Estimating the Numerical Range with a Krylov Subspace

John Urschel, Cecilia Chen

Abstract

Moment-based methods are ubiquitous in applied mathematics, and numerical linear algebra is no exception. Krylov subspace methods are an incredibly popular family of algorithms that approximate the solution to some problem involving a matrix A using the Krylov subspace

$$\mathcal{K}_m(A,b) = \operatorname{span}\{b, Ab, A^2b, \dots, A^{m-1}b\}.$$

This includes methods for approximating linear systems (conjugate gradient method, GMRES, MINRES, etc.), extremal eigenvalue problems (Arnoldi iteration, Lanczos method, etc.), and matrix functions. In many applications, m is much smaller than n, and a key benefit of this framework is that only matrix-vector products are involved, allowing for fast computation.

In this talk, we focus on the quality of estimate on the numerical range

$$W(A) = \left\{ \frac{x^* A x}{x^* x} \, \middle| \, x \in \mathbb{C}^{n \times n} \right\}$$

of a matrix A provided by the numerical range of the orthogonal projection of A onto the Krylov subspace $\mathcal{K}_m(A, b)$ for some vector b, denoted by H_m . Estimates on W(A) are important not only in the computation of extremal eigenvalues, but for error estimates for other methods. For instance, standard error bounds for the residual in the GMRES algorithm for Ax = b after m steps depend on the quantity

$$\min_{p \in \mathcal{P}_m \text{ s.t. } p(0)=1} \max_{\lambda \in \Lambda(A)} |p(\lambda)|, \tag{1}$$

where \mathcal{P}_m is the set of complex polynomials of degree at most m and $\Lambda(A)$ is the spectrum of A. If $W(H_m)$ provides a good estimate of W(A), or even $\operatorname{conv}\{\Lambda(A)\}$, then computing $W(H_m)$ and estimating the separation from zero can produce guarantees for the convergence rate.

Typically, estimates for approximating an extreme eigenvalue λ with a Krylov subspace $\mathcal{K}_{m+1}(A, b)$ depend on the quality of the initial guess b in relation to the eigenspace of λ , the eigenvector condition number, and the quantity

$$\min_{p \in \mathcal{P}_m \text{ s.t. } p(\lambda) = 1} \max_{\mu \in \Lambda(A) \setminus \lambda} |p(\mu)|.$$
(2)

The latter quantity is small when there is separation between λ and $\Lambda(A)\setminus\lambda$ in the complex plane, but can become arbitrarily close to one in many cases, producing unnecessarily pessimistic bounds. These issues can persist even when A is Hermitian. For instance, consider the tridiagonal matrix A resulting from the discretization of the Laplacian operator on an interval with Dirichlet boundary conditions (i.e., $A_{ii} = 2$, $A_{i,i+1} = A_{i+1,i} = -1$). This matrix has no small eigenvalue gaps, and so standard gap-dependent error estimates significantly over estimate the error in approximation for extreme eigenvalues.

In practice, Krylov subspace methods perform well at estimating extreme eigenvalues, even when eigenvalue gaps are small. In their 2015 paper, Musco and Musco argue that bounds on distances to any particular eigenvector are not necessarily needed in situations where the goal is only to estimate the eigenvalues. While a good estimate of an eigenvector naturally implies a good estimate of its corresponding eigenvalue, the converse is simply not true. For instance, while having two extreme eigenvalues λ and μ very close together significantly hinders eigenvector approximation, it can actually improve eigenvalue estimation, as a good approximation to any vector in the direct sum of the eigenspaces of λ and μ (assuming the matrix is well-conditioned) implies a good estimate to both eigenvalues, since they are both close together.

Kuczynski and Wozniakowski were arguably the first to fully recognize this phenomenon in a quantitative way, producing a probabilistic upper bound of the form

$$\frac{\lambda_{\max}(A) - \lambda_{\max}^{(m)}}{\lambda_{\max}(A) - \lambda_{\min}(A)} \lesssim \frac{\ln^2 n}{m^2}$$
(3)

for a real symmetric matrix $A \in \mathbb{R}^{n \times n}$, where $\lambda_{\max}^{(m)}$ is the largest eigenvalue of H_m , the orthogonal projection of A onto $\mathcal{K}_m(A, b)$, for an initial guess b sampled uniformly from the hypersphere.

In this talk, we consider the extension of this type of gap-independent result to general matrices. In particular, we consider the approximation that the numerical range of H_m provides to the numerical range of A and to the convex hull of the eigenvalues of A. This is quantified by the Hausdorff distance $d_H(W(H_m), W(A))$ between $W(H_m)$ and W(A), and the one-sided Hausdorff distance $\sup_{\mu \in \operatorname{conv}(\Lambda(A))} d(\mu, W(H_m))$ between $\operatorname{conv}(\Lambda(A))$ and $W(H_m)$. We will consider three distinct cases: normal matrices, normal matrices with their spectrum on a circle (e.g., unitary matrices), and non-normal matrices. Time permitting, we may also discuss bounds for the behavior of the quantities (1) and (2), depending on the structure of the eigenvalues of A.