Week 12: Wednesday, Nov 11

Note: Because I left town immediately after giving this lecture, the notes have been reconstructed rather after the fact. Even more so than usual, they may not completely reflect what happened in lecture.

More minimax fun

Last time, we discussed the minimax theorem, and we stated (but did not prove) the Cauchy interlace theorem. For completeness (since the proof is only in Golub and Van Loan as a citation), let us give a straightforward proof based on the minimax characterization:

Theorem 1. *Suppose A is real symmetric (or Hermitian), and let W be a matrix with m orthonormal columns. Then the eigenvalues of $L = W^*AW$ interlace the eigenvalues of A; that is, if A has eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and W^*AW has eigenvalues β_j, then*

$$\beta_j \in [\alpha_{n-m+j}, \alpha_j].$$

Proof. Suppose $A \in \mathbb{C}^{n\times n}$ and $L \in \mathbb{C}^{m\times m}$. The matrix W maps $\mathbb{C}^m$ to $\mathbb{C}^n$, so for each k-dimensional subspace $\mathcal{V} \subseteq \mathbb{C}^m$ there is a corresponding k-dimensional subspace of $W\mathcal{V} \subseteq \mathbb{C}^n$. Thus,

$$\beta_j = \max_{\dim \mathcal{V}=k} \left(\min_{0\neq v\in\mathcal{V}} \rho_L(v) \right) = \max_{\dim \mathcal{V}=k} \left(\min_{0\neq v\in W\mathcal{V}} \rho_A(v) \right) \leq \alpha_k$$

and similarly

$$\begin{aligned}\beta_j &= \min_{\dim \mathcal{V}=m-k+1} \left(\max_{0\neq v\in\mathcal{V}} \rho_L(v) \right) = \min_{\dim \mathcal{V}=m-k+1} \left(\max_{0\neq v\in W\mathcal{V}} \rho_A(v) \right) \\ &= \min_{\dim \mathcal{V}=n-(k+(n-m))+1} \left(\max_{0\neq v\in W\mathcal{V}} \rho_A(v) \right) \geq \alpha_{n-m+k}\end{aligned}$$

□

Another application of the minimax theorem is due to Weyl: if we write $\lambda_k(A)$ for the kth largest eigenvalue of a symmetric A, then for any symmetric A and E,

$$|\lambda_k(A+E) - \lambda_k(A)| \leq \|E\|_2.$$

A related theorem is the Wielandt-Hoffman theorem:

$$\sum_{i=1}^{n}(\lambda_i(A+E) - \lambda_i(A))^2 \leq \|E\|_F^2.$$

Both these theorems provide strong information about the spectrum relative to what we have in the nonsymmetric case (e.g. from Bauer-Fike). Not only do we know that each eigenvalue of $A + E$ is close to *some* eigenvalue of A, but we know that we can put the eigenvalues of A and $A + E$ into one-to-one correspondence. So for the eigenvalues in the symmetric case, small backward error implies small forward error!

As an aside, note that if $\hat{v}$ is an approximate eigenvector and $\hat{\lambda} = \rho_A(\hat{v})$ for a symmetric A, then we can find an explicit form for a backward error E such that

$$(A + E)\hat{v} = \hat{v}\hat{\lambda}.$$

by evaluate the residual $r = Av - v\lambda$ and writing $E = rv^* + vr^*$. So in the symmetric case, a small residual implies that we are near an eigenvalue. On the other hand, it says little about the corresponding eigenvector, which may still be very sensitive to perturbations if it is associated with an eigenvalue that is close to other eigenvalues.

Inertia

In Lecture 14, we described the concept of *inertia* of a matrix. The inertia $\nu(A)$ is a triple consisting of the number of positive, negative, and zero eigenvalues of A. *Sylvester's inertia theorem* says that inertia is preserved under nonsingular *congruence* transformations, i.e. transformations of the form

$$M = VAV^T.$$

Congruence transformations are significant because they are the natural transformations for *quadratic forms* defined by symmetric matrices; and the invariance of inertia under congruence says something about the invariance of the shape of a quadratic form under a change of basis. For example, if A is a positive (negative) definite matrix, then the quadratic form

$$\phi(x) = x^T Ax$$

defines a concave (convex) bowl; and $\phi(Vx) = x^T(V^TAV)x$ has the same shape.

As with almost anything else related to the symmetric eigenvalue problem, the minimax characterization is the key to proving Sylvester's inertia theorem. The key observation is that if $M = V^TAV$ and A has k positive eigenvalues, then the minimax theorem gives us a k-dimensional subspace $\mathcal{W}_+$ on which A is positive definite (i.e. if W is a basis, then $z^T(W^TAW)z > 0$ for any nonzero z). The matrix M also has a k-dimensional space on which it is positive definite, namely $V^{-1}\mathcal{W}$. Similarly, M and A both have $(n-k)$-dimensional spaces on which they are negative semidefinite. So the number of positive eigenvalues of M is k, just as the number of positive eigenvalues of A is k.

Solvers for the symmetric eigenvalue problem

Because the symmetric eigenvalue problem has so much structure, there are many more good algorithms to solve it than there are for the nonsymmetric problem. The QR iteration is still a good choice. Reduction to Hessenberg form gives us a *symmetric* Hessenberg matrix $T = Q^TAQ$; the combination of symmetry and Hessenberg shape means this matrix is *tridiagonal*. With a proper shift strategy, QR iteration converges *cubically* for symmetric problems, and if only the eigenvaleus are required, it costs $O(n)$ time per step subsequent to the initial tridiagonal reduction. There are even faster methods based on bisection (using the inertia of $T - \sigma I$ to count the number of eigenvalues greater and less than σ) and based on divide-and-conquer ideas; but since we are running low on time, we will leave those ideas for the students' leisure reading in Golub and Van Loan.