Error Bounds for the Conjugate Gradient Algorithm

Tyler Chen

This page is a work in progress.

In our derivation of the Conjugate Gradient method, we minimized the A-norm of the error over sucessive Krylov subspaces. Ideally we would like to know how quickly this method converge. That is, how many iterations are needed to reach a specified level of accuracy.

Linear algebra review

- The 2-norm of a symmetric positive definite matrix is the largest eigenvalue of the matrix
- The 2-norm is submultiplicative. That is, $\|A\|\|B\| \leq \|AB\|$
- A matrix U is called unitary if $U^*U = UU^* = I$.

Polynomial error bounds

Previously we have show that,

$$e_k \in e_0 + \operatorname{span}\{p_0, p_1, \ldots, p_{k-1}\} = e_0 + \mathcal{K}_k(A, b)$$

Observing that $r_0 = Ae_0$ we find that,

$$e_k \in e_0 + \operatorname{span}\{Ae_0, A^2e_0, \ldots, A^ke_0\}$$

Thus, we can write,

$$\|e_k\|_A = \min_{p \in \mathcal{P}_k} \|p(A)e_0\|_A, \qquad \mathcal{P}_k = \{p : p(0) = 1, \deg p \leq k\}$$

Since $A^{1/2}p(A) = p(A)A^{1/2}$ we can write,

$$\|p(A)e_0\|_A = \|A^{1/2}p(A)e_0\| = \|p(A)A^{1/2}e_0\|$$

Now, using the submultiplicative property of the 2-norm,

$$\|p(A)A^{1/2}e_0\| \leq \|p(A)\|\|A^{1/2}e_0\| = \|p(A)\|\|e_0\|_A$$

Since A is positive definite, it is diagonalizable as $U\Lambda U^*$ where U is unitary and Λ is the diagonal matrix of eigenvalues of A. Thus,

$$A^k = (U\Lambda U^*)^k = U\Lambda^k U^*$$

We can then write $p(A) = Up(\Lambda)U^*$ where $p(\Lambda)$ has diagonal entries $p(\lambda_i)$. Therefore, using the *unitary invariance* property of the 2-norm,

$$\|p(A)\| = \|Up(\Lambda)U^*\| = \|p(\Lambda)\|$$

Now, since the 2-norm of a symmetric matrix is the magnitude of the largest eigenvalue,

$$\|p(\Lambda)\| = \max_i |p(\lambda_i)|$$

Finally, putting everything together we have,

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq \min_{p \in \mathcal{P}_k} \max_i |p(\lambda_i)|$$

Since the inequality we obtained from the submultiplicativity of the 2-norm is tight, this bound is also tight in the sense that for a fixed k there exists an initial error e_0 so that equality holds.

Computing the optimal p is not trivial, but an algorithm called the Remez algorithm can be used to compute it. I discuss this in more detail below.

Let $L \subset \mathbb{R}$ be some closed set. The *minimax polynomial of degree k on L* is the polynomial satisfying,

$$\min_{p \in \mathcal{P}_k} \max_{x \in L} |p(x)|, \qquad \mathcal{P}_k = \{p : p(0) = 1, \deg p \leq k\}$$

Chebyshev bounds

The minimax polynomial on the eigenvalues of A is a bit tricky to work with. Although we can find it using the Remez algorithm, this is somewhat tedious, and requires knowledge of the whole spectrum of A. We would like to come up with a bound which depends on less informationabout A. One way to obtain such a bound is to expand the set on which we are looking for the minimax polynomial.

To this end, let $\mathcal{I} = [\lambda_{\min}, \lambda_{\max}]$. Then, since $\lambda_i \in \mathcal{I}$,

$$\min_{p \in \mathcal{P}_k} \max_i |p(\lambda_i)| \leq \min_{p \in \mathcal{P}_k} \max_{x \in \mathcal{I}} |p(x)|$$

The right hand side requires that we know the largest and smallest eigenvalues of A, but doesn't require any of the ones between. This means it can be useful in practice, since we can easily compute the top and bottom eignevalues with the power method.

The polynomials satisfying the right hand side are called the *Chebyshev Polynomials* and can be easily written down with a simple recurrence relation. If $\mathcal{I} = [-1, 1]$ then the relation is,

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x), \quad T_0 = 1, \quad T_1 = x$$

For $\mathcal{I} \neq [-1, 1]$, the above polynomials are simply stretched and shifted to the interval in question.

Let $\kappa = \lambda_{\max}/\lambda_{\min}$ (this is called the condition number). Then, from properties of these polynomials,

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$$