Week 11: Friday, Nov 1

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 eigenvalues 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. The fastest of these algorithms (the so-called "Grail") can compute all the eigenvalues of a tridiagonal matrix in $O(n)$ time, and all the eigenvectors in $O(n^2)$ – which is asymptotically optimal. By comparison, it takes $O(n^3)$ time just to reduce a symmetric matrix to tridiagonal form!

We will not go into a great deal of detail, about these algorithms, but will sketch some of the main ideas below.

Inertia and bisection

One of the useful aspects of inertia is that it can be computed *without* computing eigenvalues. For any symmetric matrix, we can compute the factorization

$$PAP^T = LDL^T$$

where L is unit lower triangular and D is diagonal. Because A and D are congruent to each other, they have the same inertia.

By computing the LDL^T factorization of $A - \sigma I$, we find that we can always count of the number of eigenvalues above and below any given shift.

This suggests a *bisection* strategy for finding eigenvalues of A: to find all the eigenvalues in some interval $[\alpha, \beta]$, recursively subdivide the interval until each subinterval either contains no eigenvalues or exactly one eigenvalue. For an interval that contains exactly one eigenvalue, we can then pin the eigenvalue down using bisection – though, in practice, we would typically switch to using something like shift-invert when we got close enough.

Note that by using methods based on inertia, we can reliably compute just the eigenvalues (and corresponding eigenvectors) in a desired interval without computing the entire spectrum.

Secular equations and divide and conquer

Another solver idea for the symmetric eigenvalue problem is the *divide-and-conquer* approach (originally due to Cuppen). The idea is as follows. Suppose we write a tridiagonal matrix T in block 2-by-2 form as

$$T = \begin{bmatrix} T_{11} & \beta e_m e_1^T \\ \beta e_1 e_m^T & T_{22} \end{bmatrix}.$$

Now suppose we have the decompositions If the diagonal blocks can be decomposed as $T_{11} - \beta e_m e_m^T = Q_1 D_1 Q_1^T$ and $T_{22} - \beta e_1 e_1^T = Q_2 D_2 Q_2^T$. Then we have

$$\begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}^T \begin{bmatrix} T_{11} & \beta e_m e_1^T \\ \beta e_1 e_m^T & T_{22} \end{bmatrix} \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} + \beta v v^T$$

where

$$v = \begin{bmatrix} Q_1^T e_m \\ Q_2^T e_1 \end{bmatrix}.$$

Now, if D is a diagonal matrix, then

$$\begin{aligned} \det(D - \lambda I + \beta v v^T) &= \det(D - \lambda I) \det(I + \beta (D - \lambda I)^{-1} v v^T) \\ &= \det(D - \lambda I)(1 + \beta v^T (D - \lambda I)^{-1} v). \end{aligned}$$

The eigenvalues of $D + vv^T$ are therefore the zeros of the rational function

$$f(\lambda) = 1 + \sum_{j=1}^{n} \frac{\beta v_j^2}{d_j - \lambda},$$

and these zeros interlace the locations of the poles (the entries of the diagonal entries of D). The equation $f(\lambda) = 0$ is sometimes called a *secular equation*. Because we know that there is exactly one root between each pair of adjacent entries of the diagonal matrix D (together with one extra), we can find these roots quite quickly. Thus, we can find all the eigenvalues of a tridiagonal matrix by solving two eigenvalue problems of half the size and then finding the roots of a secular equation