**Week 9: Friday, Oct 23**

# Eigenvalue perturbations: a 2-by-2 illustration

Consider the matrix

$$A(\epsilon) = \begin{bmatrix} \lambda & 1 \\ \epsilon & \lambda \end{bmatrix}.$$

The characteristic polynomial of $A(\epsilon)$ is $p(z) = z^2 - 2\lambda z + (\lambda^2 - \epsilon)$, which has roots $\lambda \pm \sqrt{\epsilon}$. These eigenvalues are *continuous* functions of $\epsilon$ at $\epsilon = 0$, but they are not differentiable functions. This is a more general phenomenon: an $O(\epsilon)$ perturbation to a matrix with an eigenvalue with multiplicity $m$ usually splits the eigenvalue into $m$ distinct eigenvalues, each of which is moved from the original position by $O(\epsilon^{1/m})$. We expect, then, that it will be difficult to accurately compute multiple eigenvalues of general nonsymmetric matrices in floating point. If we are properly suspicious, we should suspect that *nearly* multiple eigenvalues are almost as troublesome — and indeed they are. On the other hand, while we usually los some accuray when trying to compute nearly multiple eigenvalues, we should not always expect to lose *all* digits of accuracy.

The next lecture or two will be spent developing the perturbation theory we will need in order to figure out what we can and cannot expect from our eigenvalue computations.

# First-order perturbation theory

Suppose $A \in \mathbb{C}^{n \times n}$ has a simple[1] eigenvalue $\lambda$ with corresponding column eigenvector $v$ and row eigenvector $w^*$. We would like to understand how $\lambda$ changes under small perturbations to $A$. If we formally differentiate the eigenvalue equation $Av = v\lambda$, we have

$$(\delta A)v + A(\delta v) = (\delta v)\lambda + v(\delta \lambda).$$

If we multiply this equation by $w^*$, we have

$$w^*(\delta A)v + w^* A(\delta v) = \lambda w^*(\delta v) + w^* v(\delta \lambda).$$

[1] An eigenvalue is simple if it is not multiple.

Note that $w^*A = \lambda w^*$, so that we have

$$w^*(\delta A)v = w^*v(\delta\lambda),$$

which we rearrange to get

$$\delta\lambda = \frac{w^*(\delta A)v}{w^*v}. \tag{1}$$

This formal derivation of the first-order sensitivity of an eigenvalue only goes awry if $w^*v = 0$, which we can show is not possible if $\lambda$ is simple.

We can use formula (1) to get a condition number for the eigenvalue $\lambda$ as follows:

$$\frac{|\delta\lambda|}{|\lambda|} = \frac{|w^*(\delta A)v|}{|w^*v||\lambda|} \leq \frac{\|w\|_2\|v\|_2}{|w^*v|}\frac{\|\delta A\|_2}{|\lambda|} = \sec\theta\frac{\|\delta A\|_2}{|\lambda|}.$$

where $\theta$ is the acute angle between the spaces spanned by $v$ and by $w$. When this angle is large, very small perturbations can drastically change the eigenvalue.

# Gershgorin theory

The first-order perturbation theory outlined in the previous section is very useful, but it is also useful to consider the effects of *finite* (rather than infinitesimal) perturbations to $A$. One of our main tools in this consideration will be Gershgorin's theorem.

Here is the idea. We know that diagonally dominant matrices are nonsingular, so if $A - \lambda I$ is diagonally dominant, then $\lambda$ cannot be an eigenvalue. Contraposing this statement, $\lambda$ can be an eigenvalue only if $A - \lambda I$ is *not* diagonally dominant. The set of points where $A - \lambda I$ is not diagonally dominant is a union of sets $\cup_j G_j$, where each $G_j$ is a *Gershgorin disk*:

$$G_j = B_{\rho_j}(a_{jj}) = \left\{ z \in \mathbb{C} : |a_{jj} - z| \leq \rho_j \text{ where } \rho_j = \sum_{i \neq j} |a_{ij}| \right\}.$$

Our strategy now, which we will pursue in detail next time, is to use similarity transforms based on $A$ to make a perturbed matrix $A + E$ look "almost" diagonal, and then use Gershgorin theory to turn that "almost" diagonality into bounds on where the eigenvalues can be.