The Inverse Power Method

Assume that $A \in \mathbb{C}^{n \times n}$ has a n linearly independent eigenvectors $v_1, v_2, \dots, v_n$, and associated eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, with $|\lambda_1| > |\lambda_2| \geq |\lambda_i|,\ i = 3, 4, \dots, n$. Then (λ_1, v_1) is a dominant eigenpair of A, and for almost all $x \in \mathbb{C}^n$, $A^k x \to v_1$, $k \to \infty$.

How fast does such an iteration converge? Write $x = \sum_{i=1}^{n} c_i v_i$. Then

$$\frac{A^k x}{\lambda_1^k} = c_1 v_1 + c_2 (\frac{\lambda_2}{\lambda_1})^k v_2 + \sum_{i=3}^{n} c_i (\frac{\lambda_i}{\lambda_1})^k v_i \tag{1}$$

and it is clear (yes?) that the error gets multiplied by about $|\lambda_2/\lambda_1|$ at each step (we say that the convergence is linear with asymptotic error constant $|\lambda_2/\lambda_1|$). So the smaller the ratio $|\lambda_2/\lambda_1|$, the better, and if $|\lambda_1| \approx |\lambda_2|$ then we expect very slow convergence.

Let $B \in \mathbb{C}^{n \times n}$ have eigenvalues μ_i labeled so that $|\mu_i| \geq |\mu_{i+1}|$. The power method *applied to* B converges to the dominant eigenvector of B (if one exists) with a speed that depends on $|\mu_2/\mu_1|$.

Now if $B = A^{-1}$, then the dominant eigenpair of B is $(1/\lambda_n, v_n)$, and the power method *applied to* B converges to v_n with a speed that depends on $|\lambda_n/\lambda_{n-1}|$. This iteration is fast if $|\lambda_n|$ is small compared to the other eigenvalues. The power method applied to A^{-1} is called the *inverse power method.*

We now have methods to compute the |largest| and |smallest| eigenvectors of a matrix $A \in \mathbb{C}^{n \times n}$. How fast these methods go depends upon how relatively large or small the |largest| and |smallest| eigenvalues are. One more observation: if we take $B = (A - sI)^{-1}$, then $\mu_1 = \lambda_r - s$, where λ_r is the eigenvalue of A that is closest to s. The dominant eigenpair of B is $(1/(\lambda_r - s), v_r)$, and the power method *applied to* B converges to v_r with a speed, ρ, that depends on how close s is to λ_r relative to the next closest eigenvalue of A:

$$\rho = \frac{|\lambda_r - s|}{\min\limits_{i \neq r} |\lambda_i - s|}. \tag{2}$$

The power method applied to $(A - sI)^{-1}$ is called the *inverse power method with shift*; it is at the heart of many state-of-the-art methods.

This "shifted inverse power method" is better called the "inverse power kernel", for there are many decisions yet to be made about its implementation. For example, suppose we have an approximation s to λ_r. Then we might use Gaussian Elimination with partial pivoting to compute the factorization $P(A - sI) = LU$. Once this is done, each iteration of the inverse power method requires only about $2n^2$ flops (forsub and backsub). Do we recompute the LU factorization as our approximation to λ_r gets better? For the general eigenproblem, it is probably not efficient, but for a matrix with some special structure it may be efficient. For symmetric matrices, a reduction to tridiagonal form, followed by an inverse power method with shift, using a different shift (a Rayleigh quotient) at each iteration, is a standard technique. In fact, all of our best methods for computing eigenstuff are intimately related to this *inverse iteration.*