CS 6210: Matrix Computations
Theory for SEP and SVD

David Bindel

2025-11-03

Symmetric eigenvalue basics

The *symmetric (Hermitian) eigenvalue problem* is to find nontrivial solutions to

$$Ax = x\lambda$$

where $A = A^*$ is symmetric (Hermitian). The symmetric eigenvalue problem satisfies several properties that we do not have in the general case:

- All eigenvalues are real.

- There are no non-trivial Jordan blocks.

- Eigenvectors associated with distinct eigenvalues are orthogonal.

It is worthwhile to make some arguments for these facts, drawing on ideas we have developed already:

- For any v, $v^*Av = v^*A^*v = v^*\bar{A}v$, so v^*Av must be real; and we can write any eigenvalue as v^*Av where v is the corresponding eigenvector (normalized to unit length).

- If $(A - \lambda I)^2 v = 0$ for $\lambda \in \mathbb{R}$ and $v \neq 0$, then

$$0 = v^*(A - \lambda I)^2 v = \|(A - \lambda I)v\|^2 = 0;$$

  and so $(A - \lambda I)v = 0$ as well. But if λ is associated with a Jordan block, there must be $v \neq 0$ such that $(A - \lambda I)^2 v = 0$ and $(A - \lambda I)v \neq 0$.

- If $\lambda \neq \mu$ are eigenvalues associated with eigenvectors u and v, then

$$\lambda u^*v = u^*Av = \mu u^*v.$$

  But if $\lambda \neq \mu$, then $(\lambda - \mu)u^*v = 0$ implies that $u^*v = 0$.

We write the complete eigendecomposition of A as

$$A = U \Lambda U^*$$

where U is orthogonal or unitary and Λ is a real diagonal matrix. This is simultaneously a Schur form and a Jordan form.

More generally, if $\langle \cdot, \cdot \rangle$ is an inner product on a vector space, the *adjoint* of an operator A on that vector space is the operator A^* s.t. for any v, w

$$\langle Av, w \rangle = \langle v, A^* w \rangle.$$

If $A = A^*$, then A is said to be *self-adjoint*. If a matrix A is self-adjoint with respect to the M-inner product $\langle v, w \rangle_M = w^* M v$ where M is Hermitian positive definite, then $H = MA$ is also Hermitian. In this case, we can rewrite the eigenvalue problem

$$Ax = x\lambda$$

as

$$Hx = MAx = Mx\lambda.$$

This gives a *generalized* symmetric eigenvalue problem[1]. A standard example involves the analysis of reversible Markov chains, for which the transition matrix is self-adjoint with respect to the inner product defined by the invariant measure.

For the generalized problem involving the matrix pencil (H, M), all eigenvalues are again real and there is a complete basis of eigenvectors; but these eigenvectors are now M-orthogonal. That is, there exists U such that

$$U^* H U = \Lambda, \quad U^* M U = I.$$

Generalized eigenvalue problems arise frequently in problems from mechanics. Note that if $M = R^T R$ is a Cholesky factorization, then the generalized eigenvalue problem for (H, M) is related to a standard symmetric eigenvalue problem

$$\hat{H} = R^{-T} H R^{-1};$$

if $\hat{H} x = x\lambda$, then $Hy = My\lambda$ where $Ry = x$. We may also note that $R^{-1} \hat{H} R = M^{-1} H$; that is $\hat{H}$ is related to $A = M^{-1} H$ by a similarity transform. Particularly for the case when M is large and sparse, though, it may be preferable to work with the generalized problem directly rather than converting to a standard eigenvalue problem, whether or not the latter is symmetric.

The singular value decomposition may be associated with several different symmetric eigenvalue problems. Suppose $A \in \mathbb{R}^{n \times n}$ has the SVD $A = U \Sigma V^T$; then

$$A^T A = V \Sigma^2 V^T$$
$$A A^T = U \Sigma^2 U^T$$
$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} U & U \\ V & -V \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \begin{bmatrix} U & U \\ V & -V \end{bmatrix}^T.$$

---

[1] The case where M is allowed to be indefinite is not much nicer than the general nonsymmetric case.

The picture is marginally more complicated when A is rectangular — but only marginally.

Variational approaches

The Rayleigh quotient plays a central role in the theory of the symmetric eigenvalue problem. Recall that the Rayleigh quotient is

$$\rho_A(v) = \frac{v^* A v}{v^* v}.$$

Substituting in $A = U \Lambda U^*$ and (without loss of generality) assuming $w = U^* v$ is unit length, we have

$$\rho_A(v) = \sum_{i=1}^{N} \lambda_i |w_i|^2, \quad \text{with} \sum_{i=1}^{N} |w_i|^2 = 1.$$

That is, the Rayleigh quotient is a weighted average of the eigenvalues. Maximizing or minimizing the Rayleigh quotient therefore yields the largest and the smallest eigenvalues, respectively; more generally, for a fixed A,

$$\delta \rho_A(v) = \frac{2}{\|v\|^2} \, \delta_v^* \left(A v - v \rho_A(v) \right),$$

and so at a stationary v (where all derivatives are zero), we satisfy the eigenvalue equation

$$A v = v \rho(A).$$

The eigenvalues are the stationary values of ρ_A; the eigenvectors are stationary vectors.

The Rayleigh quotient is homogeneous of degree zero; that is, it is invariant under scaling of the argument, so $\rho_A(v) = \rho_A(\tau v)$ for any $\tau \neq 0$. Hence, rather than consider the problem of finding stationary points of ρ_A generally, we might restrict our attention to the unit sphere. That is, consider the Lagrangian function

$$L(v, \lambda) = v^* A v - \lambda(v^* v - 1);$$

taking variations gives

$$\delta L = 2 \delta v^* (A v - \lambda v) - \delta \lambda (v^* v - 1)$$

which is zero only if $A v = \lambda v$ and v is normalized to unit length. In this formulation, the eigenvalue is identical to the Lagrange multiplier that enforces the constraint.

The notion of a Rayleigh quotient generalizes to pencils. If M is Hermitian and positive definite, then

$$\rho_{A,M}(v) = \frac{v^* A v}{v^* M v}$$

is a weighted average of generalized eigenvalues, and the stationary vectors satisfy the generalized eigenvalue problem

$$Av = Mv\rho_{A,M}(v).$$

We can also restrict to the ellipsoid $\|v\|_M^2 = 1$, i.e. consider the stationary points of the Lagrangian

$$L(v, \lambda) = v^*Av - \lambda(v^*Mv - 1),$$

which again yields a generalized eigenvalue problem.

The analogous construction for the SVD is

$$\phi(u, v) = \frac{u^*Av}{\|u\|\|v\|}$$

or, thinking in terms of a constrained optimization problem,

$$L(u, v, \lambda, \mu) = u^*Av - \lambda(u^*u - 1) - \mu(v^*v - 1).$$

Taking variations gives

$$\delta L = \delta u^*(Av - 2\lambda u) + \delta v^*(A^*u - 2\mu v) - \delta\lambda(u^*u - 1) - \delta\mu(v^*v - 1),$$

and so $Av \propto u$ and $A^*u \propto v$. Combining these observations gives $A^*Av \propto v$, $AA^*u \propto u$, which we recognize as one of the standard eigenvalue problem formulations for the SVD, with the squared singular values as the constant of proportionality.

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 0.1 (Courant-Fischer). *If $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$, 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

4

$k - 1$ (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 0.2 (Cauchy interlace theorem). *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

$$\beta_j = \min_{\dim \mathcal{V} = m-k+1} \left(\max_{0 \neq v \in \mathcal{V}} \rho_L(v) \right) = \min_{\dim \mathcal{V} = m-k+1} \left(\max_{0 \neq v \in W\mathcal{V}} \rho_A(v) \right)$$

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

□

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.

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

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

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 spectrum is nicely stable. But if the spectral gap is small, the vectors may spin like crazy under perturbations.

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 $\phi(Vx) = x^*(V^* A V)x$ 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 $M = V^* A V$ 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 $z^*(W^* A W)z > 0$ 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.