# 1 Approximation of Matrix Functions via Spectral Warping

Assume a matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ is symmetric positive semi-definite. Then, $\mathbf{A}$ has a set of $N$ orthonormal eigenvectors $\{u_i\}$ and $N$ corresponding real, non-negative eigenvalues $\{\lambda_i\}$, $i = 0, 1, \cdots, N-1$, such that $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for each $i$. Thus, $\mathbf{A}$ is diagonalizable with the following spectrum decomposition:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\star,$$

where the $(i+1)_{th}$ column of $\mathbf{U}$ is $\mathbf{u}_i$, and the $(i+1)_{th}$ diagonal element of $\mathbf{\Lambda}$ is $\lambda_i$. Denote the spectrum of $\mathbf{A}$ as $\sigma(\mathbf{A}) := \{\lambda_0, \lambda_1, \cdots, \lambda_{N-1}\}$, and assume $\lambda_{\min} = \lambda_0 \le \lambda_1 \le \cdots \le \lambda_{N-1} = \lambda_{\max}$.

Given a function $f$ well-defined on $\sigma(\mathbf{A})$, the matrix function $f(\mathbf{A})$ is defined as:

$$f(\mathbf{A}) := \mathbf{U} f(\mathbf{\Lambda})\mathbf{U}^\star := \mathbf{U} \begin{bmatrix} f(\lambda_0) & & & \\ & f(\lambda_1) & & \\ & & \ddots & \\ & & & f(\lambda_{N-1}) \end{bmatrix} \mathbf{U}^\star. \tag{1}$$

In numerical computation, $f$ is often approximated with a polynomial $p$, such that

$$f(\mathbf{A}) = \mathbf{U} f(\mathbf{\Lambda})\mathbf{U}^\star \approx \mathbf{U} p(\mathbf{\Lambda})\mathbf{U}^\star.$$

Therefore, the approximation of $f(\mathbf{A})$ essentially depends on $p(\lambda)$, $\lambda \in \sigma(\mathbf{A})$.

In order to adapt general methods of approximating functions defined on a closed interval for a matrix function defined on a discrete set $\sigma(\mathbf{A})$, we introduce the idea of spectral warping. It warps the sample according to the spectrum distribution of $\mathbf{A}$ such that the distribution of the sample points resembles that of $\sigma(\mathbf{A})$. Then, we apply the interpolation or fitting methods to the warped sample to obtain an estimation of the matrix function with focus on the discrete set $\sigma(\mathbf{A})$.

Given the matrix $\mathbf{A}$ as described above, we estimate the cumulative distribution function (CDF) of its eigenvalues with the Kernel Polynomial Method (KPM). Denote the CDF as $F : [\lambda_{\min}, \lambda_{\max}] \to [0, 1]$, we can find its inverse function $F^{-1} : [0, 1] \to [\lambda_{\min}, \lambda_{\max}]$.

## 1.1 The Warped Chebyshev Interpolation

The Chebyshev polynomials $\{T_k(y)\}$ are a set of orthogonal polynomials with respect to the measure $d\lambda = \frac{dy}{\sqrt{1-y^2}}$ defined on the interval $[-1, 1]$. They follow the recurrence relationship:

$$\begin{gathered} T_0(y) = 1,\ T_1(y) = y, \\ T_k(y) = 2yT_{k-1}(y) - T_{k-2}(y),\ \ k \ge 2. \end{gathered} \tag{2}$$

Since $f$ is defined on $\sigma(\mathbf{A})$, we shift the Chebyshev polynomials to the interval $[\lambda_{\min}, \lambda_{\max}]$ by setting $x = \frac{\lambda_{\max}+\lambda_{\min}}{2} + \frac{\lambda_{\max}-\lambda_{\min}}{2} y$. Thus the shifted Chebyshev polynomials satisfy $\overline{T}_k = T_k(\frac{2x-\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}-\lambda_{\min}})$ and follow a similar recurrence relationship to (2).

To obtain an order $K$ warped Chebyshev interpolation of $f$, we begin with $K+1$ Chebyshev nodes $\{y_i\}$ shifted to $[0,1]$:

$$y_i = \frac{1}{2} + \frac{1}{2}\frac{\cos(2i-1)}{2(K+1)}, \;\; i = 1, 2, \cdots, K+1. \tag{3}$$

Next, we warp these points with the inverse CDF:

$$\{x_i\} = F^{-1}(\{y_i\}), \tag{4}$$

so that the warped points $\{x_i\}$ are in the interval $[\lambda_{\min}, \lambda_{\max}]$ with higher density in subintervals with larger eigenvalue counts. Then, we interpolate the $K+1$ warped points $(\{x_i\}, f\{x_i\})$ with the shifted Chebyshev polynomials $\{\overline{T}_0, \overline{T}_1, \cdots, \overline{T}_K\}$ for an order $K$ interpolation of $f$:

$$f(x) \approx \sum_{k=0}^{K} a_k \overline{T}_k(x), \tag{5}$$

where $\{a_k\}$ are referred to as the interpolation coefficients.

In order to attenuate the Gibbs oscillations of the interpolant, we use the technique of $\sigma$-smoothing proposed by Lanczos[ref], which mulitplies each interpolation coefficient $\{a_k\}$ by a damping coefficient $\{\gamma_k^K\}$, defined as:

$$\begin{aligned} \gamma_0^K &= 1, \\ \gamma_k^K &= \frac{\sin(\frac{k\pi}{K+1})}{\frac{k\pi}{K+1}}, \;\; k = 1, 2, \cdots, K. \end{aligned} \tag{6}$$

Thus, the damped Chebyshev interpolation can be written as:

$$f(x) \approx \sum_{k=0}^{K} \gamma_k^K a_k \overline{T}_k(x). \tag{7}$$

For any $\mathbf{b} \in \mathbb{R}^N$, the estimate for $f(\mathbf{A})\mathbf{b}$ from this interpolation is given by

$$f(\mathbf{A})\mathbf{b} \approx \sum_{k=0}^{K} \gamma_k^K a_k \overline{T}_k(\mathbf{A})\mathbf{b}, \tag{8}$$

where $\{\overline{T}_k(\mathbf{A})\mathbf{b}\}$ can be computed recursively from (2) (do we need an equation of shifted version of recurrence?).

### 1.2 The Warped PCHIP with Chebyshev Approximation

We begin with the same set of $K+1$ warped points $\{x_i\}$ from (3) and (4), and interpolate a cubic Hermite polynomial between each pair of adjacent points $\{(x_i, f(x_i)), (x_{i+1}, f(x_{i+1}))\}$, $i = 1, 2, \cdots, K$, to build the interpolant $\hat{f}$. On each subinterval $[x_i, x_{i+1}]$, $\hat{f}$ is a cubic polynomial that preserves the shape and monotonicity of $f$ [ref], resulting in a more stable approximation without rapid oscillations, an advantage over interpolation methods on the entire interval(better ways to phrase this?) especially when $f$ is monotonic.

However, the computation of $\hat{f}(\mathbf{A})\mathbf{b}$ as an approximation to $f(\mathbf{A})\mathbf{b}$ is challenging as $\hat{f}$ is defined piecewise. Consequently, we approximate $\hat{f}$ again with a truncated Chebyshev expansion of order $K$.

Since $\{\overline{T}_k\}$ form an orthogonal basis for functions well-defined on $[\lambda_{\min}, \lambda_{\max}]$ and square integrable with respect to $d\lambda$, $\hat{f}$ can be represented as an infinite series of Chebyshev polynomial expansion, and approximated by a truncated expansion of order $K$:

$$\hat{f}(x) = \frac{1}{2}c_0 + \sum_{k=1}^{\infty} c_k \overline{T}_k(x) \approx \frac{1}{2}c_0 + \sum_{k=1}^{K} c_k \overline{T}_k(x), \tag{9}$$

where the Chebyshev expansion coefficients $\{c_k\}$ can be determined by

$$c_k := \left\langle \hat{f}, \overline{T}_k \right\rangle = \frac{2}{\pi} \int_0^{\pi} \cos(k\phi) \hat{f}\left( \frac{\lambda_{\max} + \lambda_{\min}}{2} + \frac{\lambda_{\max} - \lambda_{\min}}{2} \cos(\phi) \right) d\phi. \tag{10}$$

Thus, we have the following approximation to $f$:

$$f(x) \approx \hat{f}(x) \approx \frac{1}{2}c_0 + \sum_{k=1}^{K} c_k \overline{T}_k(x). \tag{11}$$

Similarly, we can apply $\sigma$-smoothing to the Chebyshev expansion with $\{\gamma_k^K\}$ from (6) to damp the Gibbs oscillations.

For any $\mathbf{b} \in \mathbb{R}^N$, the estimate for $f(\mathbf{A})\mathbf{b}$ from this method is given by

$$f(\mathbf{A})\mathbf{b} \approx \frac{1}{2}c_0\mathbf{b} + \sum_{k=1}^{K} \gamma_k^K c_k \overline{T}_k(\mathbf{A})\mathbf{b}, \tag{12}$$

where $\{\overline{T}_k(\mathbf{A})\mathbf{b}\}$ can be computed recursively from (2) (do we need an equation of shifted version of recurrence?).

### 1.3 The Warped Least Squares Approximation

We choose the sample size $n = N/10$, where $N$ is the size of matrix $\mathbf{A}$. (better way to state sample size? it depends on G.N and can't be smaller than K+1) Take $K = n - 1$ in (3), we begin with a set of $n$ Chebyshev nodes $\{y_i\}$ and warp them as in (4). Then, a Least Squares polynomial fitting of order $K$ is performed on the $n$ points $\{(x_i, f(x_i))\}$. Denote the fitted polynomial as $p$, we have the following approximation to $f$:

$$f(x) \approx p(x) = \sum_{k=0}^{K} \beta_k x^k. \tag{13}$$

For any $\mathbf{A} \in \mathbb{R}^{N \times N}$, the estimate for $f(\mathbf{A})$ is given by

$$f(\mathbf{A}) \approx p(\mathbf{A}) = \beta_0 \mathbf{I} + \sum_{k=1}^{K} \beta_k \mathbf{A}^k, \tag{14}$$