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

\[
\langle u, v \rangle_H = u^* Hv,
\]

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

Matrix norms

Recall that if \(A\) maps between two normed vector spaces \(V\) and \(W\), the \emph{induced norm} on \(A\) is

\[
\|A\|_{V,W} = \sup_{v \neq 0} \frac{\|Av\|_W}{\|v\|_V}.
\]

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

\[
\|A\|_{V,W} = \sup_{\|v\|_V = 1} \|Av\|_W.
\]
Note that when \( V \) is finite-dimensional (as it always is in this class), the unit ball \( \{ v \in V : \| v \| = 1 \} \) 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 \( \mathcal{L}(V, W) \) of bounded linear maps between \( V \) and \( W \) (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

\[
\| Av \| \leq \| A \| \| v \|
\]
\[
\| AB \| \leq \| A \| \| B \|.
\]

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 \( V \) and \( W \) are supplied with a 1-norm, 2-norm, or \( \infty \)-norm are simply called the matrix 1-norm, 2-norm, and \( \infty \)-norm. The matrix 1-norm and \( \infty \)-norm are given by

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

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

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

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
\[ \|A\|_2^2 = \left( \max_{\|v\|_2=1} \|Av\| \right)^2 \]
\[ = \max_{\|v\|_2=1} \|Av\|^2 \]
\[ = \max_{v^Tv=1} v^T A^T Av. \]

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
\[ L(v, \mu) = v^T A^T Av - \mu (v^Tv - 1). \]
Differentiate in an arbitrary direction \((\delta v, \delta \mu)\) to find
\[ 2\delta v^T (A^T Av - \mu v) = 0, \]
\[ \delta \mu (v^Tv - 1) = 0. \]

Therefore, the stationary points satisfy the eigenvalue problem
\[ A^T Av = \mu v. \]

The eigenvalues of \(A^T A\) are non-negative (why?), so we will call them \(\sigma_i^2\). The positive values \(\sigma_i\) 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 \(\epsilon\). Assuming \(A\) is invertible, we now bound the relative error \(\|\hat{y} - y\| / \|y\|\) in terms of \(A\) and \(\epsilon\):
\[ \frac{\|\hat{y} - y\|}{\|y\|} = \frac{\|Ex\|}{\|y\|} = \frac{\|EA^{-1}y\|}{\|y\|} \leq \|EA^{-1}\| \leq \epsilon \|A\| \|A^{-1}\|. \]

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 \(A\).