Spectrum-Adapted Polynomial Approximation for Matrix Functions

Li Fan, David I Shuman

Dept. of Math, Statistics and Computer Science, Macalester College – St. Paul, MN

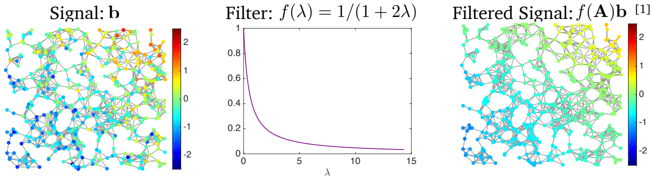
MATRIX FUNCTION × VECTOR: $f(\mathbf{A})\mathbf{b}$

- Large, sparse, Hermitian matrix $\mathbf{A} \in \mathbb{R}^{N \times N} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$
- Orthonormal eigenvectors with eigenvalues

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

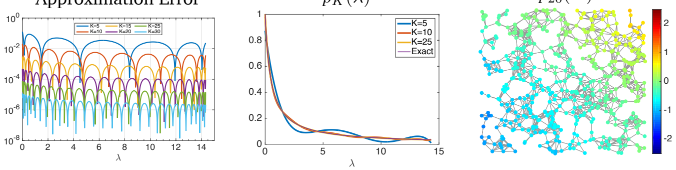
- $f(\mathbf{A})\mathbf{b}$ is widely used in signal processing, machine learning, applied math, computational science, etc., but prohibitively expensive to compute directly

Signal: $\mathbf{b}$ — Filter: $f(\lambda) = 1/(1+2\lambda)$ — Filtered Signal: $f(\mathbf{A})\mathbf{b}$ [1]



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

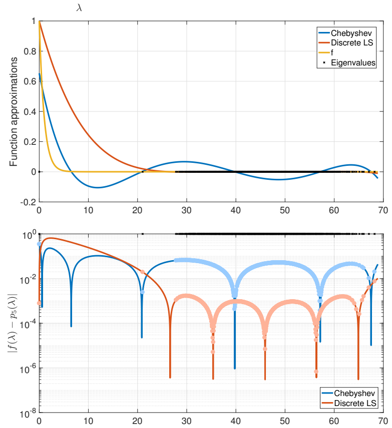
Approximation Error — $p_K(\lambda)$ — $p_{20}(\mathbf{A})\mathbf{b}$



- However, error in $f(\mathbf{A})\mathbf{b}$ depends only on residuals at eigenvalues of $\mathbf{A}$

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

- Develop methods to approximate $f(\mathbf{A})\mathbf{b}$ with $p_K(\mathbf{A})\mathbf{b}$, focusing 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

Research sponsored by the NSF and HHMI.

[1] Shuman et al., "Chebyshev polynomial approximation for distributed signal processing," DCOSS, 2011
[2] Shuman et al., "Distributed signal processing via Chebyshev polynomial approximation," SIPN, 2018
[3] Lin et al., "Approximating spectral densities of large matrices," SIAM Review, 2016
[4] Gautschi, Orthogonal Polynomials: Computation and Approximation, 2004

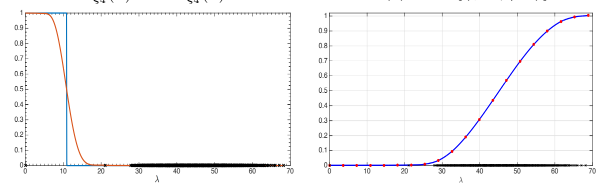
SPECTRAL DENSITY ESTIMATION: KERNAL POLYNOMIAL METHOD (KPM)

- 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

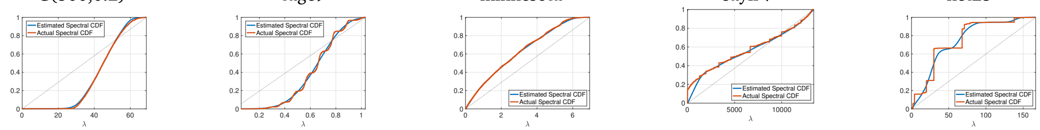
$$\eta_i = \mathrm{tr}(\Theta_{\xi_i}(\mathbf{A})) = \mathbb{E}[\mathbf{x}^\top \Theta_{\xi_i}(\mathbf{A})\mathbf{x}] \approx \frac{1}{J}\sum_{j=1}^{J} \mathbf{x}^{(j)\top} \tilde{\Theta}_{\xi_i}(\mathbf{A})\mathbf{x}^{(j)}$$

- 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 $\tilde{P}_\lambda^{-1}(z)$

$\Theta_{\xi_4}(\lambda)$ and $\tilde{\Theta}_{\xi_4}(\lambda)$ — $\tilde{P}_\lambda(z)$ with $\{(\xi_i, \eta_i/N)\}$

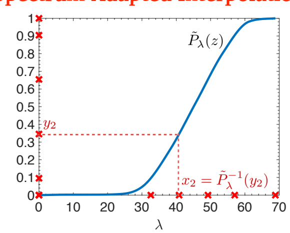


- Examples: G(500,0.2) — cage9 — minnesota — saylr4 — net25



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 $p_K(\mathbf{A})\mathbf{b} = \sum_{k=0}^{K} c_k T_k(\mathbf{A})\mathbf{b}$

II. Spectrum-Adapted Regression / Orthogonal Polynomial Expansion

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

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

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

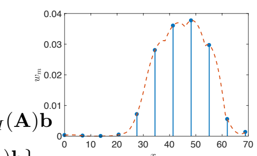
$$p_K(\lambda) = \sum_{k=0}^{K} \frac{\langle f, \pi_{k,M}\rangle_{d\lambda_M}}{\langle \pi_{k,M}, \pi_{k,M}\rangle_{d\lambda_M}} \pi_{k,M}(\lambda)$$

- $\{\pi_{k,M}(\lambda)\}$ are orthogonal polynomials with respect to the discrete measure generated from 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 $k = 0, 1, \ldots, K-1$ $(K \leq M-1)$

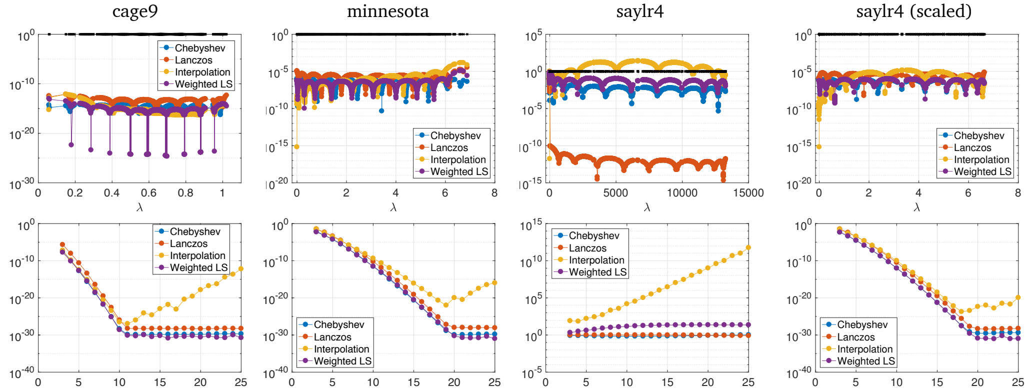
$$\pi_{k+1,M}(\mathbf{A})\mathbf{b} = (\mathbf{A} - \alpha_{k,M}\mathbf{I}_N)\pi_{k,M}(\mathbf{A})\mathbf{b} - \beta_{k,M}\pi_{k-1,M}(\mathbf{A})\mathbf{b}$$



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

NUMERICAL EXPERIMENTS

cage9 — minnesota — saylr4 — saylr4 (scaled)



- 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 outperforms the Lanczos method for matrices with many distinct interior eigenvalues.