FLATIRON INSTITUTE
Center for Computational Mathematics

Fast algorithms for hierarchical matrices

Machine Learning Seminar

May 2, 2019

What is fast?

Suppose $A \in \mathbb{C}^{N \times N}$, and, $v \in \mathbb{C}^N$

- Matvec $A \cdot v :\quad O(N^2)$
- Inversion $A^{-1} :\quad O(N^3)$
- Determinants $\det A :\quad O(N^3)$

For a given task, an algorithm is fast if it’s runtime beats the asymptotic complexity

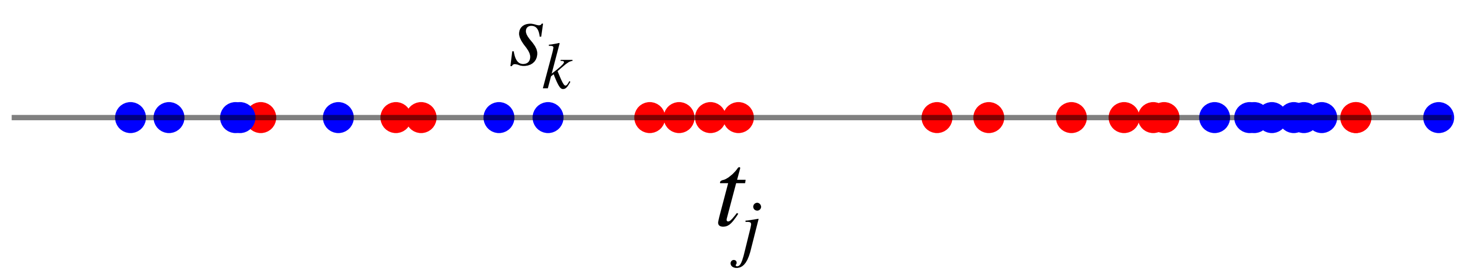
The dream: $O(N \log^s N)$

Examples

- Sparse matrices, matvecs in O(kN), if well-conditioned, inverse in O(kN)
- FFT matrices, matvecs in O(N log N), inverse analytically known, and inverse application in O(N log N)
- FMM matrices, matvecs in O(N)

Dense matrices $\neq$ Data dense

$$A_{j,k} = \delta_{j,k} + \cos(t_j - s_k)$$
$$= \delta_{j,k} + \cos(t_j)\cos(s_k) + \sin(t_j)\sin(s_k)$$



- **Matvec** $\boldsymbol{b} = A \cdot \boldsymbol{v}$: $O(N^2)$

Step 1:

$$W_1 = \sum_{k=1}^{N} \cos(s_k) v_k \,, \quad W_2 = \sum_{k=1}^{N} \sin(s_k) v_k$$

Step 2:

$$b_j = v_j + \cos(t_j) W_1 + \sin(t_j) W_2$$

O(N)!

$$A = I + UV^T \,, \quad U = \begin{bmatrix} \cos(t_1) & \sin(t_1) \\ \cos(t_2) & \sin(t_2) \\ \vdots & \vdots \\ \cos(t_N) & \sin(t_N) \end{bmatrix} , \quad V = \begin{bmatrix} \cos(s_1) & \sin(s_1) \\ \cos(s_2) & \sin(s_2) \\ \vdots & \vdots \\ \cos(s_N) & \sin(s_N) \end{bmatrix}$$

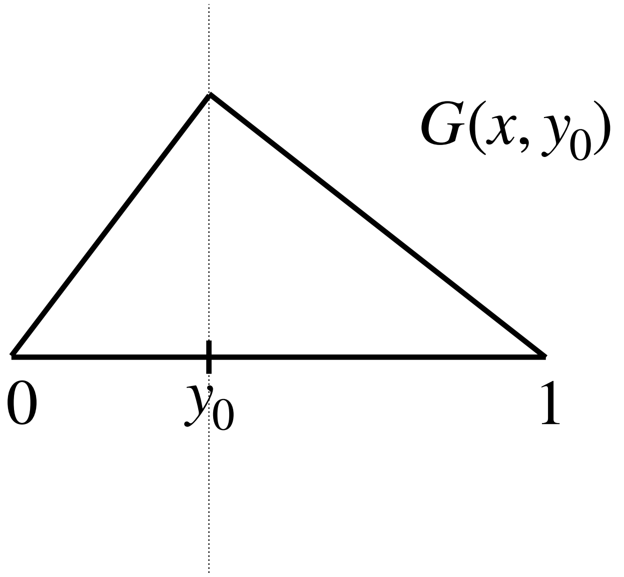
- **Inversion** A^{-1} : $O(N^3)$

  Sherman Morrison Woodbury formula: $A^{-1} = I - U(I_2 + V^T U)^{-1} V^T$ O(N)!

- **Determinants** $\det A$: $O(N^3)$

  Slyvester formula formula: $\det A = \det(I_2 + V^T U)$ O(N)!

Two-point boundary value problems



Given $p(x), q(x), f(x), \alpha, \beta,$ **find** $u(x)$ **which satisfies**

$$u''(x) + p(x)u'(x) + q(x)u(x) = f(x) \quad 0 < x < 1 .$$
$$u(0) = \alpha , \quad u(1) = \beta .$$

$u \to u(x) - \alpha - (\beta - \alpha)x$ **satisfies**

$$u''(x) + p(x)u'(x) + q(x)u(x) = \tilde{f}(x) \quad 0 < x < 1 .$$
$$u(0) = 0 , \quad u(1) = 0 .$$

$G(x, y)$: **Green's function for**

$$u''(x) = \delta_y \quad 0 < x < 1 .$$
$$u(0) = 0 , \quad u(1) = 0 .$$

$$G(x, y) = \begin{cases} x(1 - y) & 0 \leq x < y \leq 1 \\ y(1 - x) & 0 \leq y < x < \leq 1 \end{cases}$$

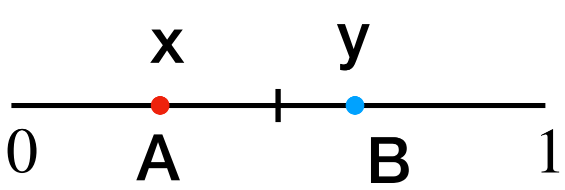
Integral formulation: $u(x) = \int_0^1 G(x, y)\sigma(y) ,$ σ **unknown density**

- **Boundary conditions** ✓
- **ODE yields following integral equation**

$$\sigma(x) + p(x)\int_0^1 \frac{\partial G}{\partial x}(x, y)\sigma(y)\, dy + q(x)\int_0^1 G(x, y)\sigma(y)\, dy = \tilde{f}, \quad 0 < x < 1$$

Structure of off-diagonal blocks

$$\sigma(x) + p(x)\int_0^1 \frac{\partial G}{\partial x}(x,y)\sigma(y)\,dy + q(x)\int_0^1 G(x,y)\sigma(y)\,dy = P\sigma = \tilde{f}$$



$$\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \end{bmatrix}$$

$$G(x,y) = \begin{cases} x(1-y) & 0 \le x < y \le 1 \\ y(1-x) & 0 \le y < x < \le 1 \end{cases}$$

$$P_{AB}\sigma_B = p(x)\int_B (1-y)\sigma(y)\,dy + q(x)\cdot x\int_B (1-y)\sigma(y)\,dy$$

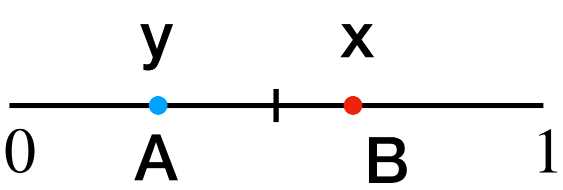
$$P_{AB} = \big(p(x) + q(x)\cdot x\big)\int_B (1-y)\cdot * \,dy$$

Rank 1

u_A $\qquad$ v_B^T

Structure of off-diagonal blocks

$$\sigma(x) + p(x)\int_0^1 \frac{\partial G}{\partial x}(x,y)\sigma(y)\,dy + q(x)\int_0^1 G(x,y)\sigma(y)\,dy = P\sigma = \tilde{f}$$



$$\begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \end{bmatrix}$$

$$G(x,y) = \begin{cases} x(1-y) & 0 \le x < y \le 1 \\ y(1-x) & 0 \le y < x < \le 1 \end{cases}$$

$$P_{AB}\sigma_B = p(x)\int_B (1-y)\sigma(y)\,dy + q(x)\cdot x\int_B (1-y)\sigma(y)\,dy$$

$$P_{AB} = \big(p(x) + q(x)\cdot x\big)\int_B (1-y)\cdot * \,dy$$

Rank 1

u_A $\qquad$ v_B^T

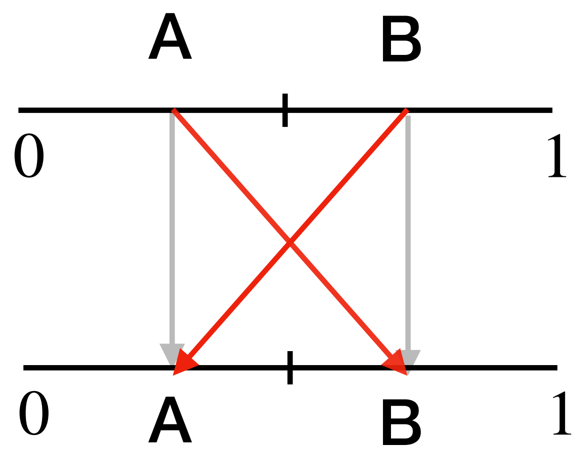
$$P_{BA}\sigma_A = -p(x)\int_A y\sigma(y)\,dy + q(x)\cdot(1-x)\int_A y\sigma(y)\,dy$$

$$P_{AB} = \big(-p(x) + q(x)\cdot(1-x)\big)\int_A y\cdot * \,dy$$

Rank 1

u_B $\qquad$ v_A^T

Faster? Matvec and inverse



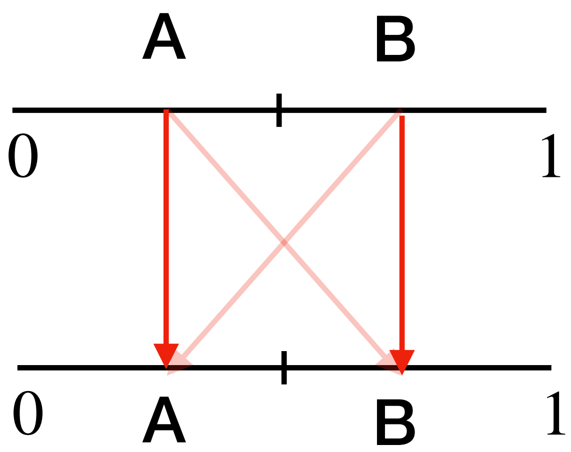
$$\begin{bmatrix} P_{AA} & u_A v_B^T \\ u_B v_A^T & P_{BB} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \end{bmatrix}$$

Matvec

Step 1:

Compute $u_B v_A^T \sigma_A, u_A v_B^T \sigma_B$ $\quad O(N)$

Faster? Matvec and inverse



$$\begin{bmatrix} P_{AA} & u_A v_B^T \\ u_B v_A^T & P_{BB} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \end{bmatrix}$$

Matvec

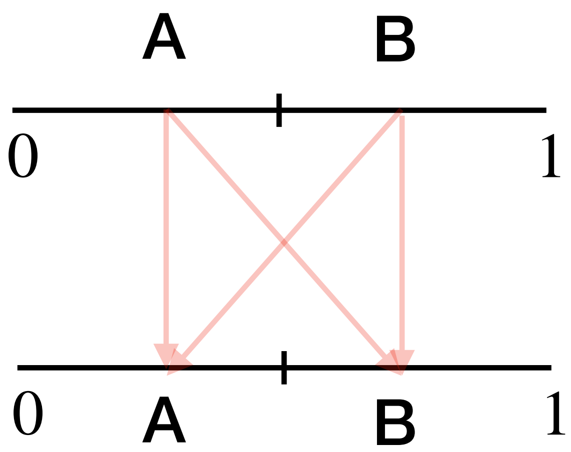
Step 1:

Compute $u_B v_A^T \sigma_A \, , u_A v_B^T \sigma_B$ $O(N)$

Step 2:

Compute $P_{AA}\sigma_A \, , P_{BB}\sigma_B$ $O(N^2/2)$

Faster? Matvec and inverse



$$\begin{bmatrix} P_{AA} & u_A v_B^T \\ u_B v_A^T & P_{BB} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \end{bmatrix}$$

Matvec

Step 1:

Compute $u_B v_A^T \sigma_A \, , u_A v_B^T \sigma_B$ $\quad O(N)$

Step 2:

Compute $P_{AA}\sigma_A \, , P_{BB}\sigma_B$ $\quad O(N^2/2)$

Inverse

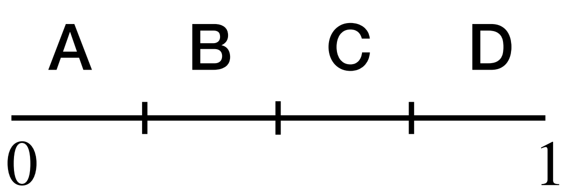
$$\begin{bmatrix} P_{AA} & u_A v_B^T \\ u_B v_A^T & P_{BB} \end{bmatrix} = \underbrace{\begin{bmatrix} P_{AA} & 0 \\ 0 & P_{BB} \end{bmatrix}}_{D} + \underbrace{\underbrace{\begin{bmatrix} u_A & 0 \\ 0 & u_B \end{bmatrix}}_{N \times 2} \underbrace{\begin{bmatrix} v_A^T & 0 \\ 0 & v_B^T \end{bmatrix}}_{2 \times N}}_{UV^T}$$

$$P^{-1} = \begin{bmatrix} P_{AA}^{-1} & 0 \\ 0 & P_{BB}^{-1} \end{bmatrix} - \begin{bmatrix} P_{AA}^{-1} u_A & 0 \\ 0 & P_{BB}^{-1} u_B \end{bmatrix} \left(I_2 + \begin{bmatrix} v_A^T P_{AA}^{-1} u_A & 0 \\ 0 & v_B^T P_{BB}^{-1} u_B \end{bmatrix} \right)^{-1} \begin{bmatrix} v_A^T P_{AA}^{-1} & 0 \\ 0 & v_B^T P_{BB}^{-1} \end{bmatrix} \quad O(N^3/4)$$

$$D^{-1} - D^{-1} U (I + V^T D^{-1} U)^{-1} V^T D^{-1}$$

Structure of off-diagonal blocks

$$\begin{bmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} \\ P_{BA} & P_{BB} & P_{BC} & P_{BD} \\ P_{CA} & P_{CB} & P_{CC} & P_{CD} \\ P_{DA} & P_{DB} & P_{DC} & P_{DD} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \\ \sigma_C \\ \sigma_D \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \\ \tilde{f}_C \\ \tilde{f}_D \end{bmatrix}$$



All off-diagonal blocks are rank 1

$$\begin{bmatrix} P_{AA} & u_{A,R}v_{B,L}^T & u_{A,R}v_{C,L}^T & u_{A,R}v_{D,L}^T \\ u_{B,L}v_{A,R}^T & P_{BB} & u_{B,R}v_{C,L}^T & u_{B,R}v_{D,L}^T \\ u_{C,L}v_{A,R}^T & u_{C,L}v_{B,R}^T & P_{CC} & u_{C,R}v_{D,L}^T \\ u_{D,L}v_{A,R}^T & u_{D,L}v_{B,R}^T & u_{D,L}v_{C,R}^T & P_{DD} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \\ \sigma_C \\ \sigma_D \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \\ \tilde{f}_C \\ \tilde{f}_D \end{bmatrix}$$

$$u_{I,L} = -p(x) + (1-x)\cdot q(x)\,, \quad x \in I$$

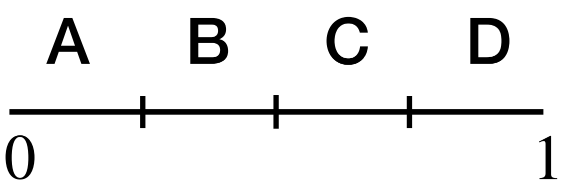
$$v_{I,L} = 1 - y\,, \quad y \in I$$

$$u_{I,R} = p(x) + x \cdot q(x)\,, \quad x \in I$$

$$v_{I,R} = y\,, \quad y \in I$$

Block separable form

$$\begin{bmatrix} P_{AA} & u_{A,R}v_{B,L}^T & u_{A,R}v_{C,L}^T & u_{A,R}v_{D,L}^T \\ u_{B,L}v_{A,R}^T & P_{BB} & u_{B,R}v_{C,L}^T & u_{B,R}v_{D,L}^T \\ u_{C,L}v_{A,R}^T & u_{C,L}v_{B,R}^T & P_{CC} & u_{C,R}v_{D,L}^T \\ u_{D,L}v_{A,R}^T & u_{D,L}v_{B,R}^T & u_{D,L}v_{C,R}^T & P_{DD} \end{bmatrix}$$



$$\begin{bmatrix} P_{AA} & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_B^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_B \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & P_{BB} & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & P_{CC} & u_C \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_C^T & P_{DD} \end{bmatrix}$$

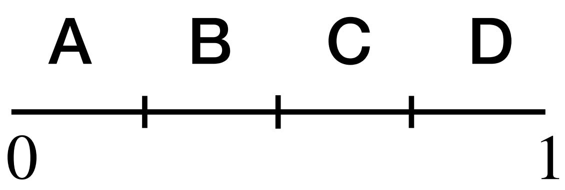
$$u_I = (u_{I,L} \quad u_{I,R})$$

$$v_I^T = \begin{pmatrix} v_{I,L}^T \\ v_{I,R}^T \end{pmatrix}$$

$$P_{i,j} = u_i S_{i,j} v_j^T$$

Block separable form

$$\begin{bmatrix} P_{AA} & u_{A,R}v^T_{B,L} & u_{A,R}v^T_{C,L} & u_{A,R}v^T_{D,L} \\ u_{B,L}v^T_{A,R} & P_{BB} & u_{B,R}v^T_{C,L} & u_{B,R}v^T_{D,L} \\ u_{C,L}v^T_{A,R} & u_{C,L}v^T_{B,R} & P_{CC} & u_{C,R}v^T_{D,L} \\ u_{D,L}v^T_{A,R} & u_{D,L}v^T_{B,R} & u_{D,L}v^T_{C,R} & P_{DD} \end{bmatrix}$$



$$\begin{bmatrix} P_{AA} & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_B^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_B \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & P_{BB} & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & P_{CC} & u_C \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_C^T & P_{DD} \end{bmatrix}$$

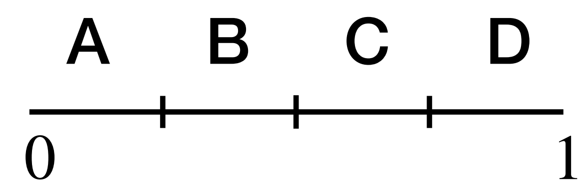
$$u_I = (u_{I,L} \quad u_{I,R})$$

$$v_I^T = \begin{pmatrix} v_{I,L}^T \\ v_{I,R}^T \end{pmatrix}$$

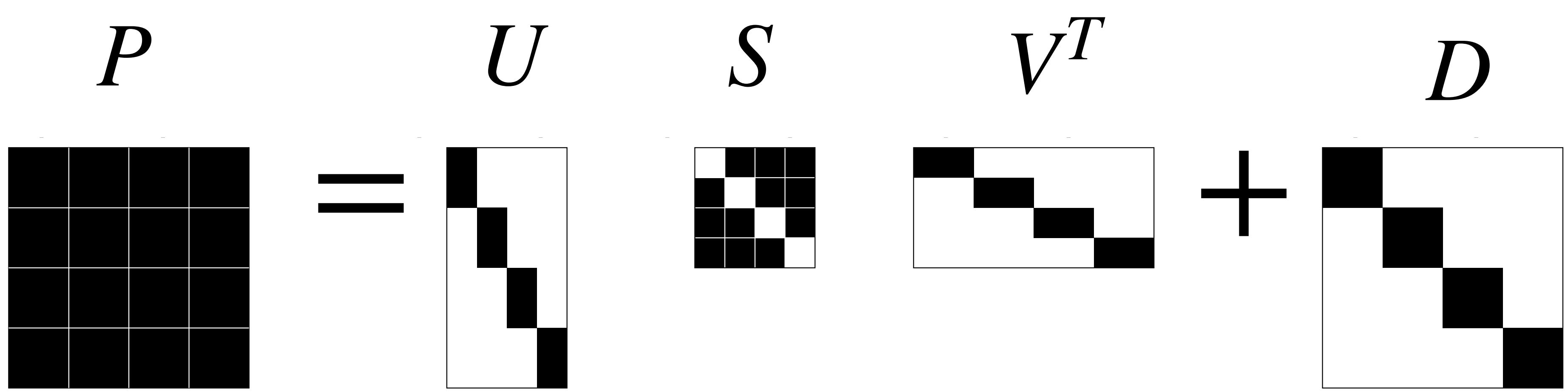
$$P_{i,j} = u_i S_{i,j} v_j^T$$

Block separable form

$$
\begin{bmatrix}
P_{AA} & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_B^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\
u_B \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & P_{BB} & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\
u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & P_{CC} & u_C \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\
u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_C^T & P_{DD}
\end{bmatrix}
$$

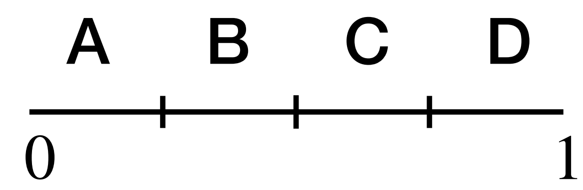


$$P_{i,j} = u_i S_{i,j} v_j^T$$



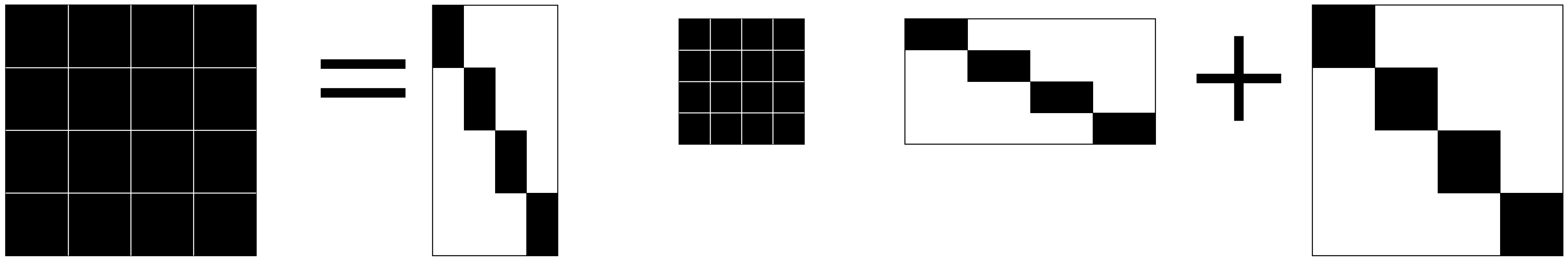
Block separable form

$$\begin{bmatrix} P_{AA} & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_B^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_B \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & P_{BB} & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & P_{CC} & u_C \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\ u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_C^T & P_{DD} \end{bmatrix}$$



$$P_{i,j} = u_i S_{i,j} v_j^T$$

$$P^{-1} = E \, (S + \hat{D})^{-1} \, F^T + G$$

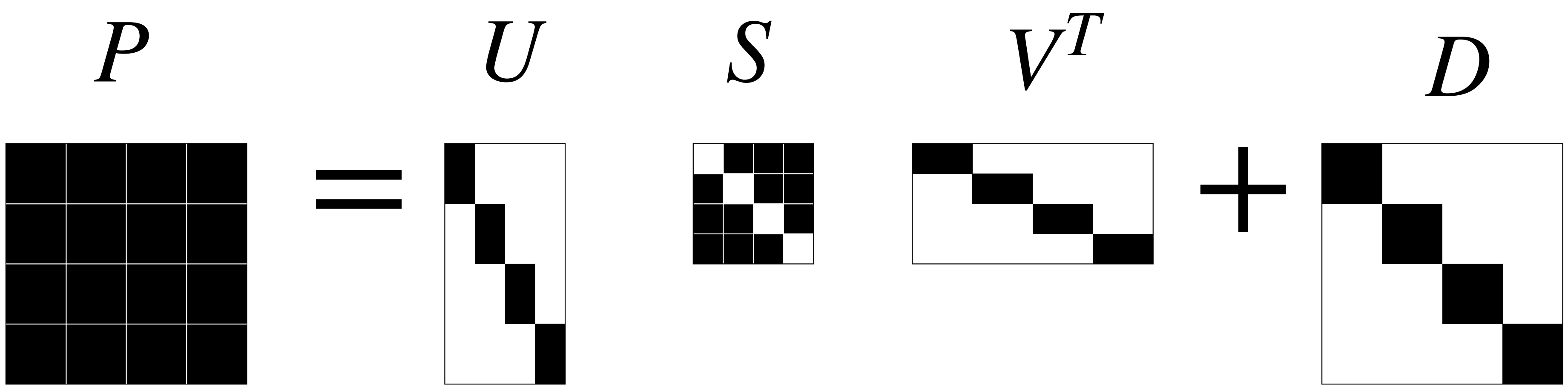


$$\hat{D} = (V^T D^{-1} U)^{-1}, \quad E = D^{-1} U \hat{D}, \quad F = (\hat{D} V^T D^{-1})^T, \quad G = D^{-1} - D^{-1} U \hat{D} V^T D^{-1}$$

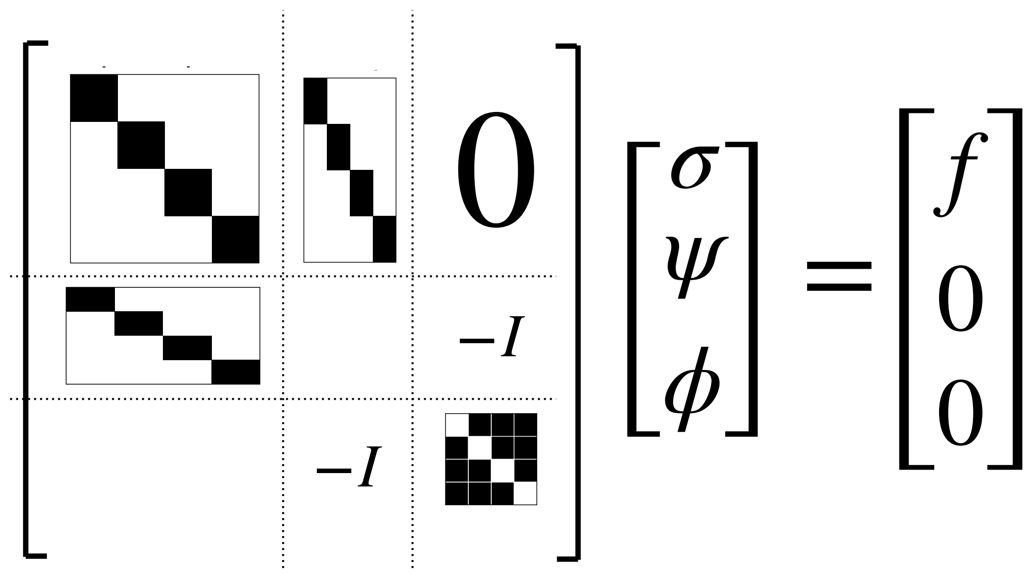
Recall inverse with two intervals: $P^{-1} = \begin{bmatrix} P_{AA}^{-1} & 0 \\ 0 & P_{BB}^{-1} \end{bmatrix} - \begin{bmatrix} P_{AA}^{-1} u_A & 0 \\ 0 & P_{BB}^{-1} u_B \end{bmatrix} \left(I_2 + \begin{bmatrix} v_A^T P_{AA}^{-1} u_A & 0 \\ 0 & v_B^T P_{BB}^{-1} u_B \end{bmatrix} \right)^{-1} \begin{bmatrix} v_A^T P_{AA}^{-1} & 0 \\ 0 & v_B^T P_{BB}^{-1} \end{bmatrix}$

Sparse matrix embedding of P

$$P\sigma = f$$



$$\phi = V^T\sigma, \quad \psi = S\phi$$



Matvec: $O(N^2/4)$

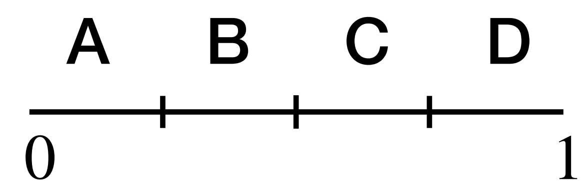
Inverse: $O(N^3/32)$

Not fast enough!

S admits a similar factorization

Structure of S

$$
\begin{bmatrix}
P_{AA} & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_B^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_A \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\
u_B \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & P_{BB} & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_C^T & u_B \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\
u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_C \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & P_{CC} & u_C \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} v_D^T \\
u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_A^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_B^T & u_D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} v_C^T & P_{DD}
\end{bmatrix}
$$

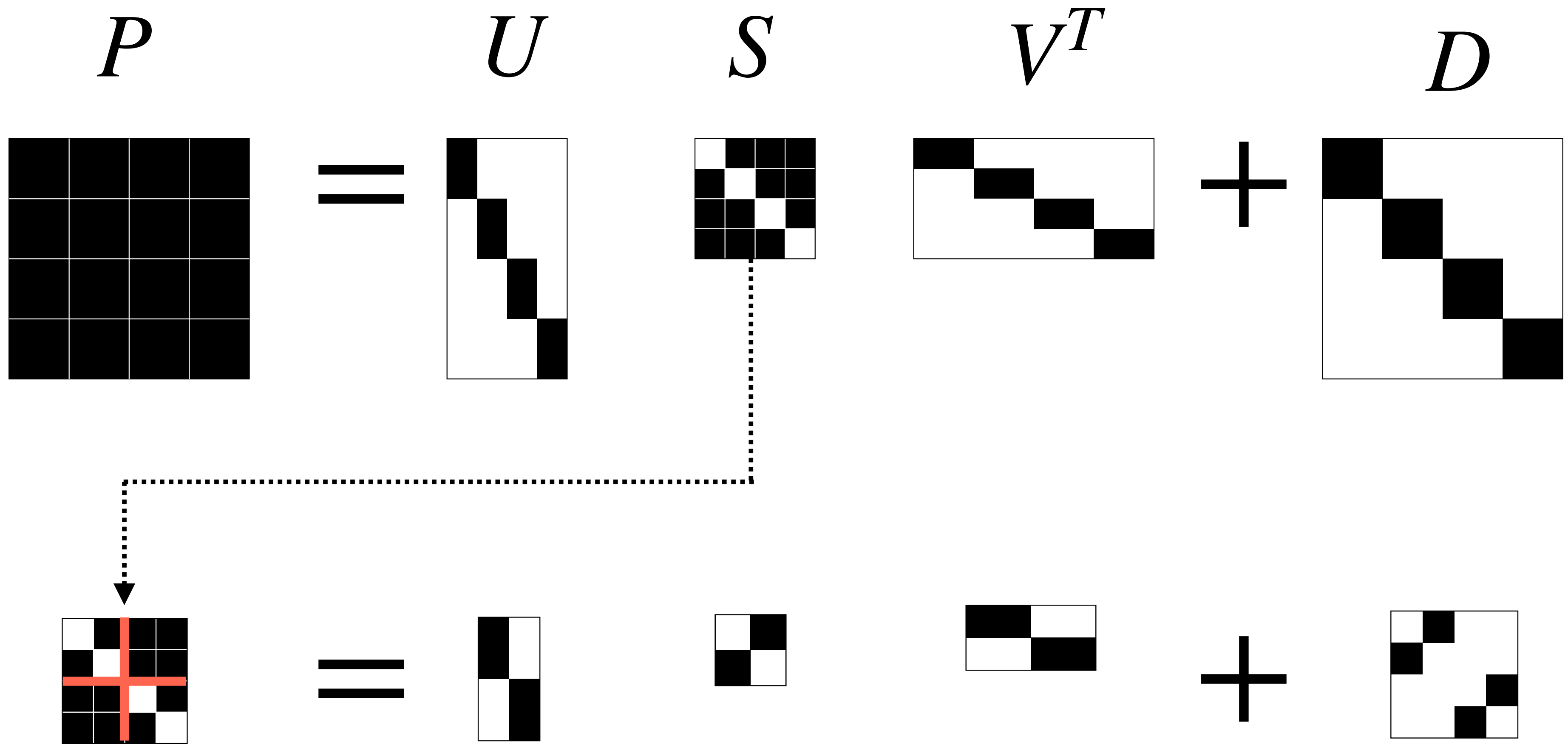


$$
P_{i,j} = u_i S_{i,j} v_j^T
$$

$$
S = \begin{bmatrix}
\begin{matrix} & \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} & \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \\
\begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} & \begin{matrix} & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \end{matrix}
\end{bmatrix}
= \begin{bmatrix}
\begin{matrix} & \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} & \end{matrix} & \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix} \\
\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 1 \end{pmatrix} & \begin{matrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \\ & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \end{matrix}
\end{bmatrix}
$$

Heirarchical block separable form for P

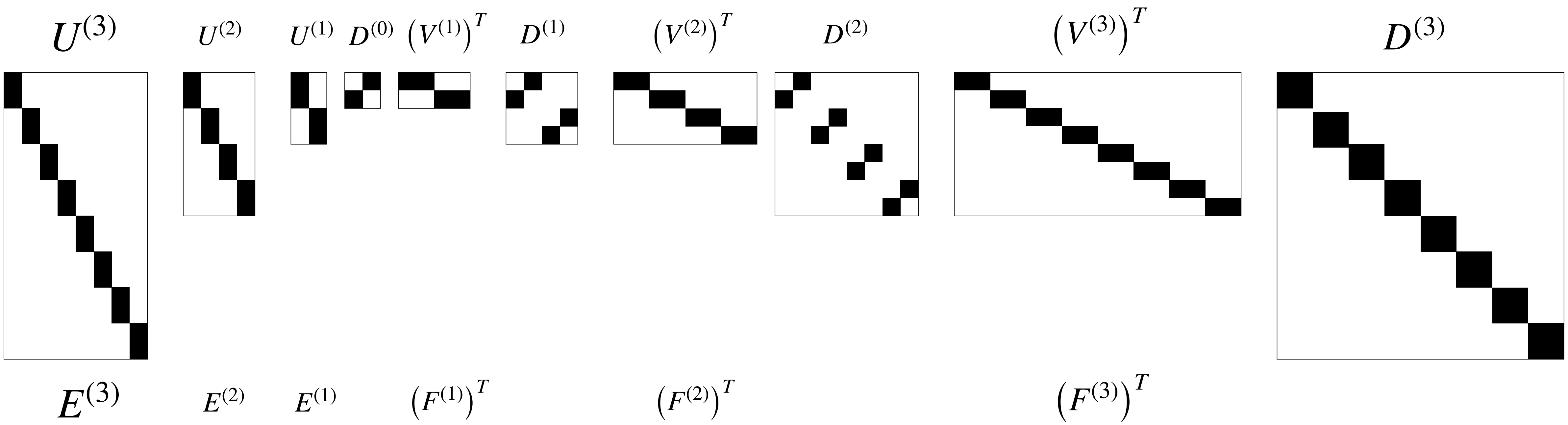
$$P = U^{(\ell)}\left(U^{(\ell-1)}\left(\ldots U^{(1)}D^{(0)}\left(V^{(1)}\right)^T + D^{(1)}\right)\ldots\left(V^{(\ell-1)}\right)^T + D^{(\ell-1)} + \right)\left(V^{(\ell)}\right)^T + D^{(\ell)}$$



Hierarchical block separable (HBS) form of P

$$P = U^{(3)}\left(U^{(2)}\left(U^{(1)}D^{(0)}\left(V^{(1)}\right)^T + D^{(1)}\right)\left(V^{(2)}\right)^T + D^{(2)}\right)\left(V^{(3)}\right)^T + D^{(3)}$$

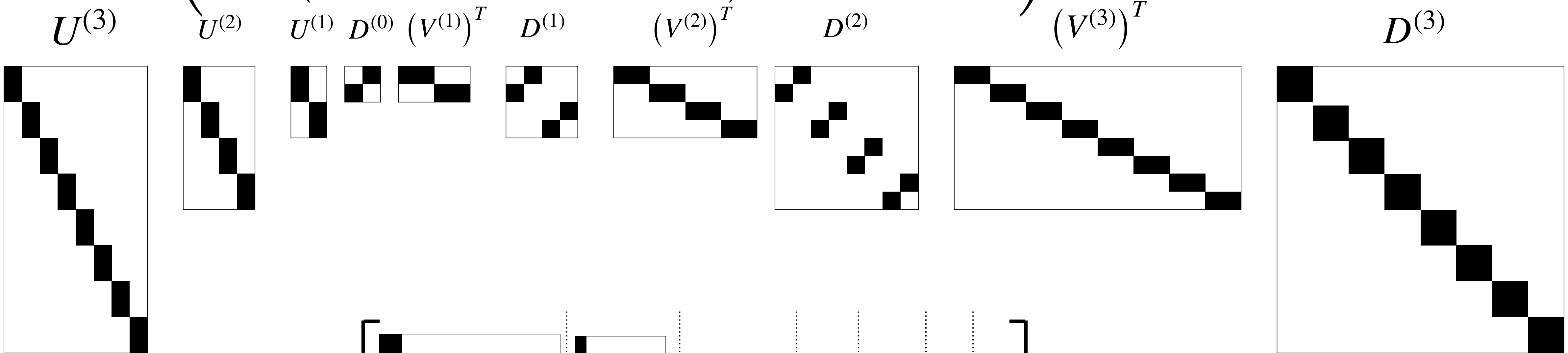
$$P^{-1} = E^{(3)}\left(E^{(2)}\left(E^{(1)}G^{(0)}\left(F^{(1)}\right)^T + G^{(1)}\right)\left(F^{(2)}\right)^T + G^{(2)}\right)\left(F^{(3)}\right)^T + G^{(3)}$$

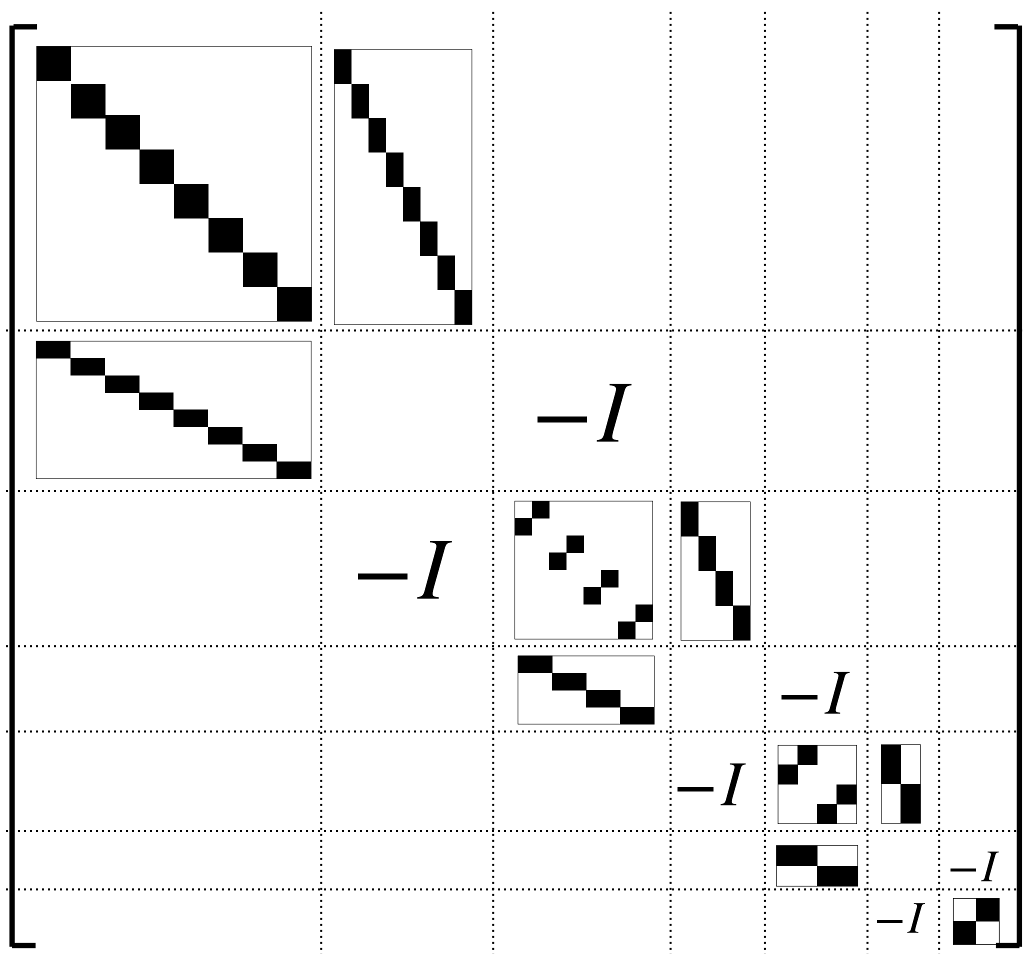


All matrices except $G^{(0)}$ are block diagonal

Sparse embedding of HBS form of P

$$P = U^{(3)}\left(U^{(2)}\left(U^{(1)}D^{(0)}\left(V^{(1)}\right)^T + D^{(1)}\right)\left(V^{(2)}\right)^T + D^{(2)}\right)\left(V^{(3)}\right)^T + D^{(3)}$$

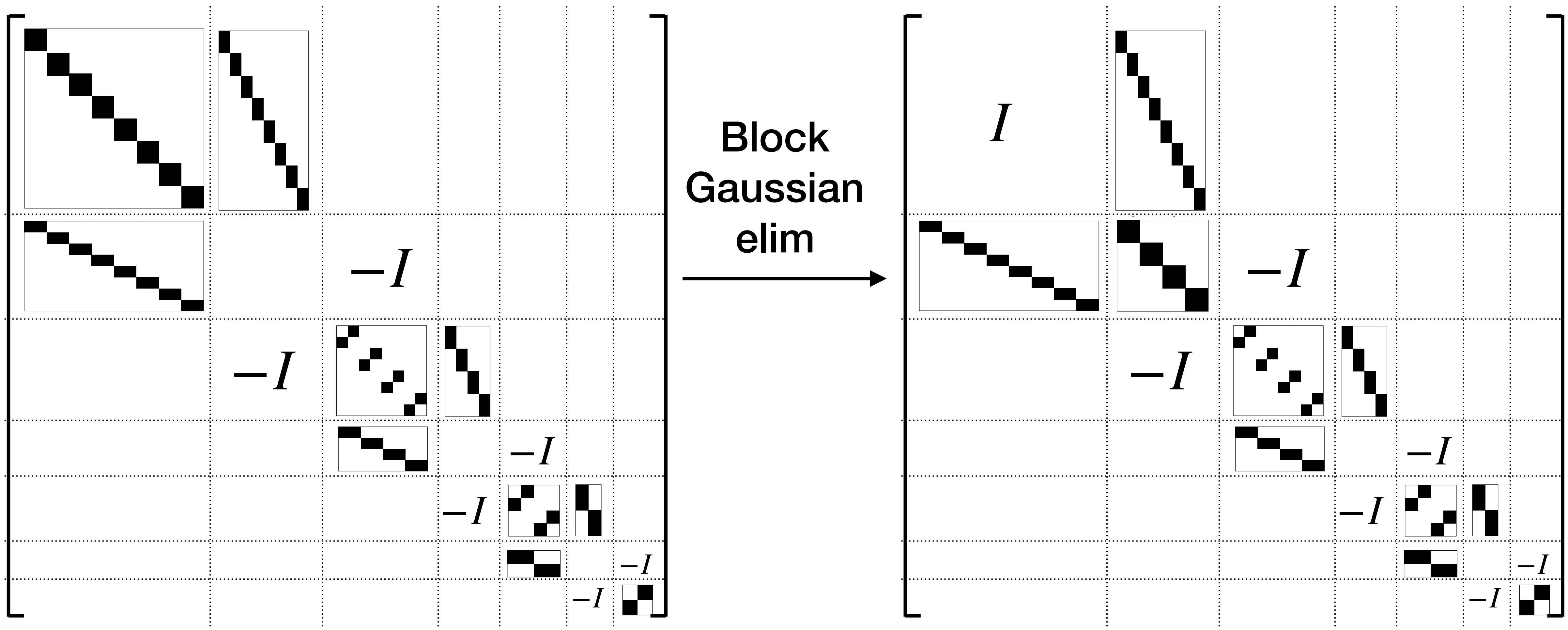




Matvec: $O(N)$

Sparse embedding of HBS form of P - inversion

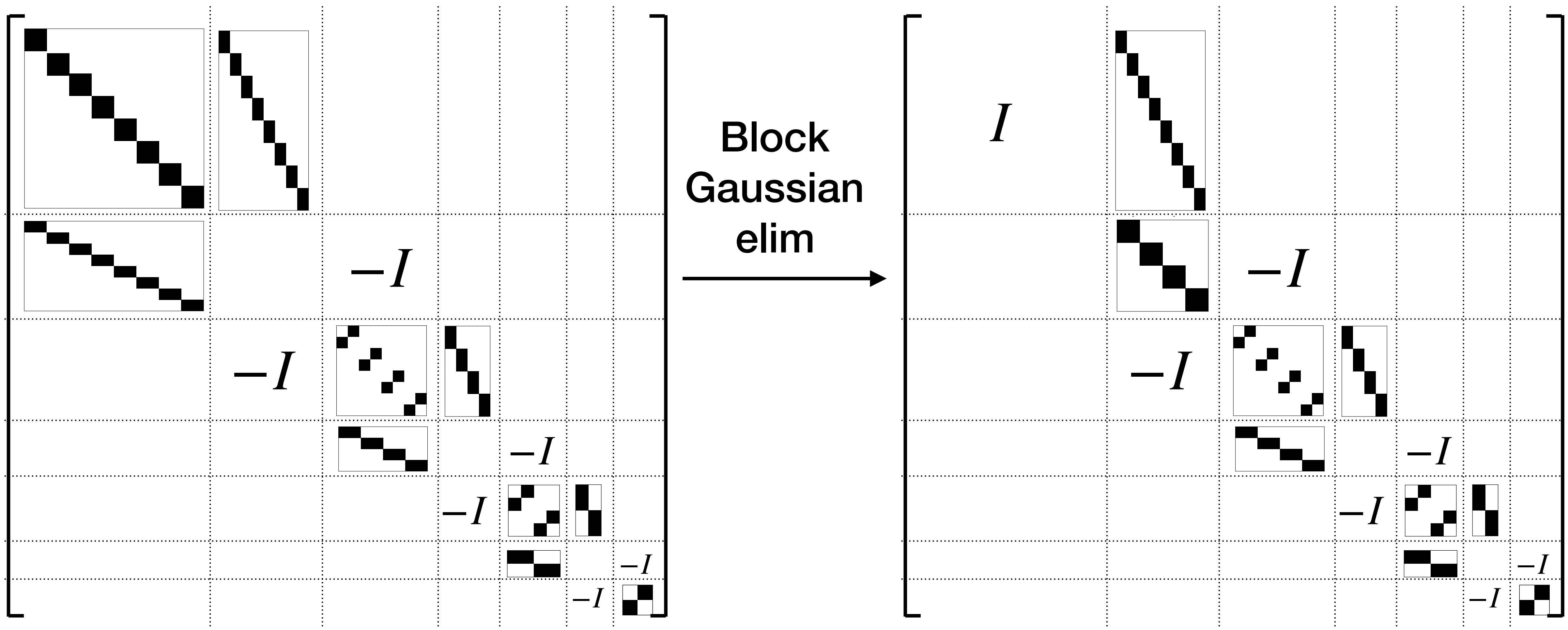
$$P^{-1} = E^{(3)}\left(E^{(2)}\left(E^{(1)}G^{(0)}\left(F^{(1)}\right)^T + G^{(1)}\right)\left(F^{(2)}\right)^T + G^{(2)}\right)\left(F^{(3)}\right)^T + G^{(3)}$$



Inversion: $O(N)$

Sparse embedding of HBS form of P - inversion

$$P^{-1} = E^{(3)}\left(E^{(2)}\left(E^{(1)}G^{(0)}\left(F^{(1)}\right)^T + G^{(1)}\right)\left(F^{(2)}\right)^T + G^{(2)}\right)\left(F^{(3)}\right)^T + G^{(3)}$$

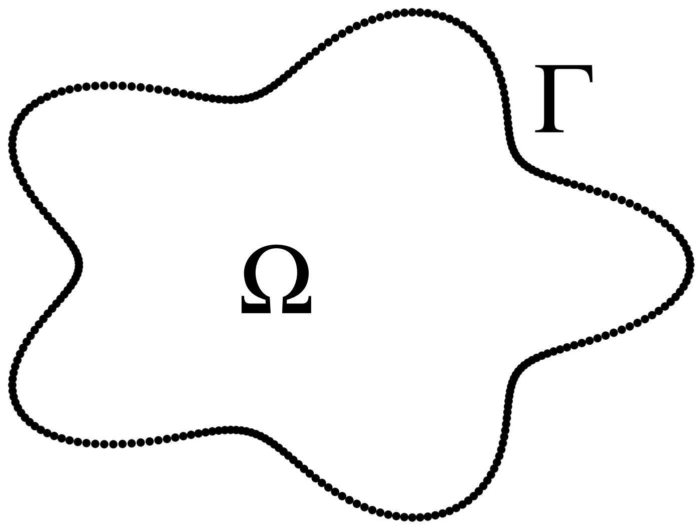


Inversion: $O(N)$

Higher dimensions?

Given $f(x)$, find $u(x)$ which satisfies

$$\Delta u(x) = 0 \quad x \in \Omega .$$
$$u(x) = f(x) , \quad x \in \Gamma .$$



$G(x,y)$: Green's function for

$$\Delta u(x) = \delta_y$$

$$G(x,y) = -\frac{1}{2\pi} \log |x-y|$$

Integral formulation: $u(x) = \int_\Gamma \frac{\partial}{\partial \nu} G(x,y)\sigma(y) \quad x \in \Omega ,$ σ unknown density

$$\int_\Gamma \frac{\partial}{\partial \nu} G(x,y)\sigma(y)\, dS_y = \text{Re} \int_\Gamma \frac{\sigma(\xi)}{z-\xi} d\xi \quad z = x_1 + ix_2 , \quad \xi = y_1 + iy_2$$

- PDE ✓
- Boundary conditions yields following integral equation

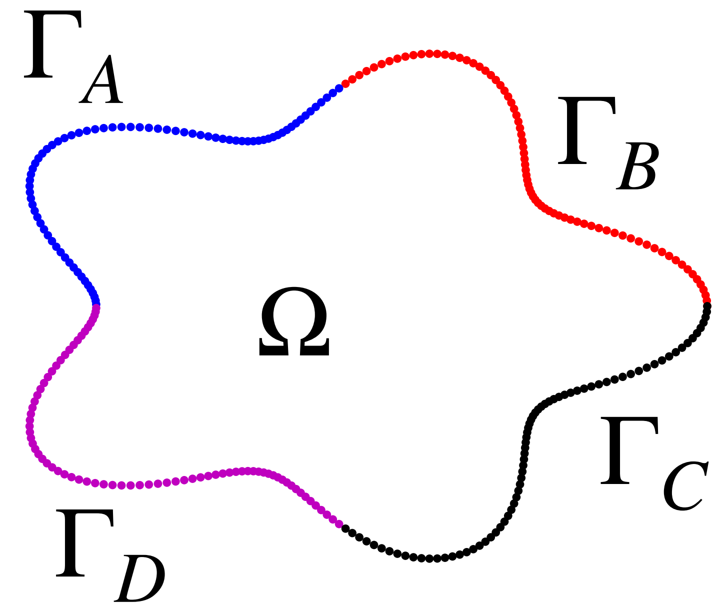
$$\sigma(x) + 2\int_\Gamma \frac{\partial}{\partial \nu} G(x,y)\sigma(y)\, dS_y = 2f(x)$$

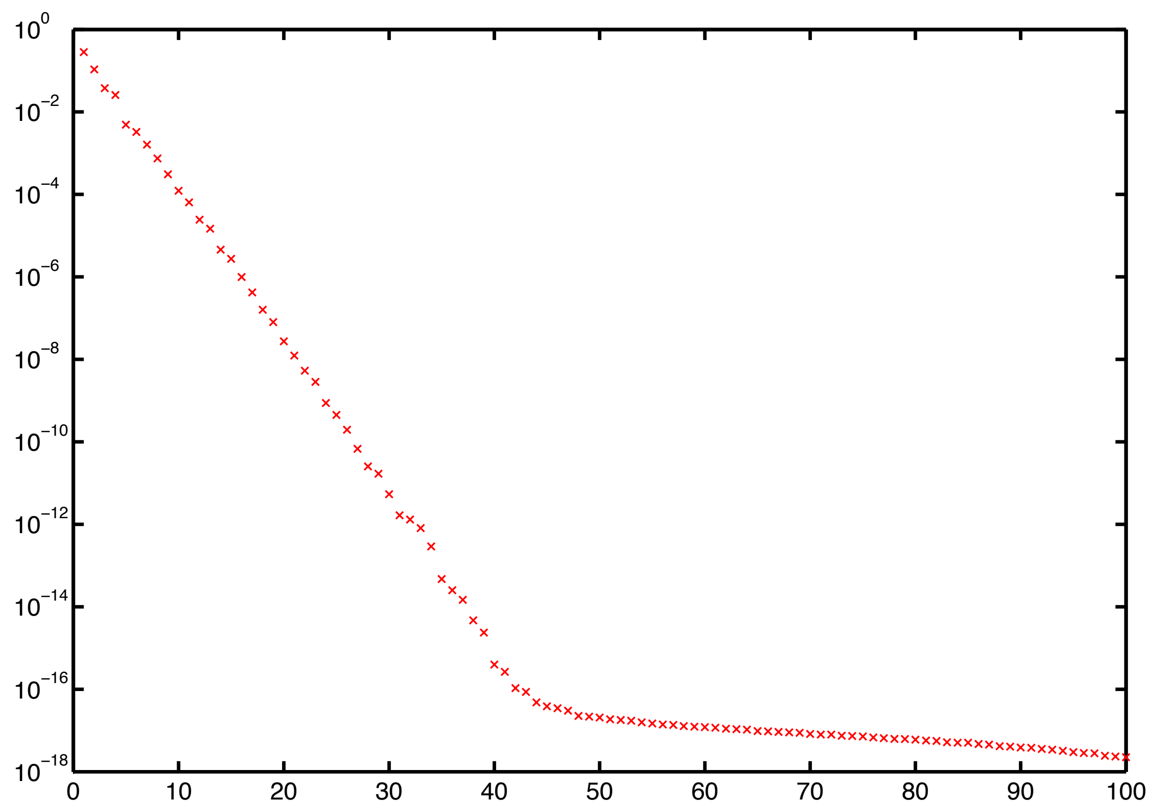
$$P\sigma = f$$

P is also HBS compressible!

Off-diagonal blocks

$$\begin{bmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} \\ {\color{red}P_{BA}} & P_{BB} & {\color{red}P_{BC}} & {\color{red}P_{BD}} \\ P_{CA} & P_{CB} & P_{CC} & P_{CD} \\ P_{DA} & P_{DB} & P_{DC} & P_{DD} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \\ \sigma_C \\ \sigma_D \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \\ \tilde{f}_C \\ \tilde{f}_D \end{bmatrix}$$





Singular values of matrix : $\begin{bmatrix} P_{BA} & 0 & P_{BC} & P_{BD} \end{bmatrix}$

$$P_{i,j} = u_i S_{i,j} v_j^T$$

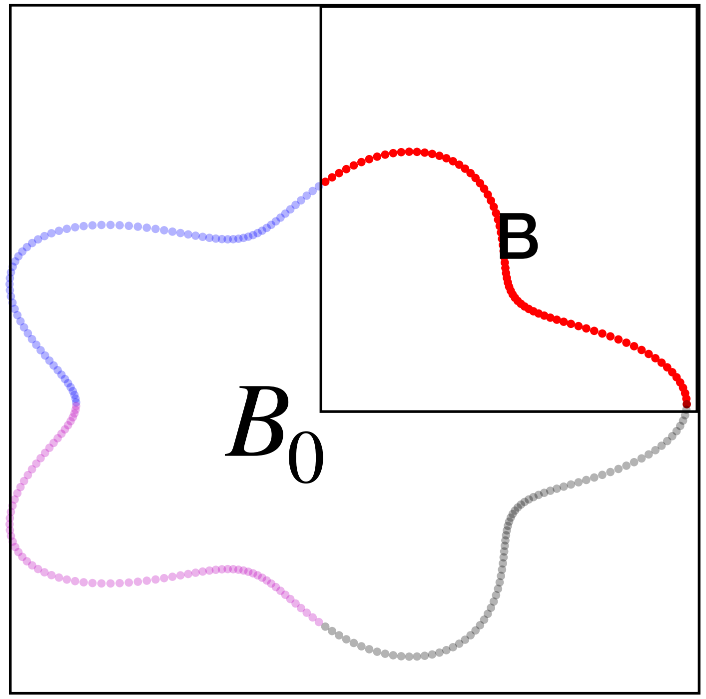
$u_i, v_j,$ **now rank-k matrices**

How to compute u_i, v_j?

Option 1: Use analytical FMM expansions — matrix no longer HBS then, but $\mathscr{H}^2$

Option 2: Use numerical compression techniques, like SVD, ID

Low-rank approximations - Functional SVDs



Suppose $P\sigma = \sum_j K(x_i, y_j)\sigma_j \ , \quad x_i \in B \quad y_j \in B_0 \backslash B$

$Tf = \int_B K(x,y)f(y)dy \ , \quad T : \mathbb{L}^2(B_0 \backslash B) \to \mathbb{L}^2(B)$ with

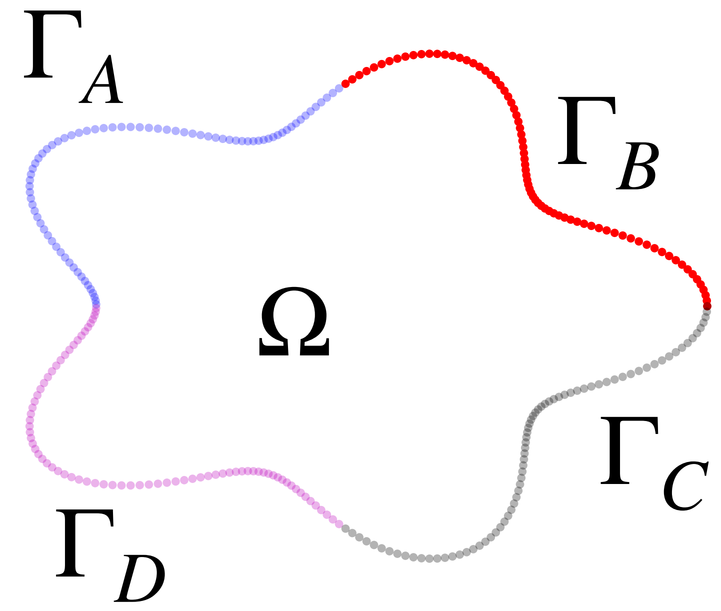
$$\int_{B_0 \backslash B} \int_B |K(x,y)|^2 \, dx \, dy < \infty$$

Then, $K(x,y) = \sum_{i=1}^{p} u_i(x) s_i v_i(y) + O(\varepsilon)$.

- Computing the functional SVD can be numerically intensive, particularly beyond d=2,3
- Costs can be amortized for translationally invariant kernels $K(|x-y|)$ and/or homogeneous kernels $K(\lambda x, \lambda y) = \lambda^r K(x,y)$
- Computational savings if kernel satisfies Green's identities (Proxy surfaces)
- FMM-like translation operators through SVDs for recompressing S

Low rank approximations - Randomized algorithms

$$\begin{bmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} \\ P_{BA} & P_{BB} & P_{BC} & P_{BD} \\ P_{CA} & P_{CB} & P_{CC} & P_{CD} \\ P_{DA} & P_{DB} & P_{DC} & P_{DD} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \\ \sigma_C \\ \sigma_D \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \\ \tilde{f}_C \\ \tilde{f}_D \end{bmatrix}$$



$$P_B$$

$$\begin{bmatrix} P_{AB} & 0 & P_{CB} & P_{DB} \end{bmatrix} = U_B \begin{bmatrix} \tilde{V}^T_{BA} & 0 & \tilde{V}^T_{BC} & \tilde{V}^T_{BD} \end{bmatrix}$$

$$p \times N \qquad k \times N \qquad k \times p$$

Randomized algorithms:

$W = \mathbb{R}^{N \times (k+r)}$, **random Gaussian matrix, FFT matrix**

$Y = P_B W$	$Y \in \mathbb{R}^{p \times (k+r)}$	Sample range of matrix
$Y = QR$		Orthogonalize sampled range
$T = Q^* P_B$	$T \in \mathbb{R}^{(k+r) \times p}$	Change of basis
$T = \hat{U} S V^T$		SVD of reduced matrix
$P_B \approx Q \hat{U} S V^T$		

Randomized algorithms - error analysis and performance

$$P_B$$

$$\begin{bmatrix} P_{AB} & 0 & P_{CB} & P_{DB} \end{bmatrix} = U_B \begin{bmatrix} \tilde{V}_{BA}^T & 0 & \tilde{V}_{BC}^T & \tilde{V}_{BD}^T \end{bmatrix}$$

$$p \times N \qquad k \times N \qquad k \times p$$

Randomized algorithms:

$W = \mathbb{R}^{N\times(k+r)}$, random Gaussian matrix, FFT matrix

$Y = P_B W$	$Y \in \mathbb{R}^{p\times(k+r)}$	Sample range of matrix	$O(N \cdot (k+r) \cdot p)$
$Y = QR$		Orthogonalize sampled range	$O(p \cdot (k+r)^2)$
$T = Q^* P_B$	$T \in \mathbb{R}^{(k+r)\times N}$	Change of basis	$O(N \cdot (k+r) \cdot p)$
$T = \hat{U} S V^T$		SVD of reduced matrix	$O((k+r)^2 \cdot N)$
$P_B \approx Q \hat{U} S V^T$			

$$\|P_B - Q\hat{U}SV^T\| = \|P_B - QT\| = \|P_B - QQ^*P_B\|$$

$\|P_B - QQ^*P_B\| \le \left(1 + C\sqrt{N}\right) s_{k+1}$ with probability $1 - 6r^{-r}$

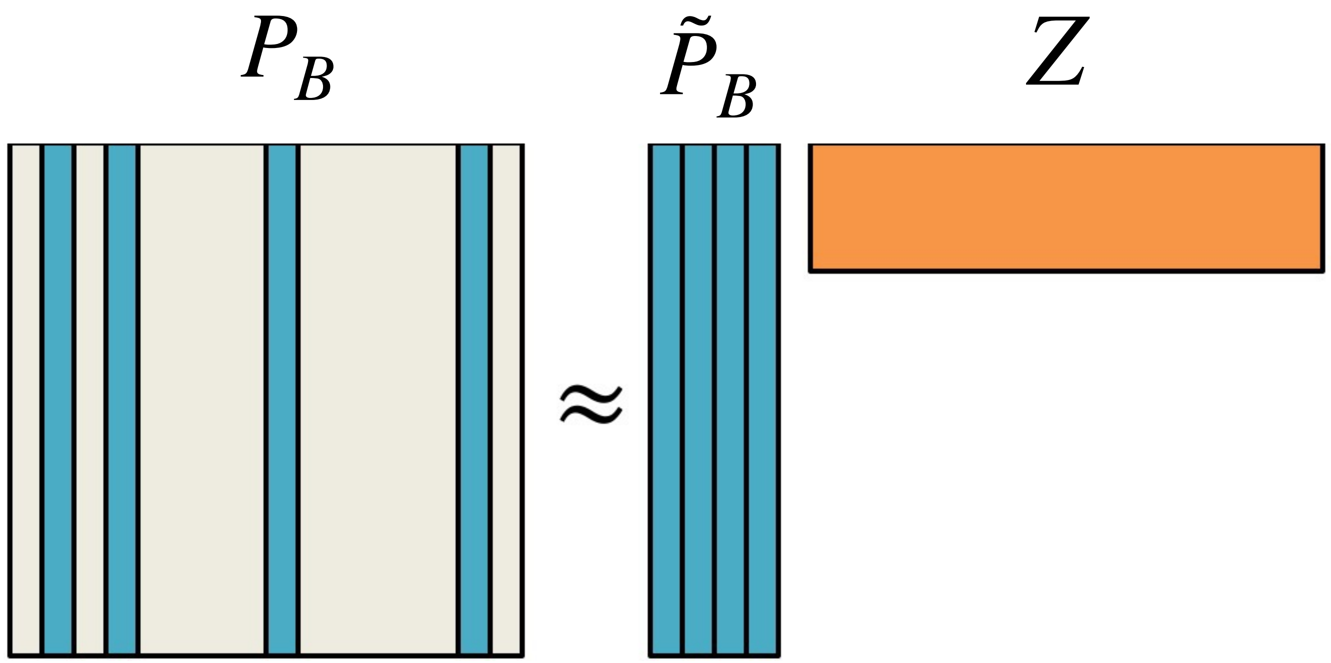
Issues:

Cost of compressing one block of rows: $O(N \cdot (k+r) \cdot p)$

N such factorizations needed $\implies$ cost of factorization: $O(N^2)$

Lack of interpretability of S at next layer

Interpolative Decomposition (ID)



$$\|P_B - \tilde{P}_B Z\| \leq (1 + \sqrt{k(n-k)})s_{k+1}$$

$$|Z_{i,j}| \leq 1$$

Combinatorial search, exponential cost

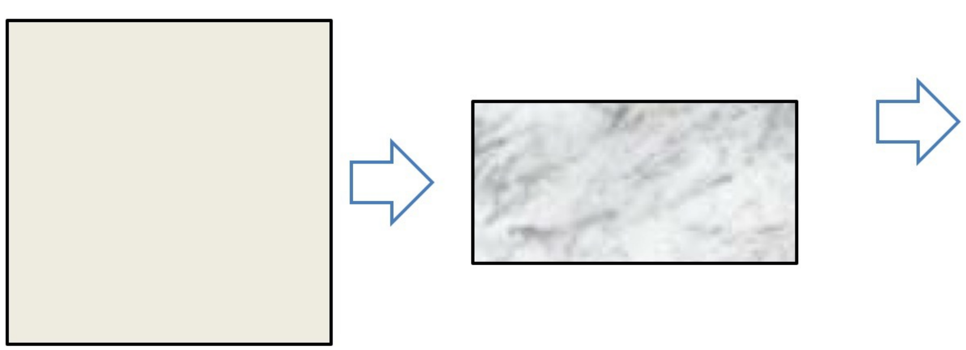
$$\|P_B - \tilde{P}_B Z\| \leq (1 + \sqrt{k(n-k)})s_{k+1}$$

$$|Z_{i,j}| \leq 2$$

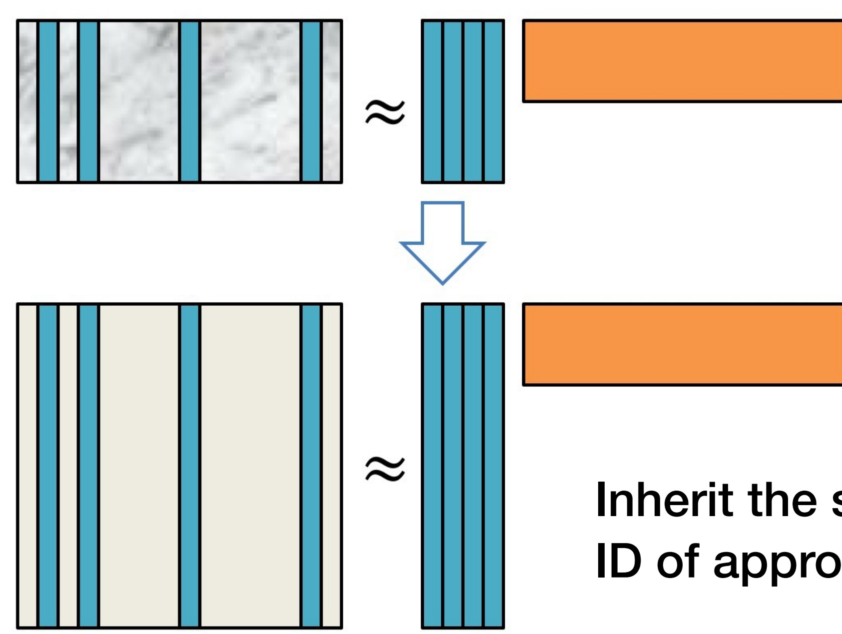
$$O(N \cdot p^2 \log N)$$

In practice, rank revealing QR works fine

Low rank approximation that uses columns of input matrix



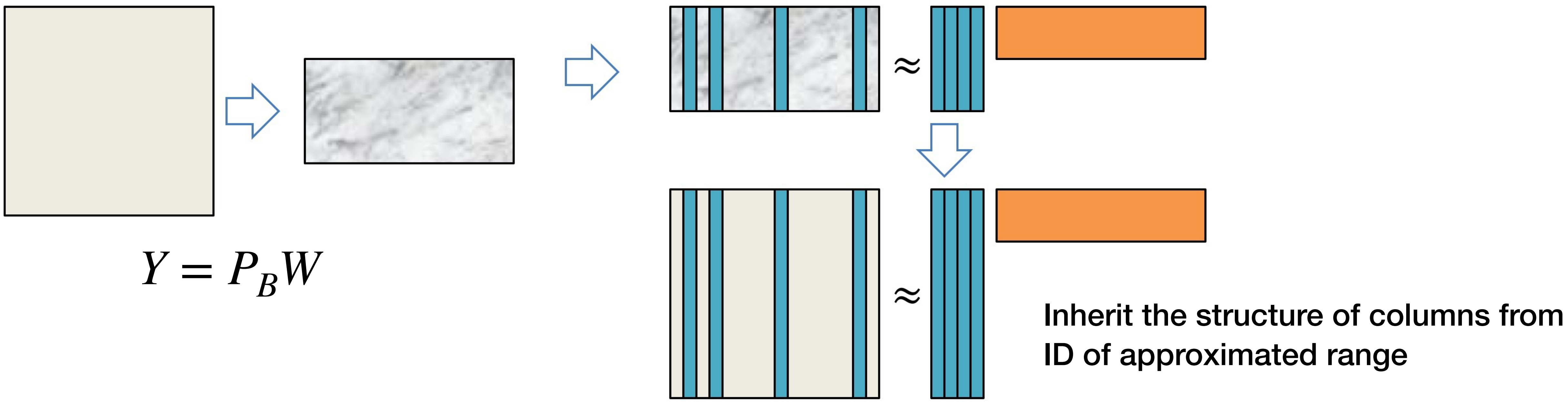
$$Y = P_B W$$



Inherit the structure of columns from ID of approximated range

Randomized approach for computing ID

Interpolative Decomposition (ID)



$Y = P_B W$

Randomized approach for computing ID

Issues:

Cost of compressing one block of rows: $O(N \cdot (k + r) \cdot p)$

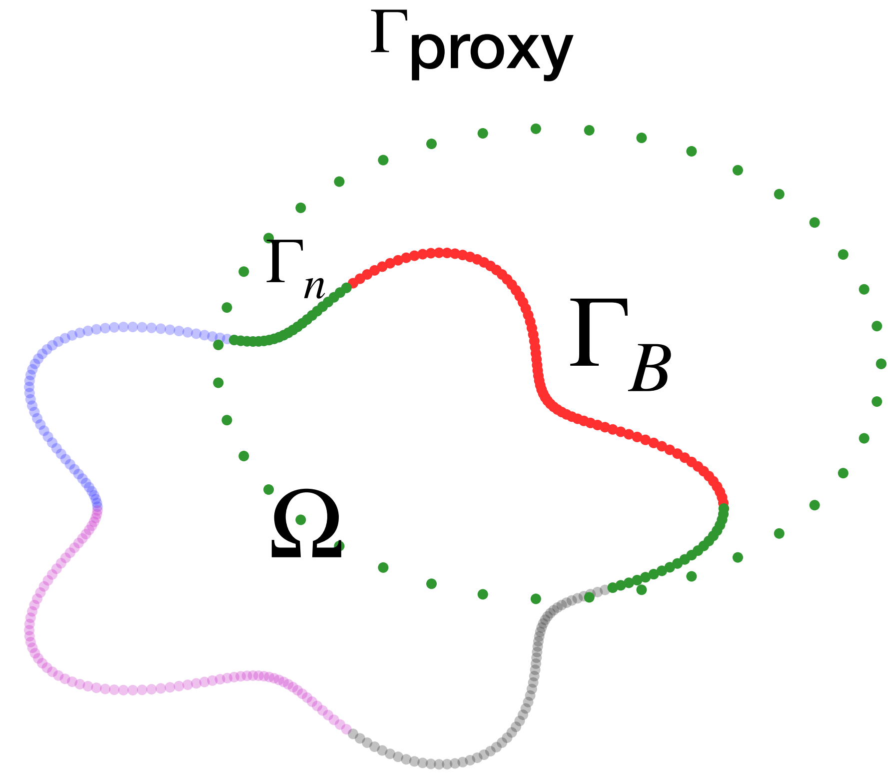
N such factorizations needed $\implies$ cost of factorization: $O(N^2)$

~~Lack of interpretability of S at next layer~~

Entries of S are sub-blocks of the original matrix

Low rank approximations - Proxy surfaces

$$\begin{bmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} \\ \color{red}{P_{BA}} & P_{BB} & \color{red}{P_{BC}} & \color{red}{P_{BD}} \\ P_{CA} & P_{CB} & P_{CC} & P_{CD} \\ P_{DA} & P_{DB} & P_{DC} & P_{DD} \end{bmatrix} \begin{bmatrix} \sigma_A \\ \sigma_B \\ \sigma_C \\ \sigma_D \end{bmatrix} = \begin{bmatrix} \tilde{f}_A \\ \tilde{f}_B \\ \tilde{f}_C \\ \tilde{f}_D \end{bmatrix}$$



Instead of compressing P_B, compress $\underbrace{\begin{bmatrix} P_{B,\Gamma_n} & P_{B,\Gamma_{\text{proxy}}} \end{bmatrix}}_{p \times (n_{\text{proxy}} + n_{\text{near}})}$

Works when matrix entries from Kernel satisfying Green's identity

General idea: identify smaller collection of columns which approximate bulk of matrix

Issues:

~~Cost of compressing one block of rows: $O(N \cdot (k+r) \cdot p)$~~

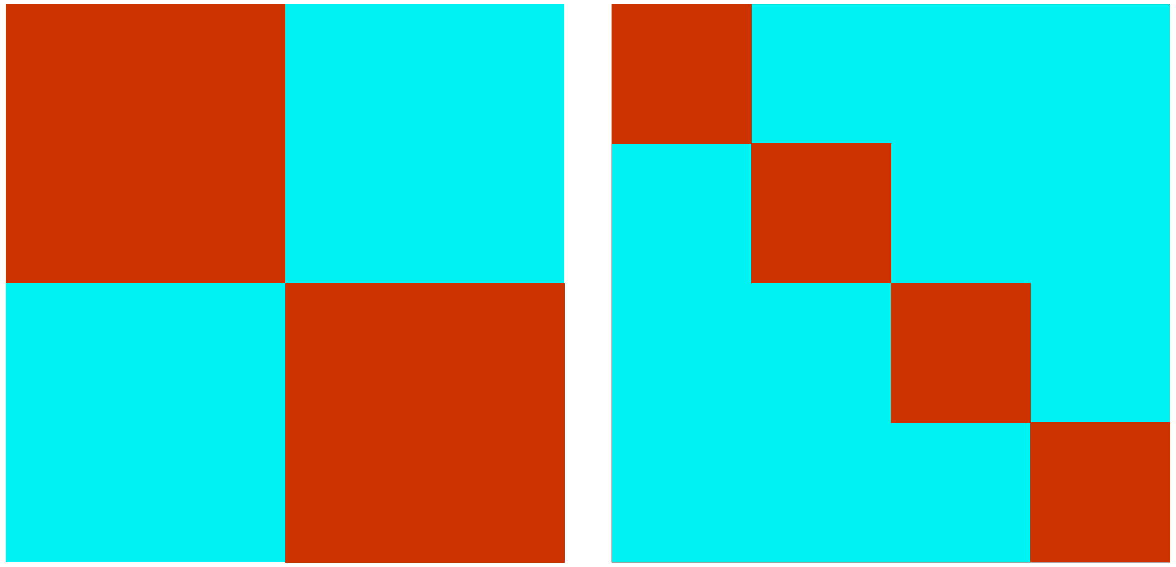
Cost of compressing one block of rows: $O((n_{\text{proxy}} + n_{\text{near}}) \cdot (k+r) \cdot p)$

~~N such factorizations needed $\Longrightarrow$ cost of factorization: $O(N^2)$~~

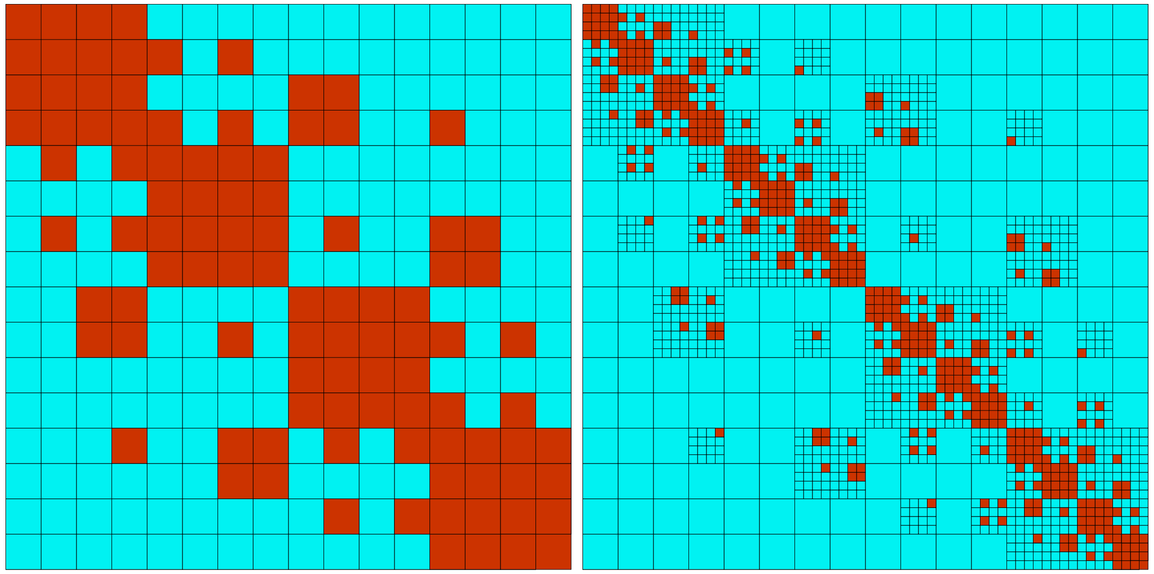
~~Lack of interpretability of S at next layer~~

Entries of S are sub-blocks of the original matrix

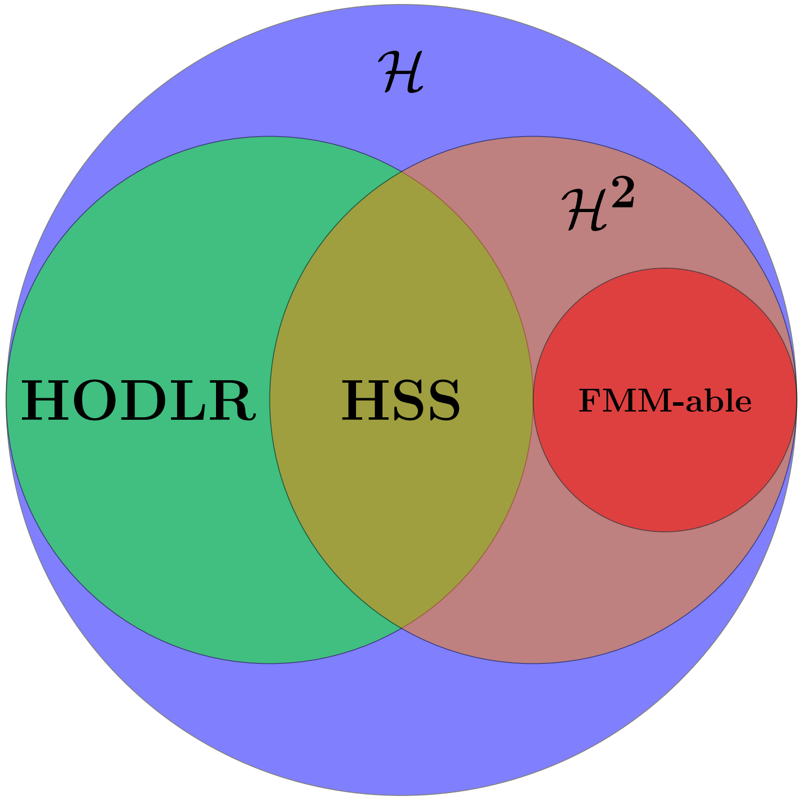
The zoo of matrix factorizations



HODLR/HSS matrices



FMM/$\mathscr{H}^2$ matrices



Butterfly/FFT matrices

Nested basis →

Low-rank structure ↓

Low-rank structure	No	Yes
Strong	HODLR	HSS
Weak	$\mathcal{H}$	$\mathcal{H}^2$

Hierarchical matrices in Neural Networks

A multiscale neural network based on hierarchical nested bases

Yuwei Fan[*] Jordi Feliu-Fabà[†] Lin Lin[‡] Lexing Ying[§] Leonardo Zepeda-Núñez[¶]

Using $\mathscr{H}^2$ in layers of locally connected networks

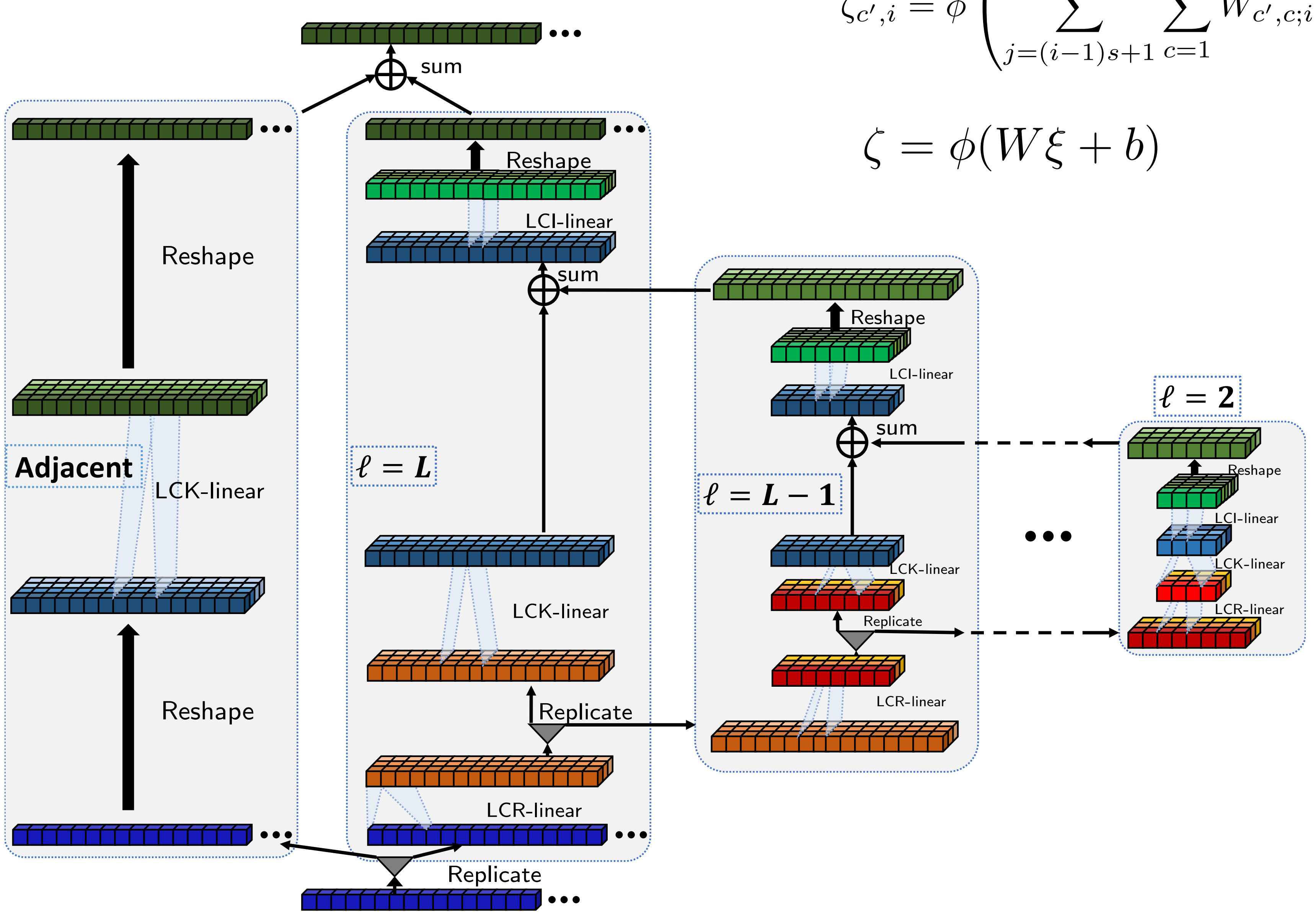
A multiscale neural network based on hierarchical matrices

Yuwei Fan[*] Lin Lin[†] Lexing Ying[‡] Leonardo Zepeda-Núñez[§]

Using $\mathscr{H}$ in layers of locally connected networks

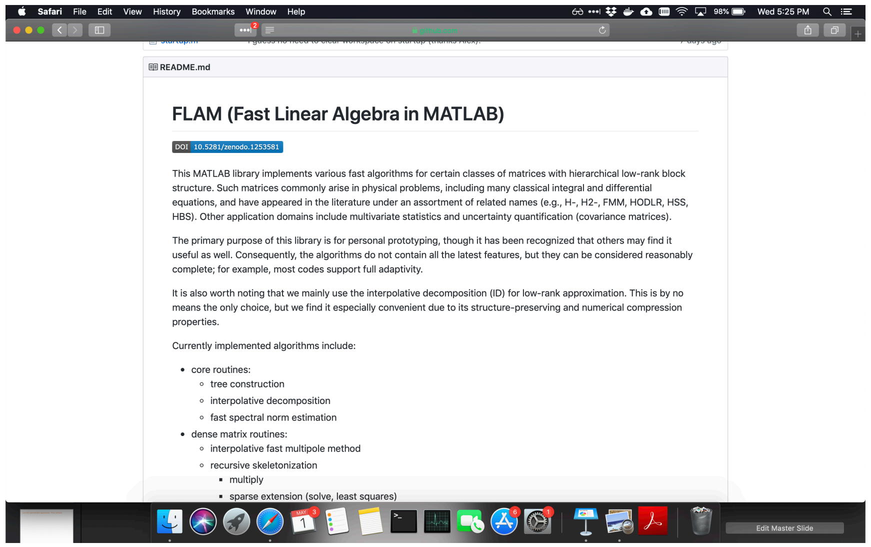
$$\zeta_{c',i} = \phi\left(\sum_{j=(i-1)s+1}^{(i-1)s+w} \sum_{c=1}^{\alpha} W_{c',c;i,j}\xi_{c,j} + b_{c',i}\right), \quad i = 1, \dots, N_x', \; c' = 1, \dots, \alpha'$$

$$\zeta = \phi(W\xi + b)$$

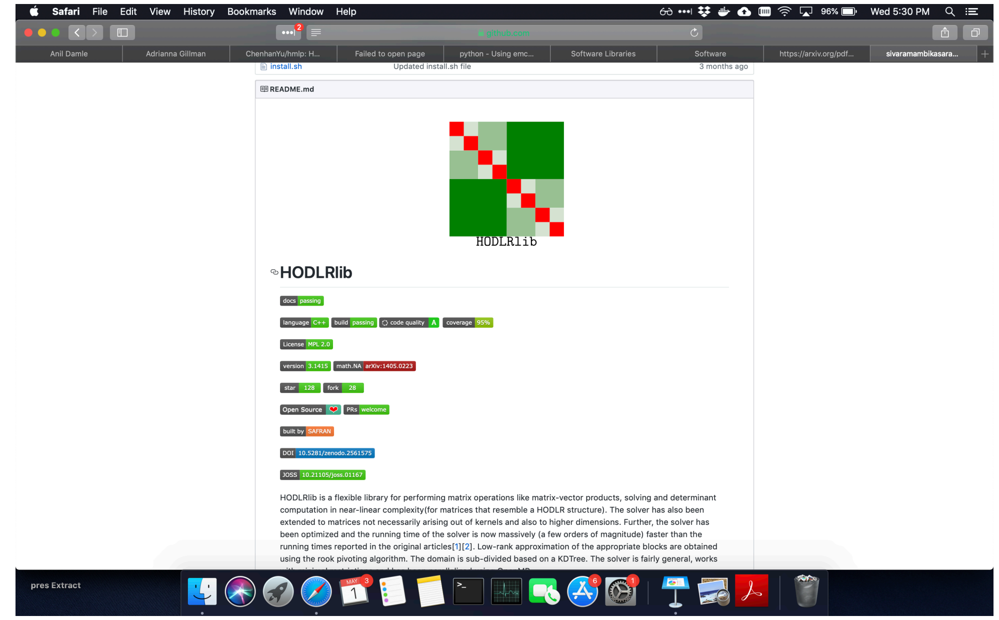


Affords fewer number of parameters in Neural net representation and fast application of the forward network

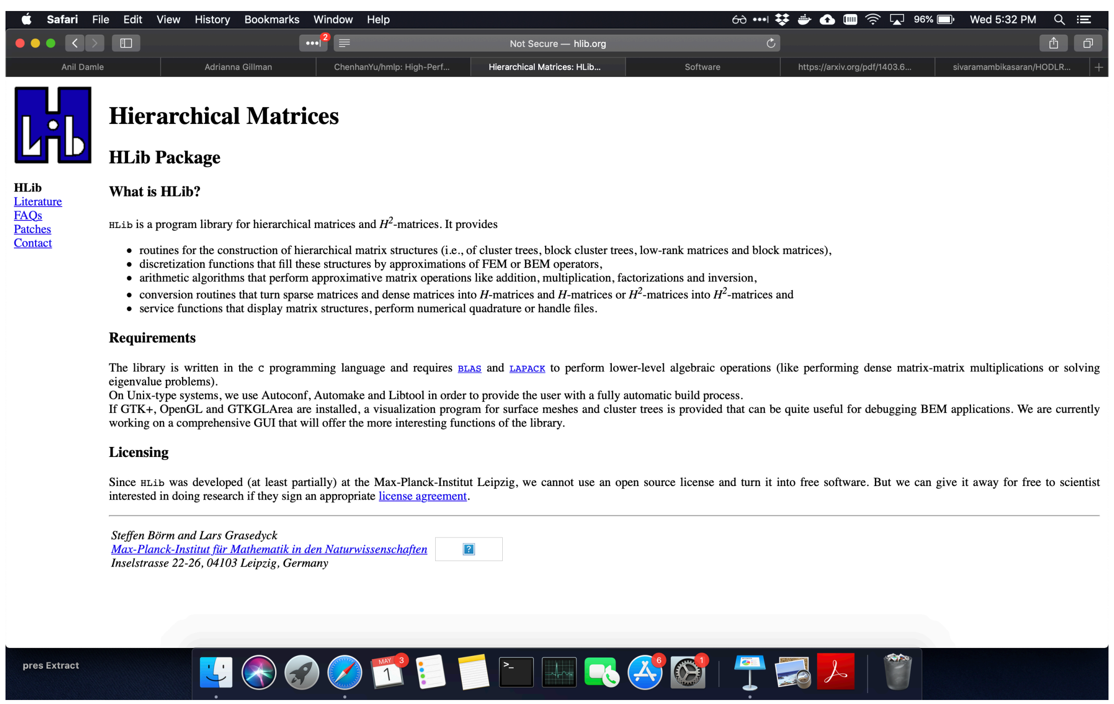
Software



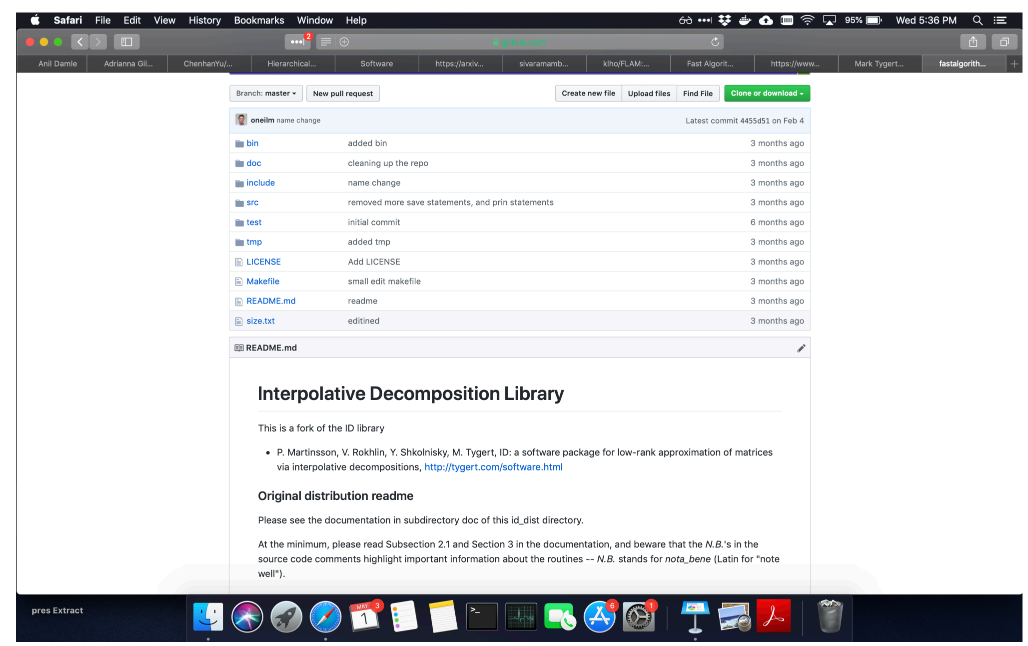
https://github.com/klho/FLAM



https://github.com/sivaramambikasaran/HODLR

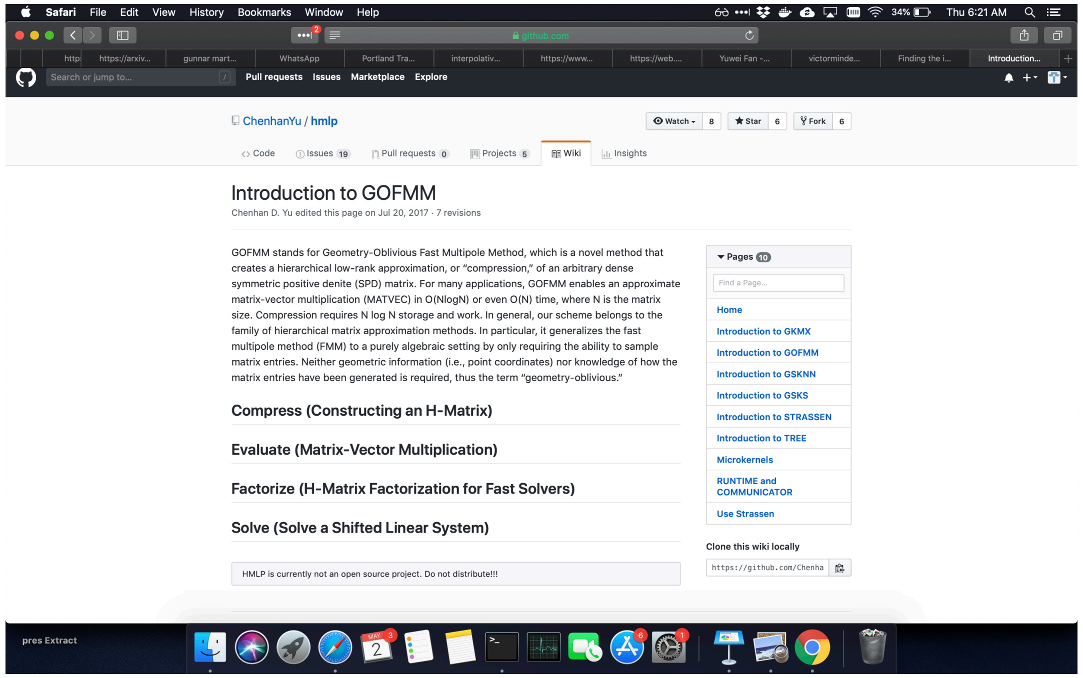


http://www.hlib.org

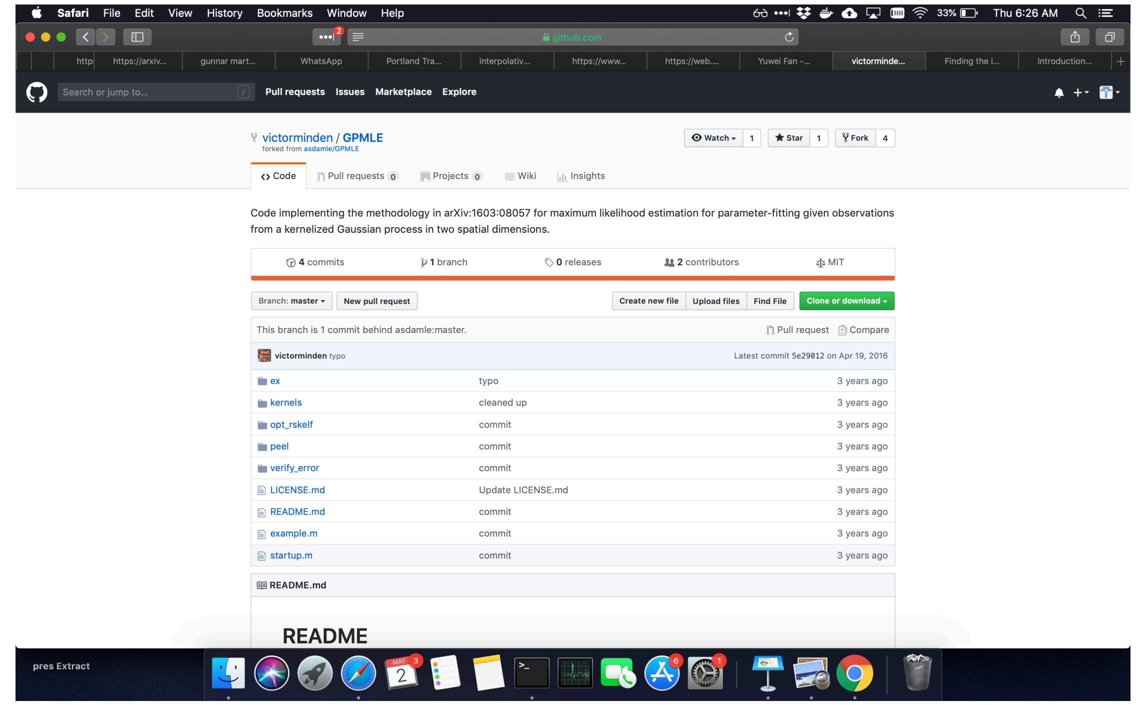


https://github.com/fastalgorithms/libid

More resources



https://github.com/ChenhanYu/hmlp/wiki/Introduction-to-GOFMM



https://github.com/victorminden/GPMLE

- Video lectures by Gunnar - https://www.youtube.com/playlist?list=PLPDZ9rclfxyOrlpcu_D1PRcyK-o2iofwW
- Excellent review article on randomized methods for low rank approximations - Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: https://arxiv.org/pdf/0909.4061.pdf
- Some of the illustrations courtesy: Sivaram Ambikasaran, Per-Gunnar Martinsson, Ken Ho, Lesliie Greengard, Lexing Ying, Adrianna Gillman

References

- Ho, Kenneth L., and Leslie Greengard. "A fast direct solver for structured linear systems by recursive skeletonization." *SIAM Journal on Scientific Computing* 34.5 (2012): A2507-A2532.
- Yu, Chenhan D., et al. "Geometry-oblivious FMM for compressing dense SPD matrices." *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM, 2017.
- Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. (2016). Fast direct methods for Gaussian processes. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), 252-265.
- Gillman, A., Young, P. M., & Martinsson, P. G. (2012). A direct solver with O (N) complexity for integral equations on one-dimensional domains. *Frontiers of Mathematics in China*, *7*(2), 217-247.
- Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, *53*(2), 217-288.
- Greengard, L., & Rokhlin, V. (1991). On the numerical solution of two‐point boundary value problems. *Communications on Pure and Applied Mathematics*, *44*(4), 419-452.
- Fan, Y., Lin, L., Ying, L., & Zepeda-Núnez, L. (2018). A multiscale neural network based on hierarchical matrices. *arXiv preprint arXiv:1807.01883*.
- Minden, V., Damle, A., Ho, K. L., & Ying, L. (2017). Fast spatial gaussian process maximum likelihood estimation via skeletonization factorizations. *Multiscale Modeling & Simulation*, *15*(4), 1584-1611.
- Martinsson, P. G., Rokhlin, V., Shkolnisky, Y., & Tygert, M. (2008). ID: A software package for low-rank approximation of matrices via interpolative decompositions, Version 0.2.
- Corona, E., Martinsson, P. G., & Zorin, D. (2015). An O (N) direct solver for integral equations on the plane. *Applied and Computational Harmonic Analysis*, *38*(2), 284-317.
- Gimbutas, Z., & Rokhlin, V. (2003). A generalized fast multipole method for nonoscillatory kernels. *SIAM Journal on Scientific Computing*, *24*(3), 796-817.

Not an exhaustive list