Notes for 2016-11-11

1 Quasi-optimality

We quantify the stability of a subspace approximation method via a *quasi-optimality bound*:
$$\|x^* - \hat{x}\| \leq C \min_{v \in \mathcal{V}} \|x^* - v\|.$$
That is, the approximation $\hat{x}$ is quasi-optimal if it has error within some factor C of the best error possible within the space.

To derive quasi-optimality results, it is useful to think of all of our methods as defining a *solution projector* that maps x^* to the approximate solution to $A\hat{x} = Ax^* = b$. From the (Petrov-)Galerkin perspective, if $W \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{n \times k}$ are bases for the trial space $\mathcal{W}$ and $\mathcal{V}$, respectively, then we have

$$\begin{aligned} W^T AV\hat{y} = W^T b, \quad \hat{x} = V\hat{y} \\ \hat{x} &= V(W^T AV)^{-1} W^T b \\ &= V(W^T AV)^{-1} W^T Ax^*. \\ &= \Pi x^*. \end{aligned}$$

The *error projector* $I - \Pi$ maps x^* to the error $\hat{x} - x^*$ in approximately solving $A\hat{x} \approx Ax^* = b$. There is no error iff x^* is actually in $\mathcal{V}$; that is, $\mathcal{V}$ is the null space of $I - \Pi$. Hence, if $\tilde{x}$ is any vector in $\mathcal{V}$, then

$$\hat{e} = (I - \Pi)x = (I - \Pi)(x - \tilde{x}) = (I - \Pi)\tilde{e}.$$

Therefore we have

$$\|x - \hat{x}\| \leq \|I - \Pi\| \min_{\tilde{x} \in \mathcal{V}} \|x - \tilde{x}\|,$$

and a bound on $\|I - \Pi\|$ gives a quasi-optimality result.

For any operator norm, we have

$$|I - \Pi\| \leq 1 + \|\Pi\| \leq 1 + \|V\| \|(W^T AV)^{-1}\| \|W^T A\|;$$

and in any Euclidean norm, if V and W are chosen to have orthonormal columns, then
$$\|I - \Pi\| \leq 1 + \|(W^T AV)^{-1}\| \|A\|.$$

If A is symmetric and positive definite and $V = W$, then the interlace theorem gives $\|(V^T AV)^{-1}\| \leq \|A^{-1}\|$, and the quasi-optimality constant is bounded by $1 + \kappa(A)$. In more general settings, though, we may have no guarantee that the projected matrix $W^T AV$ is far from singular, even if A itself is nonsingular. To guarantee boundedness of $(W^T AV)^{-1}$ *a priori* requires a compatibility condition relating $\mathcal{W}$, $\mathcal{V}$, and A; such a condition is sometimes called the *LBB* condition (for Ladyzhenskaya-Babuška-Brezzi) or the *inf-sup* condition, so named because (as we have discussed previously)

$$\sigma_{\min}(W^T AV) = \inf_{w \in \mathcal{W}} \sup_{v \in \mathcal{V}} \frac{w^T Av}{\|w\|\|v\|}.$$

The LBB condition plays an important role when Galerkin methods are used to solve large-scale PDE problems, since there it is easy to choose the spaces $\mathcal{V}$ and $\mathcal{W}$ in a way that leads to very bad conditioning. But for iterative solvers of the type we discuss in this course (Krylov subspace solvers), such pathologies are a more rare occurrence. In this setting, we may prefer to monitor $\|(W^T AV)^{-1}\|$ directly as we go along, and to simply increase the dimension of the space if we ever run into trouble.

2 Model reduction

Our focus in this section is methods for solving a single linear system at a time. Often, we want to solve many closely-related linear systems with different matrices. As a simple example, we might want to evaluate

$$(A - \sigma I)x(\sigma) = b$$

for several different values of σ within some range; more generally, we might want to solve linear systems $A(s)x(s) = b(s)$ where A and b depend smoothly on some low-dimensional parameter vector s that varies over a bounded set. In such settings, one often finds (and can sometimes prove via interpolation theory) that $x(s)$ lies close to a space $\mathcal{V}$ that can be computed. For example, we might find that an adequate space $\mathcal{V}$ spanned by sample solutions $x(s_1), x(s_2), \ldots$; we could then choose a corresponding trial space $\mathcal{W}$ as the basis for a Galerkin scheme. Hence, we may estimate $x(\sigma)$ very quickly (online) after a more expensive computation to construct a basis for an appropriate approximation space (offline).

There are a wide-variety of techniques that employ this general idea. These include model reduction methods from control theory (moment-matching methods that use Krylov subspaces, truncated balanced realization methods that involve solving Sylvester equations, etc); global-basis methods for the solution of PDEs (e.g. the so-called *empirical interpolation method*); and many other methods for both linear and nonlinear problems. While it is not a focus for this course, the approach is so simple and broadly applicable that I would feel bad if you left not knowing about it.

3 Krylov subspaces

The *Krylov subspace* of dimension k generated by $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$ is

$$\mathcal{K}_k(A,b) = \operatorname{span}\{b, Ab, \ldots, A^{k-1}b\} = \{p(A)b : p \in \mathcal{P}_{k-1}\}.$$

Krylov subspaces are a natural choice for subspace-based methods for approximate linear solves, for two reasons:

- If all you are allowed to do with A is compute matrix-vector products, and the only vector at hand is b, what else would you do?
- The Krylov subspaces have excellent approximation properties.

Krylov subspaces have several properties that are worthy of comment. Because the vectors $A^j b$ are proportional to the vectors obtained in power iteration, one might reasonably (and correctly) assume that the space quickly contains good approximations to the eigenvectors associated with the largest magnitude eigenvalues. Krylov subspaces are also *shift-invariant*, i.e. for any σ

$$\mathcal{K}_k(A - \sigma I, b) = \mathcal{K}_k(A, b).$$

By choosing different shifts, we can see that the Krylov subspaces tend to quickly contain not only good approximations to the eigenvector associated with the largest magnitude eigenvalue, but to all “extremal” eigenvalues.

Most arguments about the approximation properties of Krylov subspaces derive from the characterization of the space as all vectors $p(A)b$ where $p \in \mathcal{P}_{k-1}$ and from the spectral mapping theorem, which says that if $A = V \Lambda V^{-1}$ then $p(A) = V p(\Lambda) V^{-1}$. Hence, the distance between an arbitrary vector (say d) and the Krylov subspace is

$$\min_{p \in \mathcal{P}_{k-1}} \left\| V \left[p(\Lambda) V^{-1} b - V^{-1} d \right] \right\|.$$

As a specific example, suppose that we want to choose $\hat{x}$ in a Krylov subspace in order to minimize the residual $A\hat{x} - b$. Writing $\hat{x} = p(A)b$, we have that we want to minimize

$$\|[Ap(A) - I]b\| = \|q(A)b\|$$

where $q(z)$ is a polynomial of degree at most k such that $q(1) = 1$. The best possible residual in this case is bounded by

$$\|q(A)b\| \leq \kappa(V)\|q(\Lambda)\|\|b\|,$$

and so the relative residual can be bounded in terms of the condition number of V and the minimum value that can bound q on the spectrum of A subject to the constraint that $q(0) = 1$.