Spectrum-Adapted Polynomial Approximation for Matrix Functions

Li Fan, David I Shuman

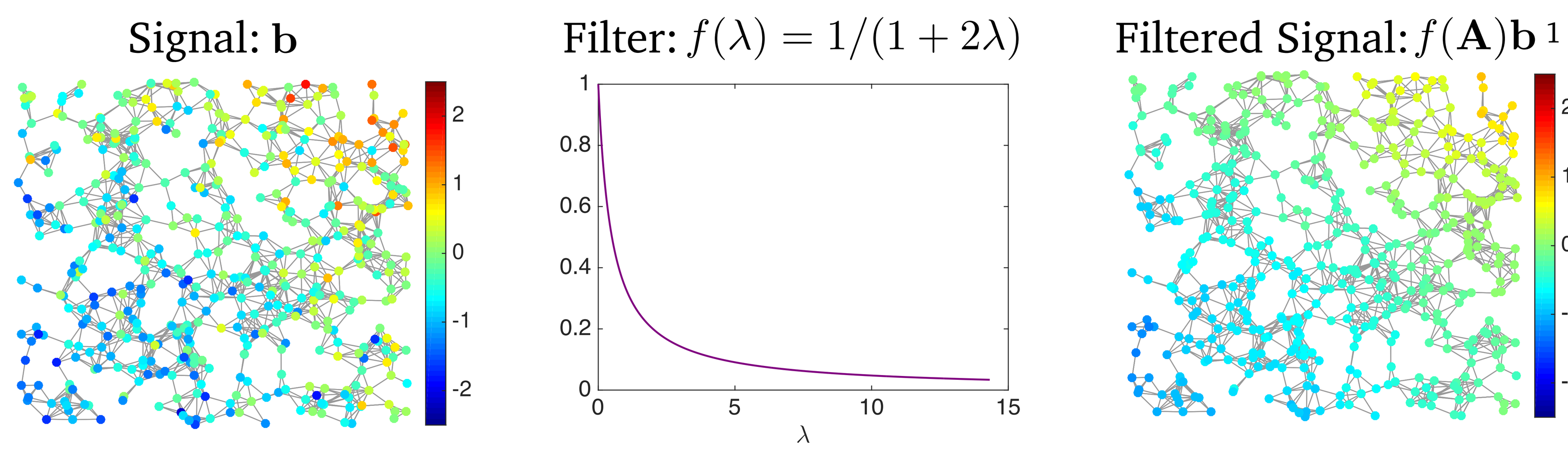
Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, MN

MATRIX FUNCTION × VECTOR: $f(\mathbf{A})\mathbf{b}$

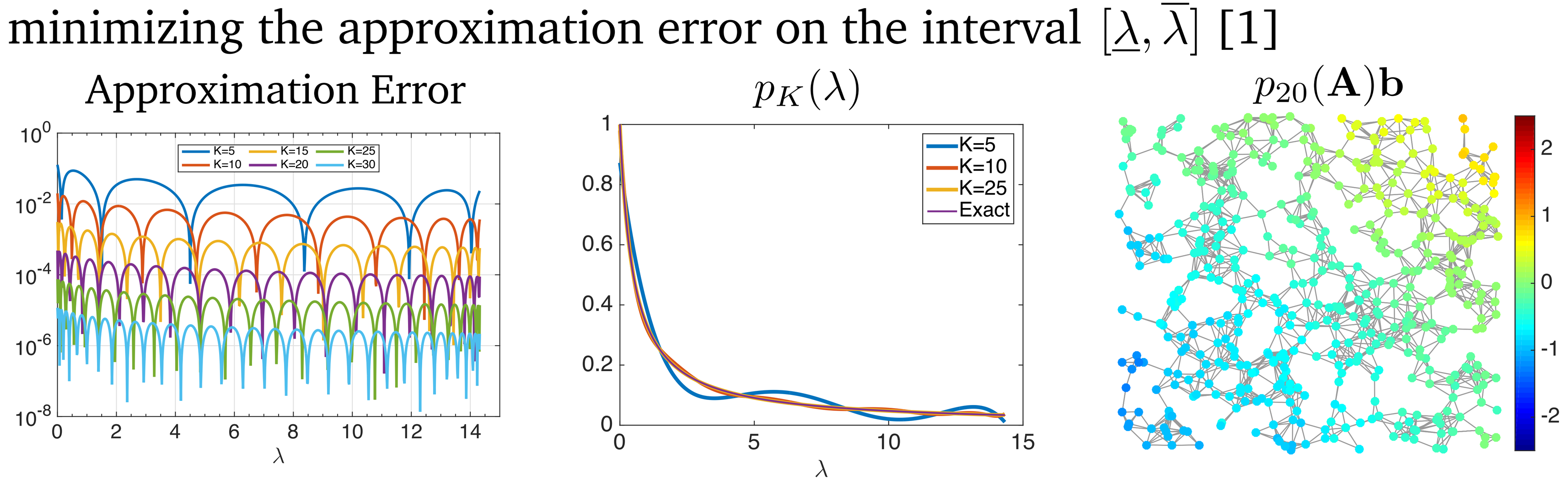
- Large, sparse, Hermitian matrix $\mathbf{A} \in \mathbb{R}^{N\times N} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$
- Orthonormal eigenvectors $\{\mathbf{v}_i\}_{i=1,\cdots,N}$ with eigenvalues $\{\lambda_i\}_{i=1,\cdots,N}$

$$f(\mathbf{A}) := \mathbf{V}f(\mathbf{\Lambda})\mathbf{V}^\top = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_N \\ | & & | \end{bmatrix} \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_N) \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1 & - \\ & \vdots & \\ - & \mathbf{v}_N & - \end{bmatrix}$$

- $f(\mathbf{A})\mathbf{b}$ is widely used in signal processing, machine learning, applied math and science, etc., but impractical to compute directly: complexity $\mathcal{O}(N^3)$

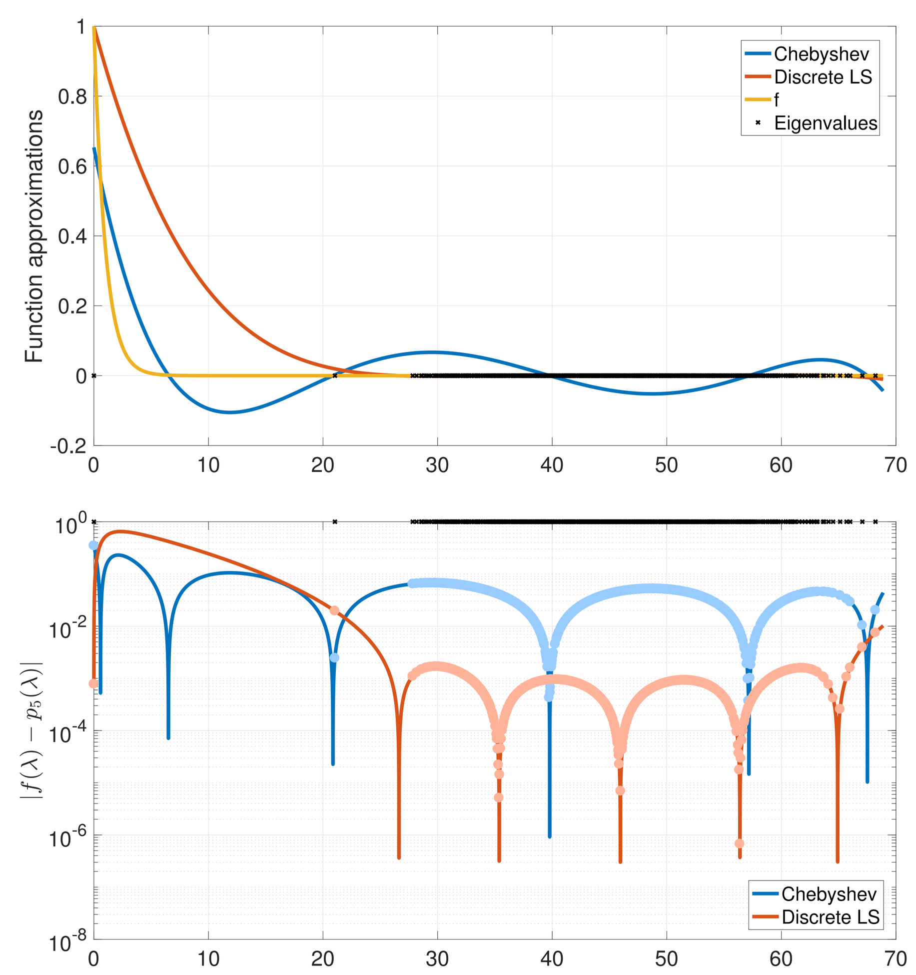


- Classical methods approximate $f(\lambda)$ with an order K polynomial $p_K(\lambda)$, by minimizing the approximation error on the interval $[\underline{\lambda}, \overline{\lambda}]$ [1]



- Polynomial methods have complexity of $\mathcal{O}(KZ)$ where Z is the number of nonzero entries in $\mathbf{A}$; faster than the direct computation
- However, the error in $f(\mathbf{A})\mathbf{b}$ depends only on the residuals at the eigenvalues of $\mathbf{A}$

$$||f(\mathbf{A}) - p_K(\mathbf{A})||_2 = \max_{l=1,2,\ldots,N} |f(\lambda_l) - p_K(\lambda_l)| \le \sup_{\lambda\in[\underline{\lambda},\overline{\lambda}]} |f(\lambda) - p_K(\lambda)|$$



- Approximate $f(\mathbf{A})\mathbf{b}$ with $p_K(\mathbf{A})\mathbf{b}$ which provides a better approximation on regions with higher density of eigenvalues

1 In this example, $\mathbf{A}$ is the graph Laplacian matrix of the displayed sensor network.

ONGOING WORK

- Test proposed methods on further applications, such as the estimation of the log-determinant of a large sparse Hermitian matrix
- Investigate convergence theory and error analysis
- Adapt the approximation to matrix function in addition to spectral density
- Explore efficient methods for computing interior eigenvalues
- Include iterative steps in the approximation of spectral density

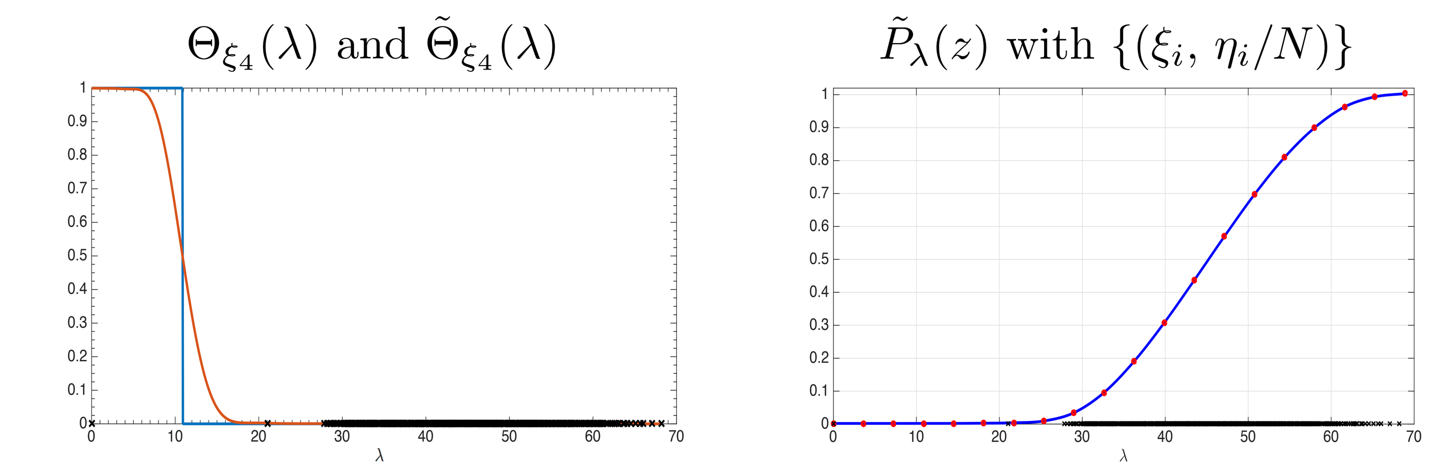
[1] Druskin and Knizhnerman, "Two polynomial methods of calculating functions of symmetric matrices," *U.S.S.R. Comput. Maths. Math. Phys.*, 1989.
[2] Lin et al., "Approximating spectral densities of large matrices," *SIAM Review*, 2016.
[3] Shuman et al., "Distributed signal processing via Chebyshev polynomial approximation," *IEEE T-SIPN*, 2018.
[4] Forsythe, "Generation and use of orthogonal polynomials for data-fitting with a digital computer," *J. SIAM*, 1957.
[5] Gautschi, Orthogonal Polynomials: Computation and Approximation, 2004.

SPECTRAL DENSITY ESTIMATION: KERNAL POLYNOMIAL METHOD [2,3]

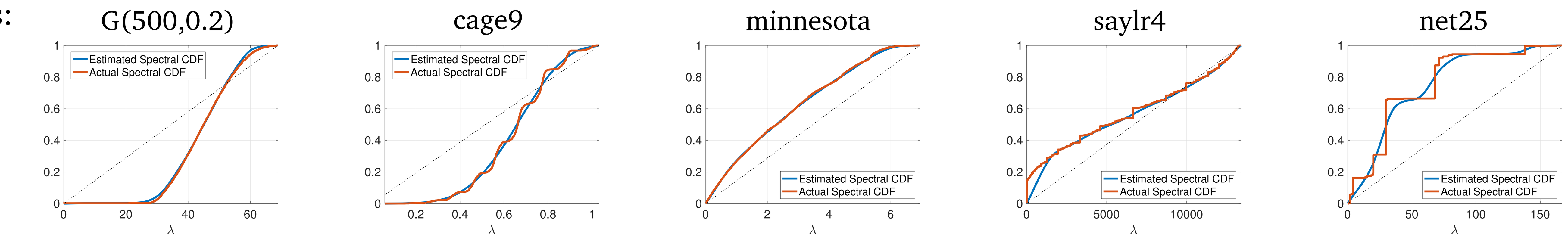
- Generate evenly spaced points $\{\xi_i\}$ on $[\underline{\lambda}, \overline{\lambda}]$
- Build low pass filter $\Theta_{\xi_i}(\lambda)$ to count eigenvalues below ξ_i

$$\eta_i = \mathrm{tr}\big(\Theta_{\xi_i}(\mathbf{A})\big) = \mathbb{E}[\mathbf{x}^\top \Theta_{\xi_i}(\mathbf{A})\mathbf{x}] \approx \frac{1}{J}\sum_{j=1}^{J} \mathbf{x}^{(j)\top} \tilde{\Theta}_{\xi_i}(\mathbf{A})\mathbf{x}^{(j)}$$

- Interpolate a monotonic piecewise cubic function $\tilde{P}_\lambda(z)$ through points $\{(\xi_i, \eta_i/N)\}$ to estimate spectral CDF
- Compute spectral PDF $\tilde{p}_\lambda(z)$ and inverse CDF $\tilde{P}_\lambda^{-1}(z)$

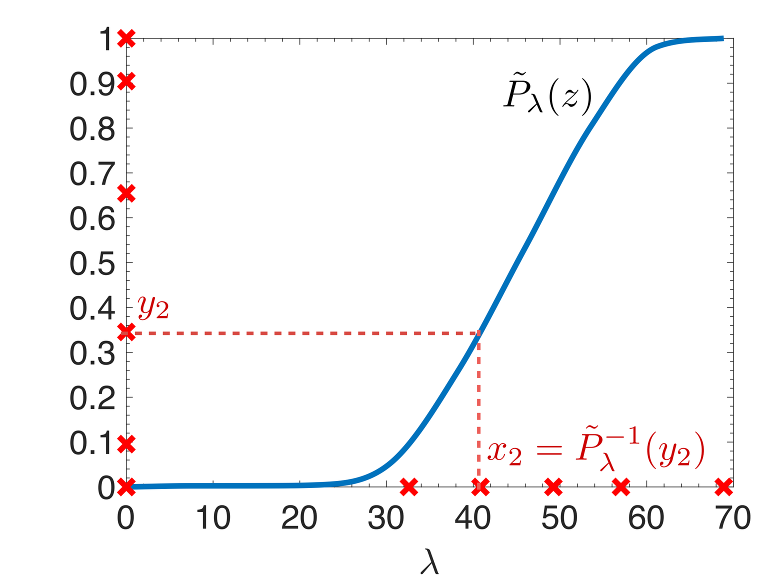


- Examples:



SPECTRUM-ADAPTED METHODS

I. Spectrum-Adapted Interpolation



- $K+1$ Chebyshev nodes $\{y_i\}$ on $[0,1]$
- Warp them via $x_i = \tilde{P}_\lambda^{-1}(y_i)$
- Find the unique order K polynomial $p_K(\lambda)$ through points $\{(x_i, f(x_i)\}$
- Compute $p_K(\mathbf{A})\mathbf{b}$ recursively

II. Spectrum-Adapted Regression / Orthogonal Polynomial Expansion

- Weighted least squares problem, with M abscissae $\{x_m\}$ and weights from the spectral PDF $\{w_m\} = \tilde{p}_\lambda(\{x_m\})$

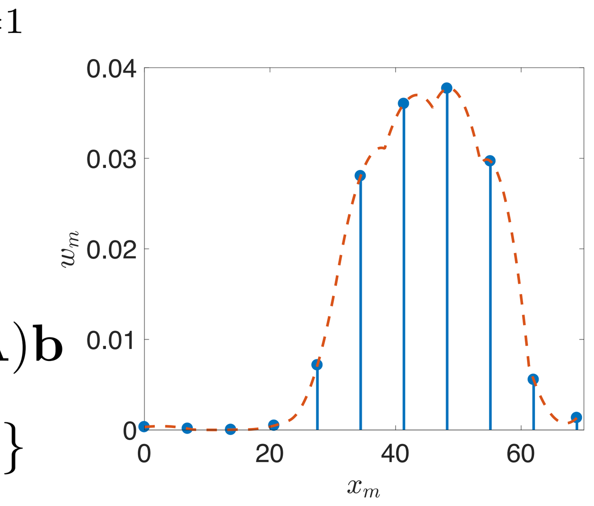
$$p_K(\lambda) = \min_{p\in\mathcal{P}_K} \sum_{m=1}^{M} w_m\,[f(x_m) - p(x_m)]^2$$

- Equivalently, a truncated expansion in a basis of polynomials $\{\pi_{k,M}(\lambda)\}$ [4,5]

$$= \sum_{k=0}^{K} \frac{\langle f, \pi_{k,M}\rangle_{d\lambda_M}}{\langle \pi_{k,M}, \pi_{k,M}\rangle_{d\lambda_M}} \pi_{k,M}(\lambda)$$

- $\{\pi_{k,M}(\lambda)\}$ are orthogonal polynomials with respect to the discrete measure generated from the spectral PDF

$$\langle f, g\rangle_{d\lambda_M} = \sum_{m=1}^{M} w_m f(x_m) g(x_m)$$

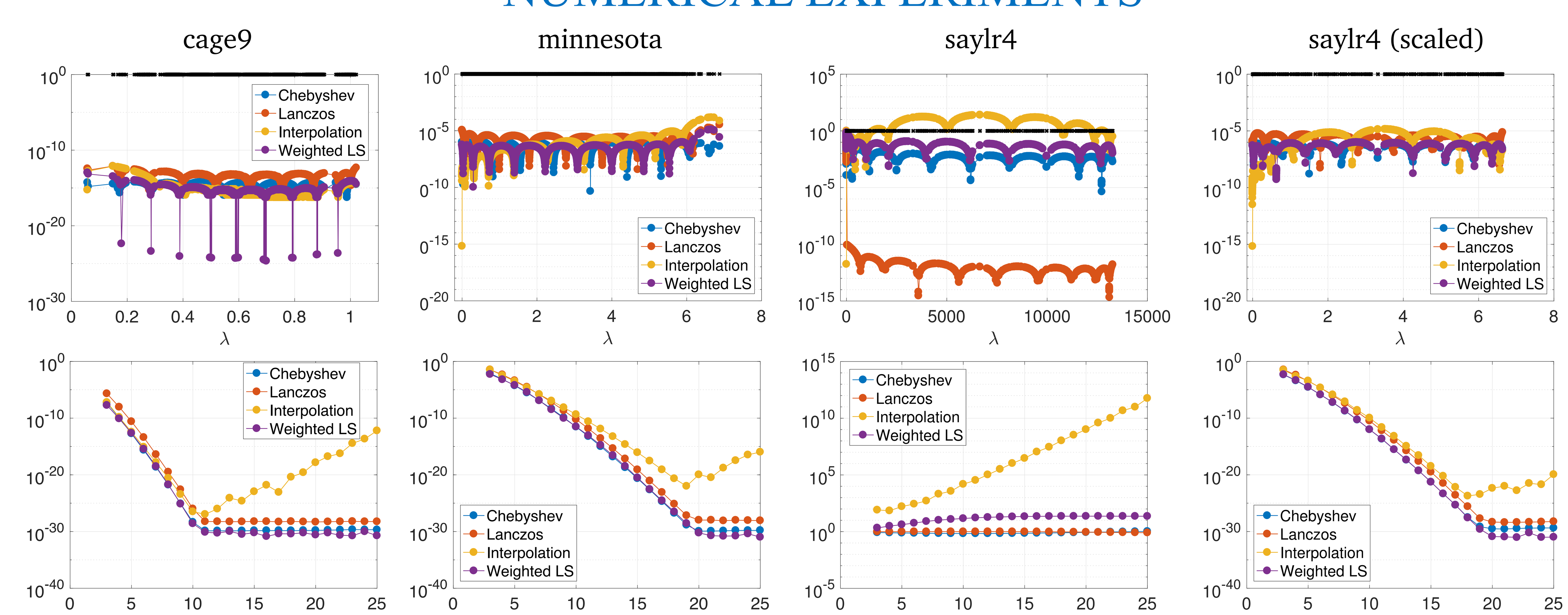


- Evaluate each $\pi_{k,M}(\mathbf{A})\mathbf{b}$ via a three-term recursion for $k = 0, 1, \ldots, K-1$ ($K \le M-1$)

$$\pi_{k+1,M}(\mathbf{A})\mathbf{b} = (\mathbf{A} - \alpha_{k,M}\mathbf{I}_N)\pi_{k,M}(\mathbf{A})\mathbf{b} - \beta_{k,M}\pi_{k-1,M}(\mathbf{A})\mathbf{b}$$

- Compute $p_K(\mathbf{A})\mathbf{b}$ as a linear combination of $\{\pi_{k,M}(\mathbf{A})\mathbf{b}\}$

NUMERICAL EXPERIMENTS



- Spectrum-adapted interpolation works well for low polynomial orders, but is unstable at higher orders due to ill-conditioning
- The Lanczos method is more stable with respect to the width of the spectrum
- Spectrum-adapted regression tends to outperform the Lanczos method for matrices with many distinct interior eigenvalues