Week 11: Wednesday, Oct 31

Rayleigh quotients revisited

Recall the Rayleigh quotient iteration:

\[ \lambda_k = \rho_A(v_k) = \frac{v_k^* A v_k}{v_k^* v_k}, \]
\[ v_{k+1} r_{k+1} = (A - \lambda_k)^{-1} v_k. \]

We claimed before that as \( v_k \) becomes an increasingly good eigenvector estimate, \( \lambda_k \) becomes an increasingly good eigenvalue estimate (under some assumptions), and the combination of the two effects gives us a locally quadratically convergent algorithm.

A good way to understand Rayleigh quotient iteration is as a sort of Newton iteration for the eigenvalue equations. Write

\[ F(v, \lambda) = A v - v \lambda, \]
and differentiate to find

\[ \delta F = (A - \lambda) \delta v - (\delta \lambda) v \]

Newton iteration gives

\[ 0 = F(v_k, \lambda_k) + \delta F(v_k, \lambda_k) \]
\[ = (A - \lambda_k)(v_k + \delta v_k) - (\delta \lambda_k) v_k \]
\[ = (A - \lambda_k)v_{k+1} - (\delta \lambda_k)v_k, \]

which means that \( v_{k+1} = (\delta \lambda_k)(A - \lambda_k)^{-1} v_k \), where \( \delta \lambda_k \) is some normalizing factor. This gives a nice iteration for the vector; what about for the value?

As we introduced it, the Rayleigh quotient might look arbitrary. Here’s a way to see how naturally it occurs. Note that if \((v, \lambda)\) is an eigenpair, then

\[ \| A v - v \lambda \|_2^2 = 0. \]

Now, suppose \( \hat{v} \) is an approximate eigenvector, and I want to find the corresponding best approximate eigenvalue \( \hat{\lambda} \), in the sense that

\[ \hat{\lambda} = \arg \min_\mu \| A \hat{v} - \hat{v} \mu \|_2^2. \]
This is now a linear least squares problem in the variable $\mu$, and the normal equations give us

$$
\hat{\lambda} = \hat{v}^* A \hat{v} / \hat{v}^* \hat{v};
$$

that is, the Rayleigh quotient is the choice of $\hat{\lambda}$ that minimizes the residual norm for the eigenvalue problem.

In the same way, we can derive the block Rayleigh quotient associated with a matrix $\hat{V}$ such that $\hat{V}^* \hat{V} = I$ to be the corresponding $\hat{L}$ that minimizes $\|A \hat{V} - \hat{V} \hat{L}\|_F$. The corresponding minimizer is $\hat{L} = \hat{V}^* A \hat{V}$. Just as the Rayleigh quotient provides an approximate eigenvalue, the block Rayleigh quotient provides an approximate “block” eigenvalue; and if the residual is small, eigenvalues of $\hat{L}$ are eigenvalues are close to eigenvalues of $A$ corresponding to the invariant subspace that $\hat{V}$ approximates.

Note that we can also define block Rayleigh-quotient iteration:

$$
p_k(z) = \det(V_k^* A V_k - zI)$$

$$
V_{k+1} R_{k+1} = p_k(A)^{-1} V_k.
$$

**Rayleigh quotients, minimax, etc**

Suppose $v$ is a unit-length eigenvector of $A$ with corresponding eigenvector $\lambda$ (i.e. $Av = v\lambda$). The corresponding Rayleigh quotient is the eigenvalue. What if we consider $\delta v$ very close to $A$? Let us suppose that $v$ has unit length, and differentiate $\rho_A(v) = (v^* Av) / (v^* v)$ in a direction $\delta v$. Using the quotient rule, we have

$$
\delta \rho_A(v) = \frac{(\delta v^* Av + v^* \delta Av)(v^* v) - (v^* Av) (\delta v^* v + v^* \delta v)}{(v^* v)^2}
$$

$$
= (\delta v^* Av + v^* \delta Av) - \lambda (\delta v^* v + v^* \delta v)
$$

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

Now, note that

$$
\delta \rho_A(v) = \delta v^* (A - \lambda I) v + v^* (A - \lambda I) \delta v = v^* (A - \lambda I) \delta v.
$$

If $v^*$ is a row eigenvector of $A$ corresponding to $\lambda$, then $v$ is a stationary point of $\rho_A$. The vector $v^*$ is a row eigenvector whenever the matrix $A$ is normal.
that is, whenever \( AV = V\Lambda \) for some unitary matrix \( V \). Such stationarity implies that
\[
\rho_A(v + \delta v) = \lambda + O(\|\delta v\|^2).
\]
This is a strong statement; it implies that when we have a first-order accurate estimate of an eigenvector, we have a second-order accurate estimate of the corresponding eigenvalue.

Any real symmetric (or complex Hermitian) matrices is normal; and for a real symmetric matrix, we have that all the eigenvalues are real, and that they are critical points for \( \rho_A \). This variational characterization of the eigenvalues of \( A \) means, in particular, that \( \lambda_{\max} = \max_{v \neq 0} \rho_A(v) \) and \( \lambda_{\min} = \min_{v \neq 0} \rho_A(v) \). We can go one step further with 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 \( \lambda_k \) when we choose \( v = u_k \). \( \square \)
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 $\lambda_k$ and an upper bound on $\lambda_{n-k+1}$. 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 $V^*AV$ interlace the eigenvalues of $A$; that is, if $A$ has eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and $V^*AV$ has eigenvalues $\beta_j$, then

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