Week 2: Wednesday, Sep 2

Vector norms

At the end of the last lecture, we discussed three vector norms:

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

Also, note that 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.

There is actually another type of norm that will be important for us later in the class: the norm induced by an inner product. In general, 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. More generally, any Hermitian positive definite matrix H induces an inner product

$$\langle u, v \rangle_H = u^* H v,$$

and thus a norm $\|u\|_H = \sqrt{\langle u, u \rangle_H}$.

Matrix norms

Recall that 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

$$\begin{aligned} \|A\|_1 &= \max_j \sum_i |A_{ij}| \\ \|A\|_\infty &= \max_i \sum_j \max |A_{ij}|. \end{aligned}$$

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.

The 2-norm

The matrix 2-norm is very useful, but it is also not so straightforward to compute. However, it has an interesting characterization. If A is a real matrix, then we have

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

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

$$L(v, \mu) = v^TA^TAv - \mu(v^Tv - 1).$$

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

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

Therefore, the stationary points satisfy the eigenvalue problem

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

The eigenvalues of A^TA are non-negative (why?), so we will call them σ_i^2. The positive values σ_i are called the *singular values* of A, and the largest of these singular values is $\|A\|_2$. We will return to the idea of singular values, and the properties we can infer from them, in the not-too-distant future.

Back to a perturbation theorem

Recall that we started talking about norms so that we could analyze the difference between $\hat{y} = (A + E)x$ and $y = Ax$ when E is "small." By smallness, we will mean that $\|E\| \leq \epsilon\|A\|$ for some small ϵ. Assuming A is invertible, we now bound the relative error $\|\hat{y} - y\|/\|y\|$ in terms of A and ϵ:

$$\frac{\|\hat{y} - y\|}{\|y\|} = \frac{\|Ex\|}{\|y\|} = \frac{\|EA^{-1}y\|}{\|y\|} \leq \|EA^{-1}\| \leq \epsilon\|A\|\|A^{-1}\|.$$

The number $\kappa(A) = \|A\|\|A^{-1}\|$ is a *condition number* that characterizes the (relative) sensitivity of y to (relatively) small changes in the matrix A.