Multiresolution Matrix Factorization (ICML 2014) — Supplement

RISI KONDOR, NEDELINA TENEVA AND VIKAS K. GARG

Proof of Proposition 1. By the nestedness of $S_0 \supseteq S_1 \supseteq \ldots \supseteq S_L$, for some sequence of permutation matrices $\Pi_1, \ldots, \Pi_L$, H decomposes recursively as

$$[H]_{S_\ell,S_\ell} = \Pi_\ell \begin{pmatrix} [H]_{S_{\ell+1},S_{\ell+1}} & [H]_{S_{\ell+1},\,J_{\ell+1}} \\ [H]_{J_{\ell+1},\,S_{\ell+1}} & [H]_{J_{\ell+1},\,J_{\ell+1}} \end{pmatrix} \Pi_\ell^\top$$

Unwrapping this recursion tells us that $\|H\|^2_{\text{resi}}$ is equal to

$$\sum_{\ell=1}^{L} \Big[\|[H]_{J_\ell,\,S_\ell}\|^2_{\text{Frob}} + \|[H]_{S_\ell,\,J_\ell}\|^2_{\text{Frob}} + \|[H]_{J_\ell,\,J_\ell}\|^2_{\text{off-diag}} \Big].$$

However, since the rotations $U_{\ell+1}, \ldots, U_L$ leave $\mathrm{span}(\{\, e_i \mid i \in [n] \setminus S_\ell \,\})$ invariant,

$$\|[A_\ell]_{J_\ell,\,S_\ell}\|^2_{\text{Frob}} = \|[A_{\ell+1}]_{J_\ell,\,S_\ell}\|^2_{\text{Frob}} = \ldots = \\ = \|[A_L]_{J_\ell,\,S_\ell}\|^2_{\text{Frob}} == \|[H]_{J_\ell,\,S_\ell}\|^2_{\text{Frob}}.$$

By symmetry, $\|[H]_{S_\ell,\,J_\ell}\|^2_{\text{Frob}} = \|[H]_{J_\ell,\,S_\ell}\|^2_{\text{Frob}}$. Similarly, $\|[A_\ell]_{J_\ell,\,J_\ell}\|^2_{\text{off-diag}} = \ldots = \|[H]_{J_\ell,\,J_\ell}\|^2_{\text{off-diag}}$. ■

Proof of Proposition 2. Since $J = \{i_k\}$, by Proposition 1

$$\mathcal{E}_\ell = 2 \sum_{p=1}^{k-1} [U_\ell A_{\ell-1} U_\ell^\top]^2_{i_k,i_p} + 2\, \|[U_\ell A_{\ell-1} U_\ell^\top]_{i_k,S_\ell}\,\|^2.$$

The first term can be written $2 \sum_{p=1}^{k-1} [O[A_{\ell-1}]_{\mathrm{I},\mathrm{I}} O^\top]^2_{k,p}$, while the second term is

$$2\, \|\, [O[A_{\ell-1}]_{\mathrm{I},S_\ell} [U_\ell]^\top_{S_\ell,S_\ell}]_{k,:}\, \|^2 = \\ 2\, [O[A_{\ell-1}]_{\mathrm{I},S_\ell} [U_\ell]^\top_{S_\ell,S_\ell} [U_\ell]_{S_\ell,S_\ell} [A_{\ell-1}]^\top_{\mathrm{I},S_\ell} O^\top]_{k,k} = \\ 2\, [O[A_{\ell-1}]_{\mathrm{I},S_\ell} [A_{\ell-1}]^\top_{\mathrm{I},S_\ell} O^\top]_{k,k} = 2\, [OBO^\top]_{k,k}$$

■

Proof of Proposition 3. Analogous to the proof of Proposition 2, but summed over each $\mathrm{I}_1 \times \mathrm{I}_1, \ldots, \mathrm{I}_m \times \mathrm{I}_m$ block.

Proof of Proposition 4. We want to minimize

$$\phi(\alpha) = \left(\left[O_\alpha \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} O_\alpha^\top \right]_{2,1} \right)^2 + \left[O_\alpha \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix} O_\alpha^\top \right]_{2,2}.$$

Expanding, we get

$$\phi(\alpha) = ((A_1 - A_2) \sin\alpha \cos\alpha)^2 + B_{1,1} (\sin\alpha)^2 + \\ 2 B_{1,2} \sin\alpha \cos\alpha + B_{2,2} (\cos\alpha)^2 = \\ \left(\frac{A_1 - A_2'}{2} \right)^2 (\sin(2\alpha'))^2 + \\ B_{1,2} \sin(2\alpha) + (\sin\alpha')^2 B_{1,1} + (\cos\alpha')^2 B_{2,2}.$$

Rewriting the second two terms as

$$\frac{((\sin\alpha)^2 + (\cos\alpha')^2)(B_{1,1} + B_{2,2})}{2} + \\ \frac{((\sin\alpha)^2 - (\cos\alpha')^2)(B_{1,1} - B_{2,2})}{2}$$

gives

$$\phi(\alpha) = \left(\frac{A_1 - A_2}{2} \right)^2 (\sin(2\alpha))^2 + B_{1,2} \sin(2\alpha) + \\ \frac{B_{1,1} + B_{2,2}}{2} + \frac{B_{2,2} - B_{1,1}}{2} \cos(2\alpha).$$

Introducing $d = (B_{2,2} - B_{1,1})/2$ and the other variables a, b, c, e and θ gives the new objective function

$$\psi(\theta) = a (\sin\theta)^2 + b \sin\theta + c \cos\theta + d.$$

Setting the derivative with respect to θ zero,

$$2a \sin\theta \cos\theta + b \cos\theta - c \sin\theta = 0.$$

Again using $\sin(2x) = 2 \sin x \cos x$,

$$a \sin(2\theta) + b \cos\theta - c \sin\theta = 0.$$

Now letting $e = \sqrt{b^2 + c^2}$ and $\omega = \arctan(c/b)$

$$a \sin(2\theta) + e (\cos\omega \cos\theta - \sin\omega \sin\theta) = 0.$$

Using $\cos(x + y) = \cos x \cos y - \sin x \sin y$,

$$(a/e) \sin(2\theta) + \cos(\theta + \omega) = 0,$$

which is finally equivalent to (23). ■

Proof of Theorem 1. Let ψ be a specific wavelet ψ^ℓ_m, with support $S = \{s_1, \ldots, s_K\} = \mathrm{supp}(\psi) \subseteq [n]$; f_S and ψ_S be the restriction of f and ψ to S regarded as a vectors;

and Q, D and $\tilde{Q}$ be defined as in Definition 5. The Hölder property then gives

$$f_S^\top \tilde{L} f_S = \sum_{i,j=1}^{K} \tilde{Q}_{i,j}(f(s_i)-f(s_j))^2 \leq$$
$$\leq \sum_{i,j=1}^{K} c_T\, Q_{i,j}(f(s_i)-f(s_j))^2 \leq c_T c_H K^2, \quad (26)$$

where $\tilde{L} = I - \tilde{Q}$ is the normalized Laplacian. At the same time, if ψ_m^ℓ comes from row/column i of A_ℓ, then by (11), $[A_\ell]_{:,i} = U_\ell \ldots U_1 A\psi$, and therefore

$$\psi_S^\top \tilde{Q} \psi_S \leq c_T\, \psi_S^\top Q \psi_S \leq c_T\, \psi_S^\top A_{S,:} A_{:,S} \psi_S =$$
$$= c_T \|A\psi\|^2 = c_T\, \|[A_\ell]_{:,i}\|^2 = c_T\, \|H_{:,i}\|^2 \leq c_T \epsilon \quad (27)$$

Clearly, $\tilde{Q}$ and $\tilde{L}$ share the same normalized eigenbasis $\{v_1, \ldots, v_n\}$. Letting $\lambda_1, \ldots, \lambda_K$ be the corresponding eigenvalues, $f_i = \langle f_S, v_i \rangle$ and $\psi_i = \langle \psi_S, v_i \rangle$ and taking any $\gamma > 0$

$$\sum_{i=1}^{K} \Big(\sqrt{\gamma \lambda_i}\, \psi_i - \frac{1}{\sqrt{\gamma \lambda_i}}\, f_i \Big)^2 \geq 0, \quad (28)$$

which implies

$$\langle f, \psi \rangle = \langle f_S, \psi_S \rangle \leq \frac{1}{2} \left[\gamma\, \psi_S^\top \tilde{Q} \psi_S + \gamma^{1/2} f_S^\top \tilde{Q}^{-1} f_S \right].$$

The first term on the r.h.s of this inequality is bounded by (27), while for any $c_\Lambda \geq 4/(1-(1-2\Lambda)^2)$, by (26),

$$f_S^\top \tilde{Q}^{-1} f_S = \sum_{i=1}^{K} \frac{1}{\lambda_i}\, f_i^2 \leq c_\Lambda \sum_{i=1}^{K} (1-\lambda_i) f_i^2 =$$
$$= c_\Lambda f_S^\top \tilde{L} f_S \leq c_T c_H c_\Lambda K^2$$

giving $\langle f, \psi \rangle \leq c_T(\gamma\epsilon + \gamma^{-1} c_H c_\Lambda K^2)$. Optimizing this for γ yields $\langle f, \psi \rangle \leq c_T \sqrt{c_H c_\Lambda}\, \epsilon^{1/2} K$. By flipping the $-$ sign in (28) to $+$, a similar lower bound can be derived for $-\langle f, \psi \rangle$. ■

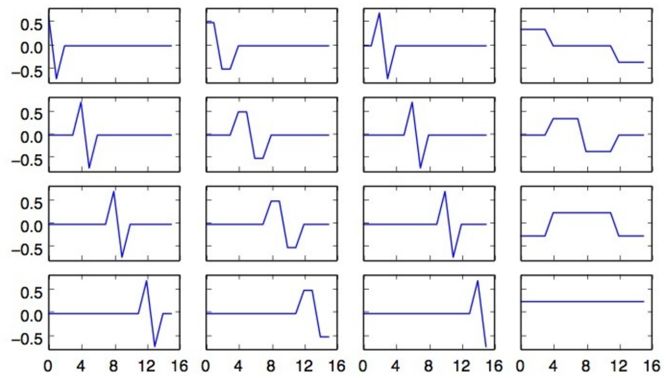

Figure 6. The MMF wavelets on a cycle graph on 16 vertices recover the Haar wavelet system.

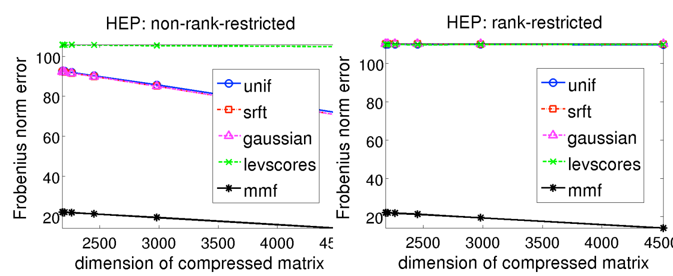

Figure 7. Comparison of the Frobenius norm error of the binary parallel MMF and Nyström approximations on the HEP dataset in the non-rank-restricted case and the rank-restricted case with $r = 60$.