Week 13: Monday, Nov 12

Beyond stationary methods

So far, we have discussed stationary iterative methods that produce a sequence of approximations

$$Mx^{(k+1)} = Kx^{(k)} + b,$$

where $A = M - K$ is some splitting. The error in these methods behaves like

$$e^{(k)} = R^k e^{(0)},$$

and asymptotically, the dominant component of the error points in the direction of the dominant eigenvector of R. Given that the errors are asymptotically become correlated with each other in a very systematic way, it seems reasonable that we could "cancel off" some of the error by taking a linear combination of the guesses $x^{(k)}$. That is, we could try something like

$$\tilde{x}^{(m)} = \sum_{k=0}^{m} \gamma_{mk} x^{(k)}$$

where the coefficients γ_{mk} sum to one. Note that

$$\tilde{x}^{(m)} - x = \sum_{k=0}^{m} \gamma_{mk}(x^{(k)} - x) = \sum_{k=0}^{m} \gamma_{mk} R^k e^{(0)} = p_m(R)e^{(0)},$$

where

$$p_m(z) = \sum_{k=0}^{m} \gamma_{mk} z^k$$

is a polynomial normalized so that $p_m(1) = 1$.

Now, if we knew all the eigenvalues R, we could take p_n to be the characteristic polynomial of R in order to get $p_n(R)e^{(0)}$ to be zero. There are just two problems with this approach. The first is that we generally don't know all the eigenvalues of the iteration matrix. The second is that even if we *did* know all the eigenvalues, applying the inverse via the characteristic polynomial would take $O(N)$ steps, and we would like to get a good answer in far fewer steps than that.

Let us now consider the special case of a stationary iteration in which R is symmetric. We may not be able to say where all the eigenvalues of R are, but we can frequently give some bound on the spectral radius, i.e. $\rho(R) \leq \alpha < 1$. In this case, we can bound

$$\|p(R)\| \leq \max_{z \in [-\alpha,\alpha]} |p(z)|,$$

so a reasonable way to choose polynomials is to make $p_m(z)$ the polynomial of degree m that minimizes the maximum of $|p_m(z)|$ on $[-\alpha, \alpha]$ subject to $p_m(z) = 1$. The solution to this problem is the *scaled Chebyshev polynomial*

$$p_m(z) = \frac{T_m(z/\alpha)}{T_m(1/\alpha)},$$

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 for $z \in [-\alpha, \alpha]$,

$$p_m(z) \leq \frac{2}{1 + m\sqrt{2/(1-\alpha)}} = 2(1 - m\sqrt{2(1-\alpha)}) + O(m^2(1-\alpha)).$$

Thus, where the number of steps for the basic stationary iteration to converge scales like $(1-\rho(R))^{-1}$, the number of steps for the Chebyshev semi-iteration to converge scales like $(1-\alpha)^{-1/2}$. On the model problem, this means we can accelerate Jacobi or symmetric Gauss Seidel from $O(N^2)$ to $O(N^{3/2})$ time with this approach, and we can scale SOR with a well-chosen relaxation parameter from $O(N^{3/2})$ to $O(N^{5/4})$.

Krylov subspaces

There were two ingredients to the Chebyshev semi-iteration:

1. Generate a space in which we expect to find good approximate solutions.
2. Pull an "optimal" solution out of that space.

If we look at the iterations generated from a stationary iterative method with the initial guess $x^{(0)} = 0$, we have

$$\begin{aligned} x^{(1)} &= c, \\ x^{(2)} &= Rc + c, \\ x^{(3)} &= R^2 c + Rc + c, \\ x^{(4)} &= R^3 c + R^2 c + Rc + c, \end{aligned}$$

and so on. In general, the things we can form from $\{x^{(1)}, \ldots, x^{(k)}\}$ live in the kth *Krylov subspace* generated by R and c:

$$\mathcal{K}_k(R, c) = \operatorname{span}\{c, Rc, \ldots, R^{k-1}c\} = \{p(R)c : p \in \mathcal{P}_{k-1}\},$$

where $\mathcal{P}_{k-1}$ is the space of polynomials of degree at most $k-1$.

As it happens, Krylov subspaces make a reasonable choice of approximation spaces even when they are generated by something other than a contraction mapping. In general, *preconditioned Krylov subspace* methods draw approximations from $\mathcal{K}_k(M^{-1}A, M^{-1}b)$.