Fast Polynomial Transforms Based on Toeplitz and Hankel Matrices

Alex Townsend[1] **Marcus Webb**[2] Sheehan Olver[3]

[1]MIT [2]University of Cambridge [3]University of Sydney

1. A Tale of Two Bases: Chebyshev & Legendre

- To **approximate a function** (e.g. a signal), we can expand in Chebyshev or Legendre polynomial series and work with the vector of coefficients:

$$f(x) = \sum_{k=0}^{N} c_k^{cheb} T_k(x) = \sum_{k=0}^{N} c_k^{leg} P_k(x), \qquad x \in [-1,1].$$

- The **Chebyshev polynomials** $T_k(x) = \cos(k \cos^{-1}(x))$ have good approximation properties and fast transforms, due to their **link to Fourier series**.
- The **Legendre polynomials** $P_k(x)$ are **orthogonal** in L^2 inner product:

$$\int_{-1}^{1} P_j(x) P_k(x)\,\mathrm{d}x = 0 \text{ if } j \neq k.$$

- Hence Legendre is **better** than Chebyshev in **some situations**:
  - Fourier transform of $P_k(x)$ is simpler $\longrightarrow$ **signal processing**
  - Faster algorithms for convolution of Legendre expansions $\longrightarrow$ **smoothing** a signal, **sums of random variables**
  - $P_k(x)$ has a rapidly decaying Cauchy transform $\longrightarrow$ Riemann-Hilbert problems, **random matrix theory** and **integrable systems**

2. Legendre-to-Chebyshev Conversion Matrix

- To **convert** from Chebyshev coefficients to Legendre coefficients, we compute a matrix-vector multiplication:

$$\underline{c}^{cheb} = M\underline{c}^{leg}, \quad M_{jk} = \begin{cases} \frac{1}{\pi}\Lambda\left(\frac{k}{2}\right)^2, & 0 = j \leq k \leq N,\ j \text{ even}, \\ \frac{2}{\pi}\Lambda\left(\frac{k-j}{2}\right)\Lambda\left(\frac{k+j}{2}\right), & 0 < j \leq k \leq N,\ k-j \text{ even}, \\ 0, & \text{otherwise}, \end{cases}$$

where $\Lambda(z) = \Gamma(z+1/2)/\Gamma(z+1)$, $\Gamma(z)$ is the gamma function.

- Directly computing $M\underline{c}^{leg}$ would take $\mathcal{O}(N^2)$ operations. **Too slow!**
- There are **quasi-linear algorithms** (i.e. $\mathcal{O}(N(\log N)^k)$ operations) due to Orszag (1986), Alpert-Rokhlin (1991), Potts-Steidl-Tasche (1998), Keiner (2009), Iserles (2011) and Hale-Townsend (2013).
- Our $\mathcal{O}(N(\log N)^2)$ algorithm is based on **decomposing the matrix** M:

$$M = D(T \circ H), \qquad D = \mathrm{diag}(1/\pi, 2/\pi, \ldots, 2/\pi)$$

$$H_{jk} = \Lambda\left(\tfrac{j+k}{2}\right), \qquad T_{jk} = \begin{cases} \Lambda\left(\frac{k-j}{2}\right), & 0 \leq j \leq k \leq N, k-j \text{ even}, \\ 0, & \text{otherwise}, \end{cases}$$

where $\circ$ is the **Hadamard product** (elementwise product). T is a **Toeplitz matrix** and H is a **Hankel matrix**.

3. Overview of New Fast Algorithm

- **Fact 1**: The Toeplitz matrix T in box 2 can be applied to a vector in quasilinear operations using the **Fast Fourier Transform** (FFT).
- **Fact 2**: Note the following identity for the Hadamard product of a matrix A with a **rank 1 matrix** $\underline{v}\,\underline{w}^T$, (where $\underline{v} = (v_0, v_1, \ldots, v_N)^T$):

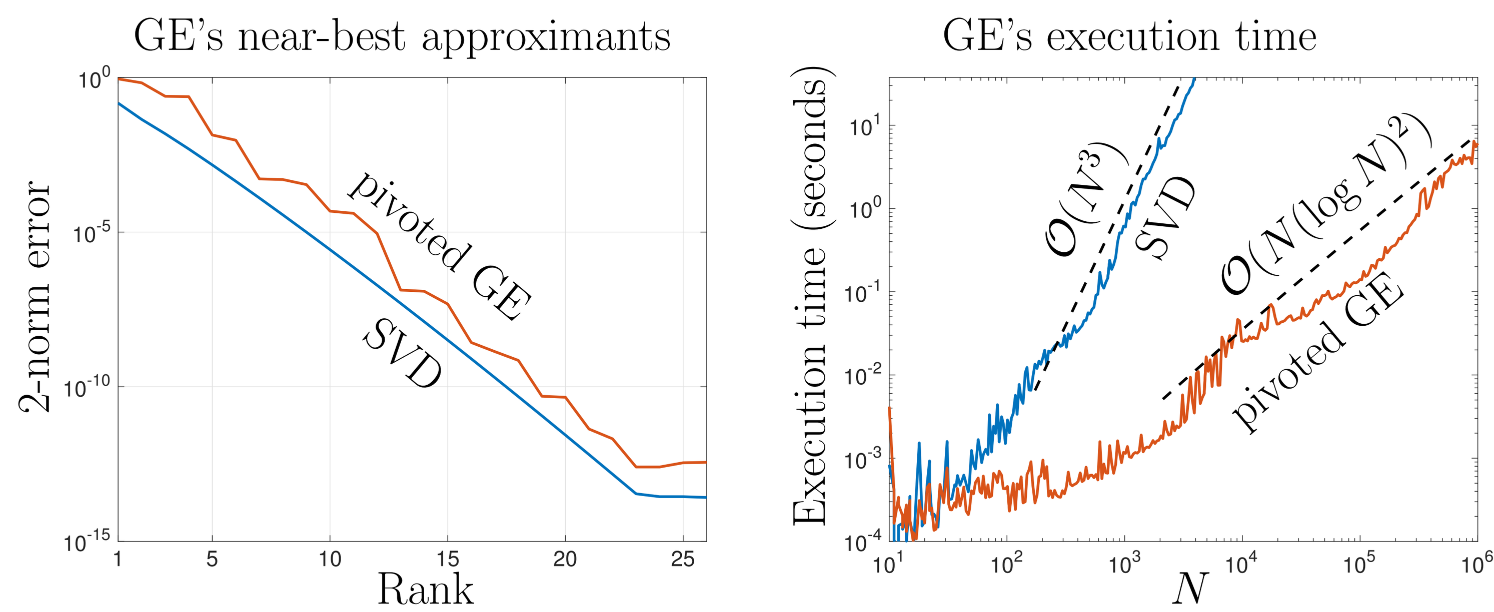
$$A \circ \underline{v}\,\underline{w}^T = D_{\underline{v}} A D_{\underline{w}},$$

where $D_{\underline{v}} = \mathrm{diag}(v_0, v_1, \ldots, v_N)$. Diagonal matrices can be applied to a vector in linear time, so matrix-vector multiplication for $A \circ \underline{v}\,\underline{w}^T$ can be **computed in quasilinear operations** if and only if it can for be for A.

Steps for computing $\underline{c}^{cheb} = M\underline{c}^{leg}$	**Cost**
1. Decompose M into $M = D(T \circ H)$ (see box 2)	$\mathcal{O}(N)$
2. Calculate low rank approx. $H \approx \sum_{j=1}^{K} a_r \underline{v}_j \underline{v}_j^T$ (see box 4)	$\mathcal{O}(N(\log N)^2)$
3. Compute $\underline{w} = (T \circ H)\underline{c}^{leg}$ (using Fact 1 and Fact 2)	$\mathcal{O}(N(\log N)^2)$
4. Compute $\underline{c}^{cheb} = D\underline{w}$	$\mathcal{O}(N)$

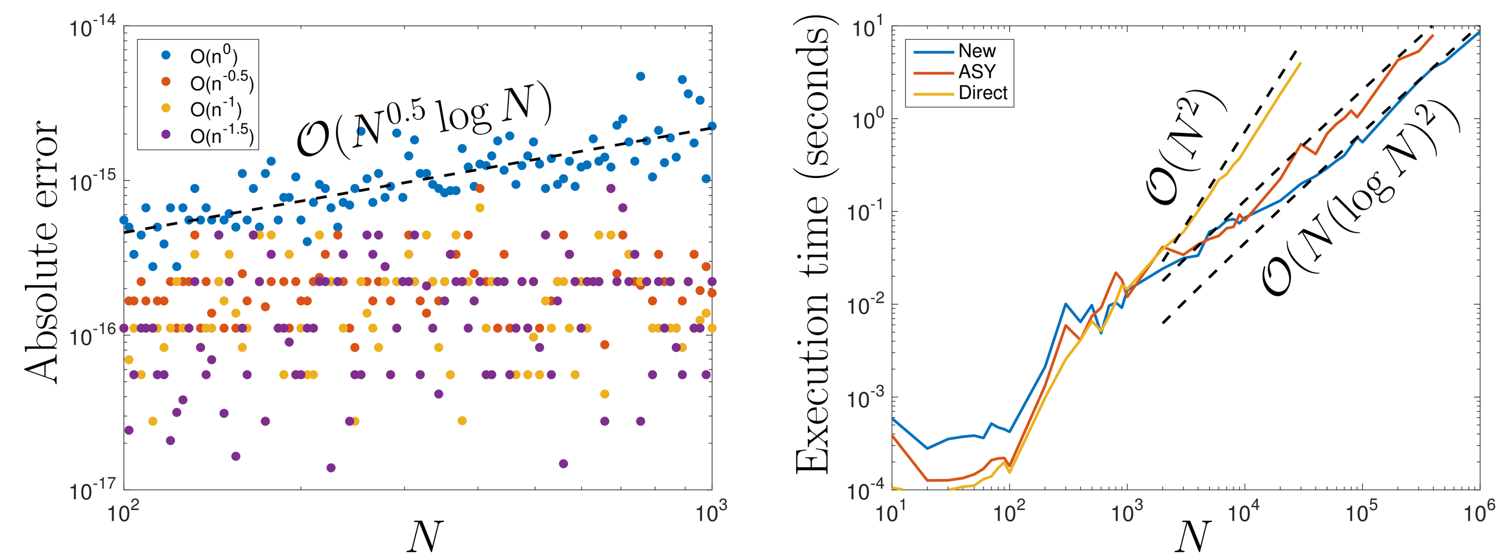
4. Computing Low Rank Approximations by Pivoted Gaussian Elimination

- **Pivoted GE:** For a matrix A, repeat the following iteration:
  1. $j, k = \underset{0 \leq s,t \leq N}{\mathrm{argmax}} |A_{st}|$ (choose pivot)
  2. $A \longleftarrow A - (A_{j,k})^{-1}(A_{*,k} A_{j,*})$ (subtract rank 1 update)
- At each iteration, the sum of those **rank 1 updates** is a **low rank approximation** to the original matrix A.
- **Theorem:** the approximate rank of the Hankel matrix H in box 2 is $\mathcal{O}(\log N)$.
- For a symmetric, positive definite matrix (like H in box 2), we can find a rank K approximant in $\mathcal{O}(K^2 N)$ operations. For our H, $K = \mathcal{O}(\log N)$.



Left: Pivoted GE's low rank approximants are **close to optimal** Singular Value Decomposition (SVD). Right: But they are much faster to compute!

5. Comparison with State of the Art



Left: Observed errors computing $\underline{c}^{cheb} = M\underline{c}^{leg}$ with various decay rates in $\underline{c}^{leg}$. Hale-Townsend (2013) method has $\mathcal{O}(N)$ error growth for $\mathcal{O}(n^0)$ decay rate. Right: Execution times between the direct (yellow), Hale-Townsend (2013) asymptotics method (red), and the new algorithm (blue).

6. Chebyshev-to-Legendre Conversion and More Polynomial Transforms

- The Chebyshev-to-Legendre conversion matrix $\underline{c}^{leg} = L\underline{c}^{cheb}$ has a similar structure:

$$L_{jk} = \begin{cases} 1, & j = k = 0, \\ \frac{\sqrt{\pi}}{2\Lambda(j)}, & 0 < j = k \leq N, \\ -k(j+\frac{1}{2})\frac{\Lambda\left(\frac{k-j-2}{2}\right)}{k-j} \cdot \frac{\Lambda\left(\frac{j+k-1}{2}\right)}{j+k+1}, & 0 \leq j < k \leq N,\ k-j \text{ even}, \end{cases}$$

and so we can use the same techniques.

- Converting between Ultraspherical (Gegenbauer) polynomials $C_k^{\lambda_1}(x)$ and $C_k^{\lambda_2}(x)$ also has this structure (cf. Keiner (2009), Cantero-Iserles (2013)):

$$A_{jk} = \begin{cases} c_{\lambda_1,\lambda_2}(j+\lambda_2)\frac{\Gamma\left(\frac{k-j}{2}+\lambda_1-\lambda_2\right)}{\Gamma\left(\frac{k-j}{2}+1\right)} \cdot \frac{\Gamma\left(\frac{k+j}{2}+\lambda_1\right)}{\Gamma\left(\frac{k+j}{2}+\lambda_2+1\right)}, & 0 \leq j \leq k,\ k-j \text{ even}, \\ 0, & \text{otherwise}. \end{cases}$$

- Converting between Jacobi polynomial expansions can also be computed in $\mathcal{O}(N(\log N)^2)$ operations with this approach! (cf. Wang-Huybrechs (2015))

LONDON MATHEMATICAL SOCIETY EST. 1865 CCA EPSRC Engineering and Physical Sciences Research Council

Funded by LMS Cecil King Travel Scholarship and Cambridge Centre for Analysis EPSRC Studentship.