Notes for 2016-11-14

1 Chebyshev polynomials

Suppose now that A is symmetric positive definite, and we seek to minimize $\|q(A)b\| \leq \|q(\Lambda)\|\|b\|$. Controlling $q(z)$ on all the eigenvalues is a pain, but it turns out to be simple to instead bound $q(z)$ over some interval $[\alpha_1, \alpha_n]$ The polynomial we want is the *scaled and shifted Chebyshev polynomial*

$$q_m(z) = \frac{T_m\left((z - \bar{\alpha})/\rho\right)}{T_m\left(-\bar{\alpha}/\rho\right)}$$

where $\bar{\alpha} = (\alpha_n + \alpha_1)/2$ and $\rho = (\alpha_n - \alpha_1)/2$.

The Chebyshev polynomials T_m are defined by the recurrence

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_{m+1}(x) &= 2xT_m(x) - T_{m-1}(x), \quad m \geq 1. \end{aligned}$$

The Chebyshev polynomials have a number of remarkable properties, but perhaps the most relevant in this setting is that

$$T_m(x) = \begin{cases} \cos(m\cos^{-1}(x)), & |x| \leq 1, \\ \cosh(m\cosh^{-1}(x)), & |x| \geq 1 \end{cases}.$$

Thus, $T_m(x)$ oscillates between ± 1 on the interval $[-1, 1]$, and then grows very quickly outside that interval. In particular,

$$T_m(1+\epsilon) \geq \frac{1}{2}(1 + m\sqrt{2\epsilon}).$$

Thus, we have that on $[\alpha_, \alpha_n]$, $|q_m| \leq \frac{2}{1+m\sqrt{2\epsilon}}$ where

$$\epsilon = \bar{\alpha}/\rho - 1 = \frac{2\alpha_1}{\alpha_n - \alpha_1} = 2\left(\kappa(A) - 1\right)^{-1},$$

and hence

$$\begin{aligned} |q_m(z)| &\leq \frac{2}{1 + 2m/\sqrt{\kappa(A) - 1}} \\ &= 2\left(1 - \frac{2m}{\sqrt{\kappa(A) - 1}}\right) + O\left(\frac{m^2}{\kappa(A-1)}\right). \end{aligned}$$

Hence, we expect to reduce the optimal residual in this case by at least about $2/\sqrt{\kappa(A)-1}$ at each step.

2 Chebyshev: Uses and Limitations

We previously sketched out an approach for analyzing the convergence of methods based on Krylov subspaces:

1. Characterize the Krylov subspace of interest in terms of polynomials, i.e. $\mathcal{K}_k(A,b) = \{p(A)b : p \in \mathcal{P}_{k-1}\}$.

2. For $\hat{x} = p(A)b$, write an associated error (or residual) in terms of a related polynomial in A.

3. Phrase the problem of minimizing the error, residual, etc. in terms of minimizing a polynomial $q(z)$ on the spectrum of A (call this $\Lambda(A)$). The polynomial q must generally satisfy some side constraints that prevent the zero polynomial from being a valid solution.

4. Let $\Lambda(A) \subset \Omega$, and write

$$\max_{\lambda \in \Lambda(A)} |q(\lambda)| \leq \max_{z \in Omega} |q(z)|.$$

   The set Ω should be simpler to work with than the set of eigenvalues. The simplest case is when A is symmetric positive definite and $\Omega = [\lambda_1, \lambda_n]$.

5. The optimization problem can usually be phrased in terms of special polynomial families. The simplest case, when Ω is just an interval, usually leads to an analysis via Chebyshev polynomials.

The analysis sketched above is the basis for the convergence analysis of the Chebyshev semi-iteration, the conjugate gradient method, and (with various twists) several other Krylov subspace methods.

The advantage of this type of analysis is that it leads to convergence bounds in terms of some relatively simple property of the matrix, such as the condition number. The disadvantage is that the approximation of the spectral set $\Lambda(A)$ by a bounding region Ω can lead to rather pessimistic bounds. In practice, the extent to which we are able to find good solutions in a Krylov subspace often depends on the "clumpiness" of the eigenvalues. Unfortunately, this "clumpiness" is rather difficult to reason about a priori! Thus, the right way to evaluate the convergence of Krylov methods in practice is usually to try them out, plot the convergence curves, and see what happens.