Notes for 2016-10-28

1 Canonical forms

Abstract linear algebra is about vector spaces and the operations on them, independent of any specific choice of basis. But while the abstract view is useful, when we compute, we are concrete, working with the vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ with a standard norm or inner product structure. A choice of basis (or choices of bases) links the abstract view to a particular representation. When working with an abstract inner product space, we often would like to choose an orthonormal basis, so that the inner product in the abstract space corresponds to the standard dot product in $\mathbb{R}^n$ or $\mathbb{C}^n$. Otherwise, the choice of basis may be arbitrary in principle — though, of course, some bases are particularly useful for revealing the structure of the operation.

For any given class of linear algebraic operations, we have *equivalence classes* of matrices that represent the operation under different choices of bases. It is useful to choose a distinguished representative for each of these equivalence classes, corresponding to a choice of basis that renders the structure of the operation particularly clear. These distinguished representatives are known as *canonical forms*. Many of these equivalence relations have special names, as do many of the canonical forms.

For spaces without and with inner product structure, the equivalence relations and canonical forms associated with an operation on $\mathcal{V}$ of dimension n and $\mathcal{W}$ of dimension n are shown in Figure 1. A major theme in the analysis of the Hermitian eigenvalue problem follows from a pun: in the Hermitian case in an inner product space, the equivalence relation for operators (unitary similarity) and for quadratic forms (unitary congruence) are the same thing!.

Abstract	Concrete	Equivalence	Canonical Form
Linear map $\mathcal{A} : \mathcal{V} \to \mathcal{W}$	$w = Av$ $A \in \mathbb{C}^{m \times n}$	$A \sim X^{-1}AY$ $X \in \mathbb{C}^{m \times m}, Y \in \mathbb{C}^{n \times n}$ invertible	$\begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$ $k = \operatorname{rank}(A)$
		$A \sim U^*AV$ $U \in \mathbb{C}^{m \times m}, V \in \mathbb{C}^{n \times n}$ orthogonal	$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_{\min(m,n)})$ SVD
Operator $\mathcal{A} : \mathcal{V} \to \mathcal{V}$	$v' = Av$ $A \in \mathbb{C}^{n \times n}$	$A \sim X^{-1}AX$ (similarity) $X \in \mathbb{C}^{n \times m}$ invertible	Jordan form
		$A \sim U^*AU$ (unitary similarity) $U \in \mathbb{C}^{n \times n}$ unitary	$T \in \mathbb{C}^{n \times n}$ upper triangular Schur form
Quadratic form $\phi : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$	$\phi = v^*Av$ $A = A^* \in \mathbb{C}^{n \times n}$	$A \sim X^*AX$ (congruence) $X \in \mathbb{C}^{n \times m}$ invertible	$\operatorname{diag}(I_{\nu_+}, 0_{\nu_0}, -I_{\nu_-})$ $\nu = (\nu_+, \nu_-, \nu_0) =$ *inertia* of A
		$A \sim U^*AU$ (unitary congruence) $U \in \mathbb{C}^{n \times m}$ unitary	$\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ Eigendecomposition

Table 1: Complex canonical forms (without and with inner product structure)

2 Minimax and interlacing

The Rayleigh quotient is a building block for a great deal of theory. One step beyond the basic characterization of eigenvalues as stationary points of a Rayleigh quotient, we have the *Courant-Fischer minimax theorem*:

Theorem 1. *If $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, then we can characterize the eigenvalues via optimizations over subspaces $\mathcal{V}$:*

$$\lambda_k = \max_{\dim \mathcal{V}=k} \left(\min_{0 \neq v \in \mathcal{V}} \rho_A(v) \right) = \min_{\dim \mathcal{V}=n-k+1} \left(\max_{0 \neq v \in \mathcal{V}} \rho_A(v) \right).$$

Proof. Write $A = U \Lambda U^*$ where U is a unitary matrix of eigenvectors. If v is a unit vector, so is $x = U^* v$, and we have

$$\rho_A(v) = x^* \Lambda x = \sum_{j=1}^n \lambda_j |x_j|^2,$$

i.e. $\rho_A(v)$ is a weighted average of the eigenvalues of A. If $\mathcal{V}$ is a k-dimensional subspace, then we can find a unit vector $v \in \mathcal{V}$ that satisfies the $k-1$ constraints $(U^* v)_j = 0$ for $j = 1$ through $k-1$ (i.e. v is orthogonal to the invariant subspace associated with the first $k-1$ eigenvectors). For this v, $\rho_A(v)$ is a weighted average of $\lambda_k, \lambda_{k+1}, \ldots, \lambda_n$, so $\rho_A(v) \leq \lambda_k$. Therefore,

$$\max_{\dim \mathcal{V}=k} \left(\min_{0 \neq v \in \mathcal{V}} \rho_A(v) \right) \leq \lambda_k.$$

Now, if $\mathcal{V}$ is the range space of the first k columns of U, then for any $v \in \mathcal{V}$ we have that $\rho_A(v)$ is a weighted average of the first k eigenvalues, which attains the minimal value λ_k when we choose $v = u_k$. □

One piece of the minimax theorem is that given any k-dimensional subspace $\mathcal{V}$, the smallest value of the Rayleigh quotient over that subspace is a *lower* bound on λ_k and an *upper* bound on λ_{n-k+1}. Taking this one step further, we have the *Cauchy interlace theorem*, which relates the eigenvalues of a block Rayleigh quotient to the eigenvalues of the corresponding matrix.

Theorem 2. *Suppose A is real symmetric (or Hermitian), and let V be a matrix with m orthonormal columns. Then the eigenvalues of 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.

3 Sensitivity of invariant subspaces

The eigenvalues of a symmetric matrix are perfectly conditioned. What of the eigenvectors (or, more generally, the invariant subspaces)? Here the picture is more complex, and involves *spectral gaps*. Suppose u is an eigenvector of A associated with eigenvalue μ, and the nearest other eigenvalue is at least γ apart. Then there is a perturbation E with $\|E\|_2 = \gamma/2$ for which the eigenvalue at μ and the nearest eigenvalue coalesce.

A more refined picture is given by Davis and Kahan and covered in many textbooks since (I recommend those of Parlett and of Stewart). Let $AU = U\Lambda$ and $\hat{A}\hat{U} = \hat{U}\hat{\Lambda}$, and define $R = \|\hat{A}U - U\Lambda\|$. Then

$$\|\sin\Theta(U, \hat(U))\|_F \leq \frac{\|R\|_F}{\delta}$$

where δ is the gap between the eigenvalues in Λ and the rest of the spectrum. If we enforce a gap between an interval containing the eigenvalues in Λ and the rest of the spectrum, we can change all the Frobenius norms into 2-norms (or any other unitarily invariant norm). The matrix $\sin\Theta(U, \hat{U})$ is the matrix of sines of the *canonical angles* between U and $\hat{U}$; if both bases are normalized, the cosines of these canonical angles are the singular values of $U^*\hat{U}$.

The punchline for this is that an eigenvector or invariant subspace for eigenvalues separated by a large spectral gap from everything else in the specturm is nicely stable. But if the spectral gap is small, the vectors may spin like crazy under perturbations.

4 Sylvester's inertia theorem

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

$$M = V^* A V$$

where V is nonsingular (but not necessarily unitary).

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

$$\phi(x) = x^* A x$$

defines a concave (convex) bowl; and $\phi(Vx) = x^*(V^*AV)x$ 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 $M = V^*AV$ 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 $z^*(W^*AW)z > 0$ 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.