any inner product can be identified with an SPD or HPD matrix:

$$\langle v, w\rangle_H = w^* H v.$$

Furthermore, any SPD or HPD matrix can be factored as

$$H = M^* M$$

where M is nonsingular. This factorization is not unique, and there are a few useful variants. We will cover one of the most common ones, the Cholesky factorization (in which M is upper triangular) next week.

An *orthogonal* matrix $Q \in \mathbb{R}^{n \times n}$ satisfies $Q^*Q = I$. The complex analogue is called a *unitary* matrix. Orthogonal matrices are special because they preserve Euclidean lengths and angles:

$$\begin{aligned}(Qw)^*(Qu) &= w^*Q^*Qu = w^*u \\ \|Qu\|^2 &= \|u\|^2.\end{aligned}$$

These matrices are also special numerically, as we will see later.

Matrix norms

Spaces of linear maps (or matrices) can also be treated as vector spaces, and the same definition of norms applies. In general, though, we would like to consider norms on spaces of linear maps that are in some way compatible with the norms on the spaces they map between.

If A maps between two normed vector spaces $\mathcal{V}$ and $\mathcal{W}$, the *induced norm* on A is

$$\|A\|_{\mathcal{V},\mathcal{W}} = \sup_{v \neq 0} \frac{\|Av\|_{\mathcal{W}}}{\|v\|_{\mathcal{V}}}.$$

Because norms are homogeneous with respect to scaling, we also have

$$\|A\|_{\mathcal{V},\mathcal{W}} = \sup_{\|v\|_{\mathcal{V}}=1} \|Av\|_{\mathcal{W}}.$$

Note that when $\mathcal{V}$ is finite-dimensional (as it always is in this class), the unit ball $\{v \in \mathcal{V} : \|v\| = 1\}$ is compact, and $\|Av\|$ is a continuous function of v, so the supremum is actually attained.

These operator norms are indeed norms on the space $\mathcal{L}(\mathcal{V}, \mathcal{W})$ of bounded linear maps between $\mathcal{V}$ and $\mathcal{W}$ (or norms on vector spaces of matrices, if you prefer). Such norms have a number of nice properties, not the least of which are the submultiplicative properties

$$\begin{aligned}\|Av\| &\leq \|A\|\|v\| \\ \|AB\| &\leq \|A\|\|B\|.\end{aligned}$$

The first property ($\|Av\| \leq \|A\|\|v\|$) is clear from the definition of the vector norm. The second property is almost as easy to prove:

$$\|AB\| = \max_{\|v\|=1} \|ABv\| \leq \max_{\|v\|=1} \|A\|\|Bv\| = \|A\|\|B\|.$$

The matrix norms induced when $\mathcal{V}$ and $\mathcal{W}$ are supplied with a 1-norm, 2-norm, or ∞-norm are simply called the matrix 1-norm, 2-norm, and ∞-norm. The matrix 1-norm and ∞-norm are given by

$$\|A\|_1 = \max_j \sum_i |A_{ij}|$$
$$\|A\|_\infty = \max_i \sum_j |A_{ij}|.$$

These norms are nice because they are easy to compute. Also easy to compute (though it's not an induced operator norm) is the *Frobenius* norm

$$\|A\|_F = \sqrt{\operatorname{tr}(A^*A)} = \sqrt{\sum_{i,j} |A_{ij}|^2}.$$

The Frobenius norm is not an operator norm, but it does satisfy the submultiplicative property (i.e. it is *consistent* with the vector 2-norm).

The 2-norm and the SVD

The matrix 2-norm is very useful, but it is also not so straightforward to compute. The computation of the matrix 2-norm leads us naturally to the *singular value decomposition*, which also often appears in data analysis (PCA). The SVD is a decomposition

$$A = U\Sigma V^T$$

where U and V are orthogonal and Σ is a diagonal matrix with positive diagonal entries, which are conventionally listed in descending order. This is useful for computing norms because we know that multiplication by orthogonal matrices doesn't change the Euclidean length. Therefore,

$$\frac{\|Ax\|}{\|x\|} = \frac{\|U\Sigma V^T x\|}{\|x\|} = \frac{\|\Sigma y\|}{\|y\|} \text{ where } y = V^T x.$$

and $\|A\| = \max_{y \neq 0} \|\Sigma y\|/\|y\| = \sigma_{\max}$, where $\sigma_{\max}$ is the largest of the singular values. Note that if $A = U\Sigma V^T$ and A is invertible, then

$$A^{-1} = V\Sigma^{-1}U^T$$

and $\|A^{-1}\| = \sigma_{\min}^{-1}$.

Here's another way of thinking about the matrix 2-norm, which again leads to singular values. We may not get to this in lecture, but I will mention it anyhow, since it previews some ideas we will need when we tackle optimization later in the class. If A is a real matrix, then we have

$$\begin{aligned} \|A\|_2^2 &= \left(\max_{\|v\|_2=1} \|Av\| \right)^2 \\ &= \max_{\|v\|_2^2=1} \|Av\|^2 \\ &= \max_{v^Tv=1} v^T A^T A v. \end{aligned}$$

This is a constrained optimization problem, to which we will apply the method of Lagrange multipliers: that is, we seek critical points for the functional

$$L(v, \mu) = v^T A^T A v - \mu(v^T v - 1).$$

Differentiate in an arbitrary direction $(\delta v, \delta\mu)$ to find

$$\begin{aligned} 2\delta v^T (A^T A v - \mu v) &= 0, \\ \delta\mu(v^T v - 1) &= 0. \end{aligned}$$

Therefore, the stationary points satisfy the eigenvalue problem

$$A^T A v = \mu v.$$

The eigenvalues of $A^T A$ are non-negative (why?), so we will call them σ_i^2. The positive values σ_i are the singular values of A; and, as we saw a moment ago, the largest of these singular values is $\|A\|_2$.