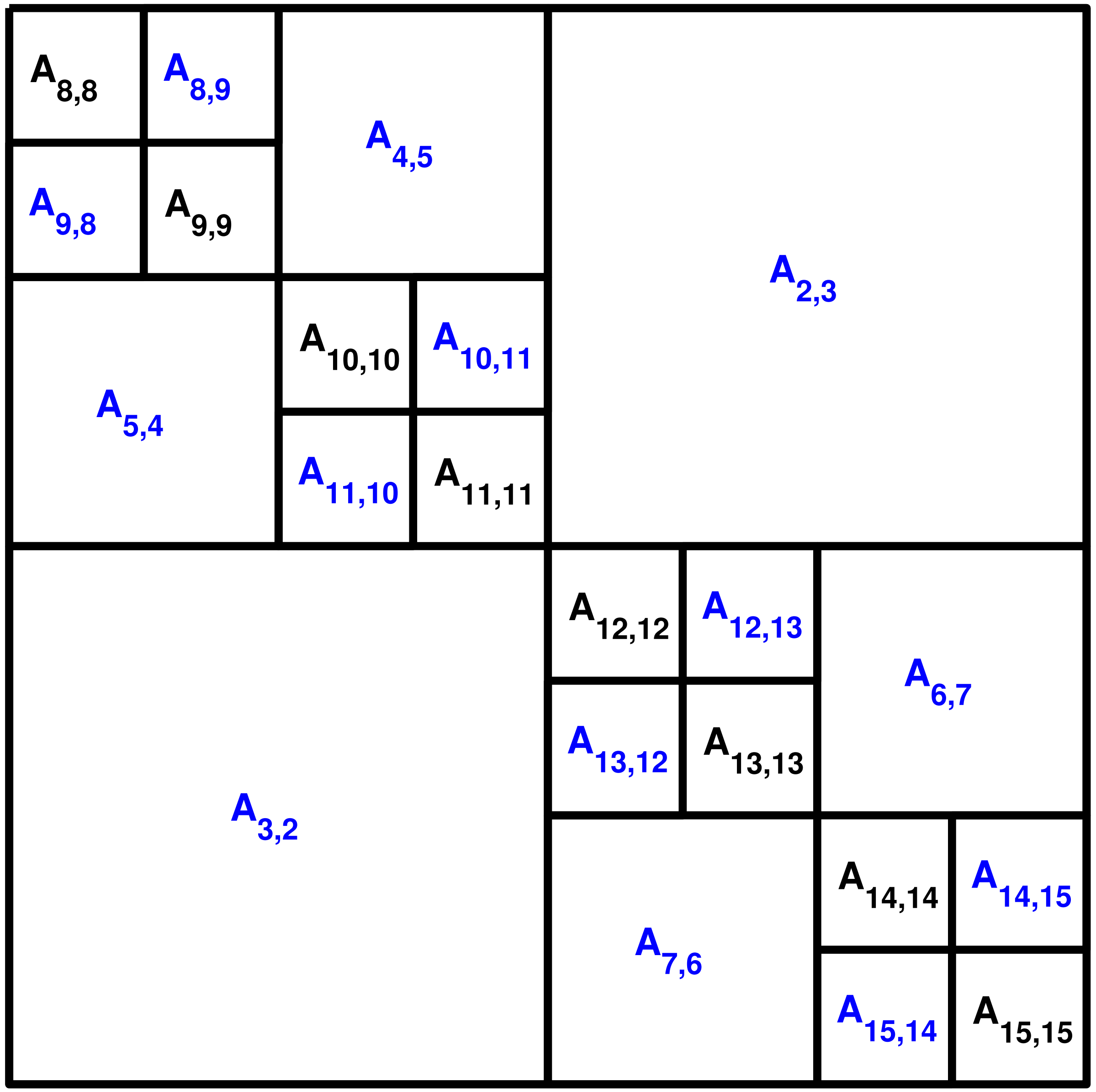
APPM 4720/5720 — week 15:

$O(N)$ inversion of rank-structured matrices

Gunnar Martinsson
The University of Colorado at Boulder

Recall: A very simple format for rank-structured matrices ...

We informally say that a matrix is in $\mathcal{S}$-format if it can be tesselated "like this":



We require that

- the diagonal blocks are of size at most $2k \times 2k$
- the off-diagonal blocks (in blue in the figure) have rank at most k.

The cost of performing a matvec is then

$$\underbrace{2 \times \frac{N}{2} k + 4 \times \frac{N}{4} k + 8 \times \frac{N}{8} k + \cdots}_{\log N \text{ terms}} \sim N \log(N)\, k.$$

Note: The "S" in "$\mathcal{S}$-matrix" is for Simple — the term is not standard by any means ...

Recall that inversion of an $\mathcal{S}$-matrix is a rather complicated operation — multiple traversals up and down the tree, various log-factors in complexity estimates, etc. To overcome these problems and attain $O(N)$ complexity, let us first introduce so called *block separable* matrices. Consider a linear system

$$\mathbf{A}\,\mathbf{q} = \mathbf{f},$$

where **A** is a "block-separable" matrix consisting of $p \times p$ blocks of size $n \times n$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{D}_4 & \mathbf{A}_{45} & \mathbf{A}_{46} & \mathbf{A}_{47} \\ \mathbf{A}_{54} & \mathbf{D}_5 & \mathbf{A}_{56} & \mathbf{A}_{57} \\ \mathbf{A}_{64} & \mathbf{A}_{65} & \mathbf{D}_6 & \mathbf{A}_{67} \\ \mathbf{A}_{74} & \mathbf{A}_{75} & \mathbf{A}_{76} & \mathbf{D}_7 \end{bmatrix}. \qquad \text{(Shown for } p = 4.\text{)}$$

Core assumption: Each off-diagonal block $\mathbf{A}_{ij}$ admits the factorization

$$\begin{array}{cccc} \mathbf{A}_{ij} & = \quad \mathbf{U}_i & \tilde{\mathbf{A}}_{ij} & \mathbf{V}_j^* \\ n \times n & n \times k & k \times k & k \times n \end{array}$$

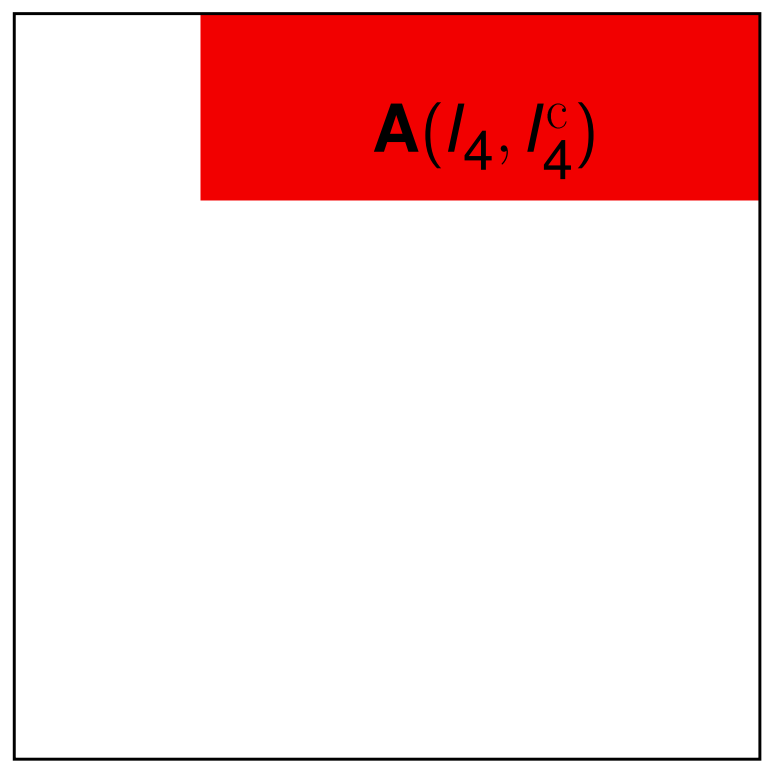
where the rank k is significantly smaller than the block size n.

The critical part of the assumption is that all off-diagonal blocks in the i'th row use the same basis matrices $\mathbf{U}_i$ for their column spaces (and analogously all blocks in the j'th column use the same basis matrices $\mathbf{V}_j$ for their row spaces).

What is the role of the basis matrices $\mathbf{U}_\tau$ and $\mathbf{V}_\tau$?

Recall our toy example:
$$\mathbf{A} = \begin{bmatrix} \mathbf{D}_4 & \mathbf{U}_4\,\tilde{\mathbf{A}}_{45}\,\mathbf{V}_5^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{46}\,\mathbf{V}_6^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{47}\,\mathbf{V}_7^* \\ \mathbf{U}_5\,\tilde{\mathbf{A}}_{54}\,\mathbf{V}_4^* & \mathbf{D}_5 & \mathbf{U}_5\,\tilde{\mathbf{A}}_{56}\,\mathbf{V}_6^* & \mathbf{U}_5\,\tilde{\mathbf{A}}_{57}\,\mathbf{V}_7^* \\ \mathbf{U}_6\,\tilde{\mathbf{A}}_{64}\,\mathbf{V}_4^* & \mathbf{U}_6\,\tilde{\mathbf{A}}_{65}\,\mathbf{V}_5^* & \mathbf{D}_6 & \mathbf{U}_6\,\tilde{\mathbf{A}}_{67}\,\mathbf{V}_7^* \\ \mathbf{U}_7\,\tilde{\mathbf{A}}_{74}\,\mathbf{V}_4^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{75}\,\mathbf{V}_5^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{76}\,\mathbf{V}_6^* & \mathbf{D}_7 \end{bmatrix}.$$

We see that the columns of $\mathbf{U}_4$ must span the column space of the matrix $\mathbf{A}(I_4, I_4^{\rm c})$ where I_4 is the index vector for the first block and $I_4^{\rm c} = I \backslash I_4$.

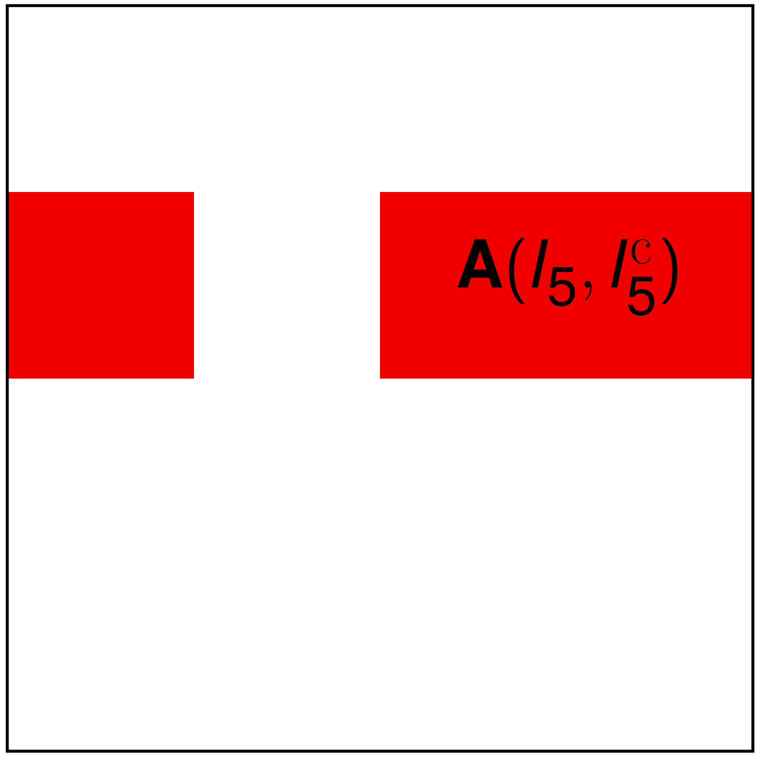


The matrix $\mathbf{A}$

What is the role of the basis matrices $\mathbf{U}_\tau$ and $\mathbf{V}_\tau$?

Recall our toy example: $\mathbf{A} = \begin{bmatrix} \mathbf{D}_4 & \mathbf{U}_4\,\tilde{\mathbf{A}}_{45}\,\mathbf{V}_5^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{46}\,\mathbf{V}_6^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{47}\,\mathbf{V}_7^* \\ \mathbf{U}_5\,\tilde{\mathbf{A}}_{54}\,\mathbf{V}_4^* & \mathbf{D}_5 & \mathbf{U}_5\,\tilde{\mathbf{A}}_{56}\,\mathbf{V}_6^* & \mathbf{U}_5\,\tilde{\mathbf{A}}_{57}\,\mathbf{V}_7^* \\ \mathbf{U}_6\,\tilde{\mathbf{A}}_{64}\,\mathbf{V}_4^* & \mathbf{U}_6\,\tilde{\mathbf{A}}_{65}\,\mathbf{V}_5^* & \mathbf{D}_6 & \mathbf{U}_6\,\tilde{\mathbf{A}}_{67}\,\mathbf{V}_7^* \\ \mathbf{U}_7\,\tilde{\mathbf{A}}_{74}\,\mathbf{V}_4^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{75}\,\mathbf{V}_5^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{76}\,\mathbf{V}_6^* & \mathbf{D}_7 \end{bmatrix}.$

We see that the columns of $\mathbf{U}_5$ must span the column space of the matrix $\mathbf{A}(I_5, I_5^{\mathrm{c}})$ where I_5 is the index vector for the first block and $I_5^{\mathrm{c}} = I \backslash I_5$.



The matrix $\mathbf{A}$

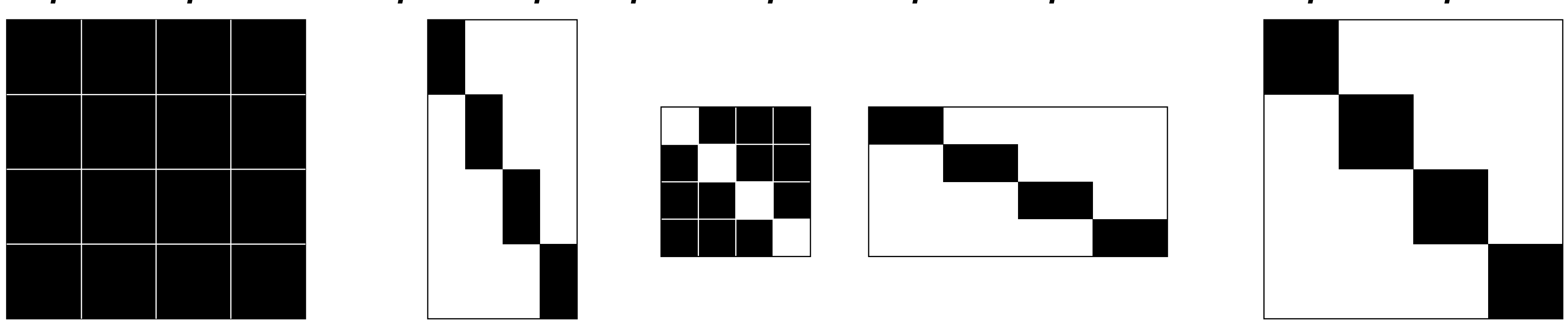
Recall $\mathbf{A} = \begin{bmatrix} \mathbf{D}_4 & \mathbf{U}_4\,\tilde{\mathbf{A}}_{45}\,\mathbf{V}_5^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{46}\,\mathbf{V}_6^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{47}\,\mathbf{V}_7^* \\ \mathbf{U}_5\,\tilde{\mathbf{A}}_{54}\,\mathbf{V}_4^* & \mathbf{D}_5 & \mathbf{U}_5\,\tilde{\mathbf{A}}_{56}\,\mathbf{V}_6^* & \mathbf{U}_5\,\tilde{\mathbf{A}}_{57}\,\mathbf{V}_7^* \\ \mathbf{U}_6\,\tilde{\mathbf{A}}_{64}\,\mathbf{V}_4^* & \mathbf{U}_6\,\tilde{\mathbf{A}}_{65}\,\mathbf{V}_5^* & \mathbf{D}_6 & \mathbf{U}_6\,\tilde{\mathbf{A}}_{67}\,\mathbf{V}_7^* \\ \mathbf{U}_7\,\tilde{\mathbf{A}}_{74}\,\mathbf{V}_4^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{75}\,\mathbf{V}_5^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{76}\,\mathbf{V}_6^* & \mathbf{D}_7 \end{bmatrix}$.

Then $\mathbf{A}$ admits the factorization:

$$\mathbf{A} = \underbrace{\begin{bmatrix} \mathbf{U}_4 & & & \\ & \mathbf{U}_5 & & \\ & & \mathbf{U}_6 & \\ & & & \mathbf{U}_7 \end{bmatrix}}_{=\mathbf{U}} \underbrace{\begin{bmatrix} \mathbf{0} & \tilde{\mathbf{A}}_{45} & \tilde{\mathbf{A}}_{46} & \tilde{\mathbf{A}}_{47} \\ \tilde{\mathbf{A}}_{54} & \mathbf{0} & \tilde{\mathbf{A}}_{56} & \tilde{\mathbf{A}}_{57} \\ \tilde{\mathbf{A}}_{64} & \tilde{\mathbf{A}}_{65} & \mathbf{0} & \tilde{\mathbf{A}}_{67} \\ \tilde{\mathbf{A}}_{74} & \tilde{\mathbf{A}}_{75} & \tilde{\mathbf{A}}_{76} & \mathbf{0} \end{bmatrix}}_{=\tilde{\mathbf{A}}} \underbrace{\begin{bmatrix} \mathbf{V}_4^* & & & \\ & \mathbf{V}_5^* & & \\ & & \mathbf{V}_6^* & \\ & & & \mathbf{V}_7^* \end{bmatrix}}_{=\mathbf{V}^*} + \underbrace{\begin{bmatrix} \mathbf{D}_4 & & & \\ & \mathbf{D}_5 & & \\ & & \mathbf{D}_6 & \\ & & & \mathbf{D}_7 \end{bmatrix}}_{=\mathbf{D}}$$

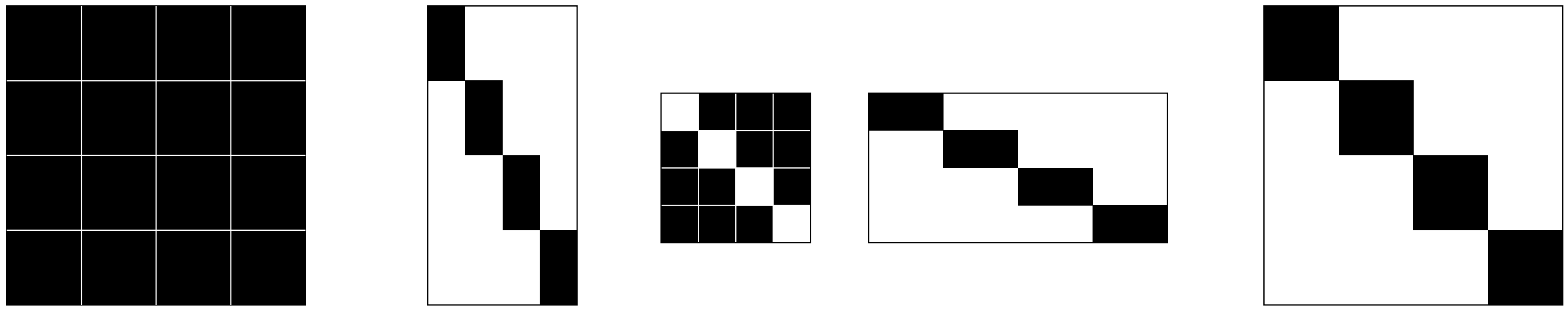
or

$$\underset{pn \times pn}{\mathbf{A}} = \underset{pn \times pk}{\mathbf{U}} \;\; \underset{pk \times pk}{\tilde{\mathbf{A}}} \;\; \underset{pk \times pn}{\mathbf{V}^*} + \underset{pn \times pn}{\mathbf{D}},$$



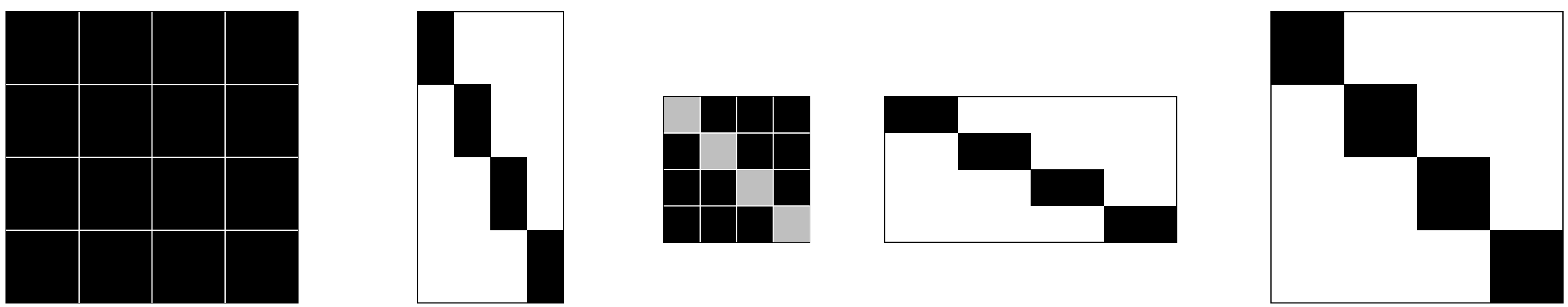
Lemma: [Variation of Woodbury] If an $N \times N$ matrix $\mathbf{A}$ admits the factorization

$$\underset{pn \times pn}{\mathbf{A}} = \underset{pn \times pk}{\mathbf{U}} \; \underset{pk \times pk}{\tilde{\mathbf{A}}} \; \underset{pk \times pn}{\mathbf{V}^*} + \underset{pn \times pn}{\mathbf{D}},$$



then

$$\underset{pn \times pn}{\mathbf{A}^{-1}} = \underset{pn \times pk}{\mathbf{E}} \; \underset{pk \times pk}{(\tilde{\mathbf{A}} + \hat{\mathbf{D}})^{-1}} \; \underset{pk \times pn}{\mathbf{F}^*} + \underset{pn \times pn}{\mathbf{G}},$$



where (provided all intermediate matrices are invertible)

$$\hat{\mathbf{D}} = \left(\mathbf{V}^* \mathbf{D}^{-1} \mathbf{U}\right)^{-1}, \quad \mathbf{E} = \mathbf{D}^{-1} \mathbf{U} \hat{\mathbf{D}}, \quad \mathbf{F} = (\hat{\mathbf{D}} \mathbf{V}^* \mathbf{D}^{-1})^*, \quad \mathbf{G} = \mathbf{D}^{-1} - \mathbf{D}^{-1} \mathbf{U} \hat{\mathbf{D}} \mathbf{V}^* \mathbf{D}^{-1}.$$

Note: All matrices set in blue are block diagonal.

Classical Woodbury: $\left(\mathbf{D} + \mathbf{U}\tilde{\mathbf{A}}\mathbf{V}^*\right)^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{U}\left(\tilde{\mathbf{A}} + \mathbf{V}^*\mathbf{D}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^*\mathbf{D}^{-1}.$

Derivation of "our" Woodbury: We consider the linear system

$$\begin{bmatrix} \mathbf{D}_4 & \mathbf{U}_4\,\tilde{\mathbf{A}}_{45}\,\mathbf{V}_5^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{46}\,\mathbf{V}_6^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{47}\,\mathbf{V}_7^* \\ \mathbf{U}_5\,\tilde{\mathbf{A}}_{54}\,\mathbf{V}_4^* & \mathbf{D}_5 & \mathbf{U}_5\,\tilde{\mathbf{A}}_{56}\,\mathbf{V}_6^* & \mathbf{U}_5\,\tilde{\mathbf{A}}_{57}\,\mathbf{V}_7^* \\ \mathbf{U}_6\,\tilde{\mathbf{A}}_{64}\,\mathbf{V}_4^* & \mathbf{U}_6\,\tilde{\mathbf{A}}_{65}\,\mathbf{V}_5^* & \mathbf{D}_6 & \mathbf{U}_6\,\tilde{\mathbf{A}}_{67}\,\mathbf{V}_7^* \\ \mathbf{U}_7\,\tilde{\mathbf{A}}_{74}\,\mathbf{V}_4^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{75}\,\mathbf{V}_5^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{76}\,\mathbf{V}_6^* & \mathbf{D}_7 \end{bmatrix} \begin{bmatrix} \mathbf{q}_4 \\ \mathbf{q}_5 \\ \mathbf{q}_6 \\ \mathbf{q}_7 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_4 \\ \mathbf{f}_5 \\ \mathbf{f}_6 \\ \mathbf{f}_7 \end{bmatrix}.$$

Introduce *reduced variables* $\tilde{\mathbf{q}}_i = \mathbf{V}_i^*\,\mathbf{q}_i$.

The system $\sum_j \mathbf{A}_{ij}\,\mathbf{q}_j = \mathbf{f}_i$ then takes the form

$$\left[\begin{array}{cccc|cccc} \mathbf{D}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{U}_4\tilde{\mathbf{A}}_{45} & \mathbf{U}_4\tilde{\mathbf{A}}_{46} & \mathbf{U}_4\tilde{\mathbf{A}}_{47} \\ \mathbf{0} & \mathbf{D}_5 & \mathbf{0} & \mathbf{0} & \mathbf{U}_5\tilde{\mathbf{A}}_{54} & \mathbf{0} & \mathbf{U}_5\tilde{\mathbf{A}}_{56} & \mathbf{U}_5\tilde{\mathbf{A}}_{57} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_6 & \mathbf{0} & \mathbf{U}_6\tilde{\mathbf{A}}_{64} & \mathbf{U}_6\tilde{\mathbf{A}}_{65} & \mathbf{0} & \mathbf{U}_6\tilde{\mathbf{A}}_{67} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_7 & \mathbf{U}_7\tilde{\mathbf{A}}_{74} & \mathbf{U}_7\tilde{\mathbf{A}}_{75} & \mathbf{U}_7\tilde{\mathbf{A}}_{76} & \mathbf{0} \\ \hline -\mathbf{V}_4^* & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{V}_5^* & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{V}_6^* & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{V}_7^* & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{array}\right] \begin{bmatrix} \mathbf{q}_4 \\ \mathbf{q}_5 \\ \mathbf{q}_6 \\ \mathbf{q}_7 \\ \tilde{\mathbf{q}}_4 \\ \tilde{\mathbf{q}}_5 \\ \tilde{\mathbf{q}}_6 \\ \tilde{\mathbf{q}}_7 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_4 \\ \mathbf{f}_5 \\ \mathbf{f}_6 \\ \mathbf{f}_7 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

Now form the Schur complement to eliminate the $\mathbf{q}_j$'s.

After eliminating the “fine-scale” variables $\mathbf{q}_i$, we obtain

$$\begin{bmatrix} \mathbf{I} & \mathbf{V}_4^*\tilde{\mathbf{A}}_{44}^{-1}\mathbf{U}_4\tilde{\mathbf{A}}_{45} & \mathbf{V}_4^*\tilde{\mathbf{A}}_{44}^{-1}\mathbf{U}_4\tilde{\mathbf{A}}_{46} & \mathbf{V}_4^*\tilde{\mathbf{A}}_{44}^{-1}\mathbf{U}_4\tilde{\mathbf{A}}_{47} \\ \mathbf{V}_5^*\tilde{\mathbf{A}}_{55}^{-1}\mathbf{U}_5\tilde{\mathbf{A}}_{54} & \mathbf{I} & \mathbf{V}_5^*\tilde{\mathbf{A}}_{55}^{-1}\mathbf{U}_5\tilde{\mathbf{A}}_{56} & \mathbf{V}_5^*\tilde{\mathbf{A}}_{55}^{-1}\mathbf{U}_5\tilde{\mathbf{A}}_{57} \\ \mathbf{V}_6^*\tilde{\mathbf{A}}_{66}^{-1}\mathbf{U}_6\tilde{\mathbf{A}}_{61} & \mathbf{V}_6^*\tilde{\mathbf{A}}_{66}^{-1}\mathbf{U}_6\tilde{\mathbf{A}}_{65} & \mathbf{I} & \mathbf{V}_6^*\tilde{\mathbf{A}}_{66}^{-1}\mathbf{U}_6\tilde{\mathbf{A}}_{67} \\ \mathbf{V}_7^*\tilde{\mathbf{A}}_{77}^{-1}\mathbf{U}_7\tilde{\mathbf{A}}_{74} & \mathbf{V}_7^*\tilde{\mathbf{A}}_{77}^{-1}\mathbf{U}_7\tilde{\mathbf{A}}_{75} & \mathbf{V}_7^*\tilde{\mathbf{A}}_{77}^{-1}\mathbf{U}_7\tilde{\mathbf{A}}_{76} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{q}}_4 \\ \tilde{\mathbf{q}}_5 \\ \tilde{\mathbf{q}}_6 \\ \tilde{\mathbf{q}}_7 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_4^*\mathbf{D}_4^{-1}\,\mathbf{f}_4 \\ \mathbf{V}_5^*\mathbf{D}_5^{-1}\,\mathbf{f}_5 \\ \mathbf{V}_6^*\mathbf{D}_6^{-1}\,\mathbf{f}_6 \\ \mathbf{V}_7^*\mathbf{D}_7^{-1}\,\mathbf{f}_7. \end{bmatrix}.$$

After eliminating the "fine-scale" variables $\mathbf{q}_i$, we obtain

$$\begin{bmatrix} \mathbf{I} & \mathbf{V}_4^*\tilde{\mathbf{A}}_{44}^{-1}\mathbf{U}_4\tilde{\mathbf{A}}_{45} & \mathbf{V}_4^*\tilde{\mathbf{A}}_{44}^{-1}\mathbf{U}_4\tilde{\mathbf{A}}_{46} & \mathbf{V}_4^*\tilde{\mathbf{A}}_{44}^{-1}\mathbf{U}_4\tilde{\mathbf{A}}_{47} \\ \mathbf{V}_5^*\tilde{\mathbf{A}}_{55}^{-1}\mathbf{U}_5\tilde{\mathbf{A}}_{54} & \mathbf{I} & \mathbf{V}_5^*\tilde{\mathbf{A}}_{55}^{-1}\mathbf{U}_5\tilde{\mathbf{A}}_{56} & \mathbf{V}_5^*\tilde{\mathbf{A}}_{55}^{-1}\mathbf{U}_5\tilde{\mathbf{A}}_{57} \\ \mathbf{V}_6^*\tilde{\mathbf{A}}_{66}^{-1}\mathbf{U}_6\tilde{\mathbf{A}}_{61} & \mathbf{V}_6^*\tilde{\mathbf{A}}_{66}^{-1}\mathbf{U}_6\tilde{\mathbf{A}}_{65} & \mathbf{I} & \mathbf{V}_6^*\tilde{\mathbf{A}}_{66}^{-1}\mathbf{U}_6\tilde{\mathbf{A}}_{67} \\ \mathbf{V}_7^*\tilde{\mathbf{A}}_{77}^{-1}\mathbf{U}_7\tilde{\mathbf{A}}_{74} & \mathbf{V}_7^*\tilde{\mathbf{A}}_{77}^{-1}\mathbf{U}_7\tilde{\mathbf{A}}_{75} & \mathbf{V}_7^*\tilde{\mathbf{A}}_{77}^{-1}\mathbf{U}_7\tilde{\mathbf{A}}_{76} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{q}}_4 \\ \tilde{\mathbf{q}}_5 \\ \tilde{\mathbf{q}}_6 \\ \tilde{\mathbf{q}}_7 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_4^*\mathbf{D}_4^{-1}\,\mathbf{f}_4 \\ \mathbf{V}_5^*\mathbf{D}_5^{-1}\,\mathbf{f}_5 \\ \mathbf{V}_6^*\mathbf{D}_6^{-1}\,\mathbf{f}_6 \\ \mathbf{V}_7^*\mathbf{D}_7^{-1}\,\mathbf{f}_7. \end{bmatrix}.$$

We set

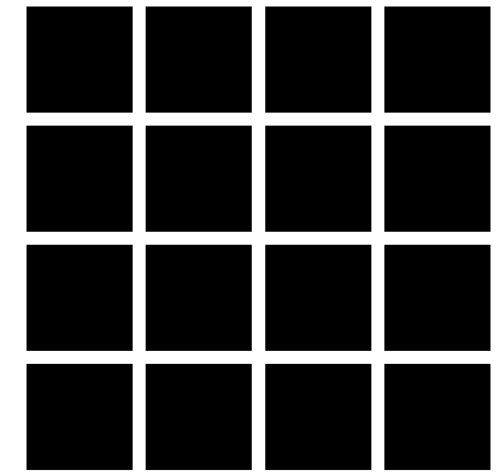
$$\tilde{\mathbf{A}}_{ii} = \left(\mathbf{V}_i^*\,\mathbf{D}_{ii}^{-1}\,\mathbf{U}_i\right)^{-1},$$

and multiply line i by $\tilde{\mathbf{A}}_{ii}$ to obtain the reduced system

$$\begin{bmatrix} \tilde{\mathbf{A}}_{44} & \tilde{\mathbf{A}}_{45} & \tilde{\mathbf{A}}_{46} & \tilde{\mathbf{A}}_{47} \\ \tilde{\mathbf{A}}_{54} & \tilde{\mathbf{A}}_{55} & \tilde{\mathbf{A}}_{56} & \tilde{\mathbf{A}}_{57} \\ \tilde{\mathbf{A}}_{64} & \tilde{\mathbf{A}}_{65} & \tilde{\mathbf{A}}_{66} & \tilde{\mathbf{A}}_{67} \\ \tilde{\mathbf{A}}_{74} & \tilde{\mathbf{A}}_{75} & \tilde{\mathbf{A}}_{76} & \tilde{\mathbf{A}}_{77} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{q}}_4 \\ \tilde{\mathbf{q}}_5 \\ \tilde{\mathbf{q}}_6 \\ \tilde{\mathbf{q}}_7 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{f}}_4 \\ \tilde{\mathbf{f}}_5 \\ \tilde{\mathbf{f}}_6 \\ \tilde{\mathbf{f}}_7 \end{bmatrix}.$$

where

$$\tilde{\mathbf{f}}_i = \tilde{\mathbf{A}}_{ii}\,\mathbf{V}_i^*\,\mathbf{D}_{ii}^{-1}\,\mathbf{f}_i.$$



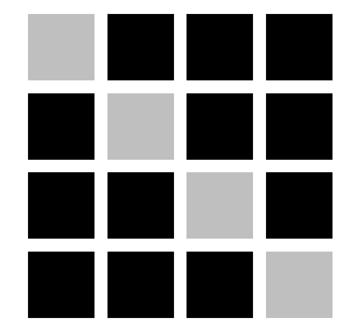
The original matrix

Before compression, we have a $pn \times pn$ linear system

$$\sum_{j=1}^{p} \mathbf{A}_{ij}\mathbf{q}_j = \mathbf{f}_i, \quad i = 1, 2, \ldots, p.$$

After compression, we have a $pk \times pk$ linear system

$$\mathbf{D}_{ii}\tilde{\mathbf{q}}_i + \sum_{i \neq j} \tilde{\mathbf{A}}_{ij}\tilde{\mathbf{q}}_j = \tilde{\mathbf{f}}_i, \quad i = 1, 2, \ldots, p.$$



The reduced matrix

Recall that k is the ε-rank of $\mathbf{A}_{i,j}$ for $i \neq j$.

The point is that $k < n$.

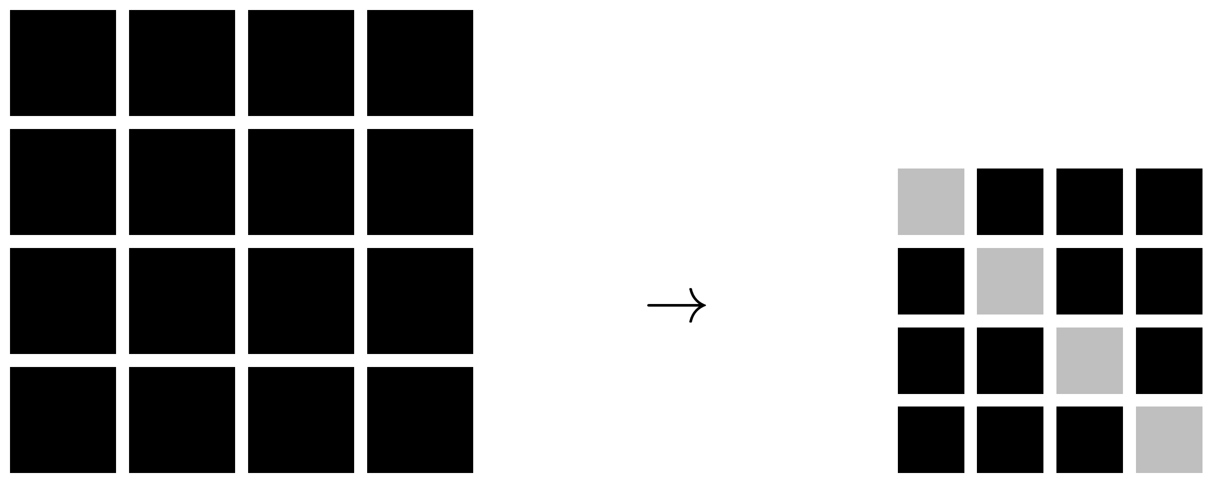
The compression algorithm needs to execute the following steps:

- Compute $\mathbf{U}_i$, $\mathbf{V}_i$, $\tilde{\mathbf{A}}_{ij}$ so that $\mathbf{A}_{ij} = \mathbf{U}_i\,\tilde{\mathbf{A}}_{ij}\,\mathbf{V}_j^*$.
- Compute the new diagonal matrices $\hat{\mathbf{D}}_{ii} = \left(\mathbf{V}_i^*\,\mathbf{A}_{ii}^{-1}\,\mathbf{U}_i\right)^{-1}$.
- Compute the new loads $\tilde{\mathbf{q}}_i = \hat{\mathbf{D}}_{ii}\,\mathbf{V}_i^*\,\mathbf{A}_{ii}^{-1}\,\mathbf{q}_i$.

For the algorithm to be efficient, it has to be able to carry out these steps *locally*.

To achieve this, we use interpolative representations, then $\tilde{\mathbf{A}}_{i,j} = \mathbf{A}(\tilde{I}_i, \tilde{I}_j)$.

We have built a scheme for reducing a system of size $pn \times pn$ to one of size $pk \times pk$.

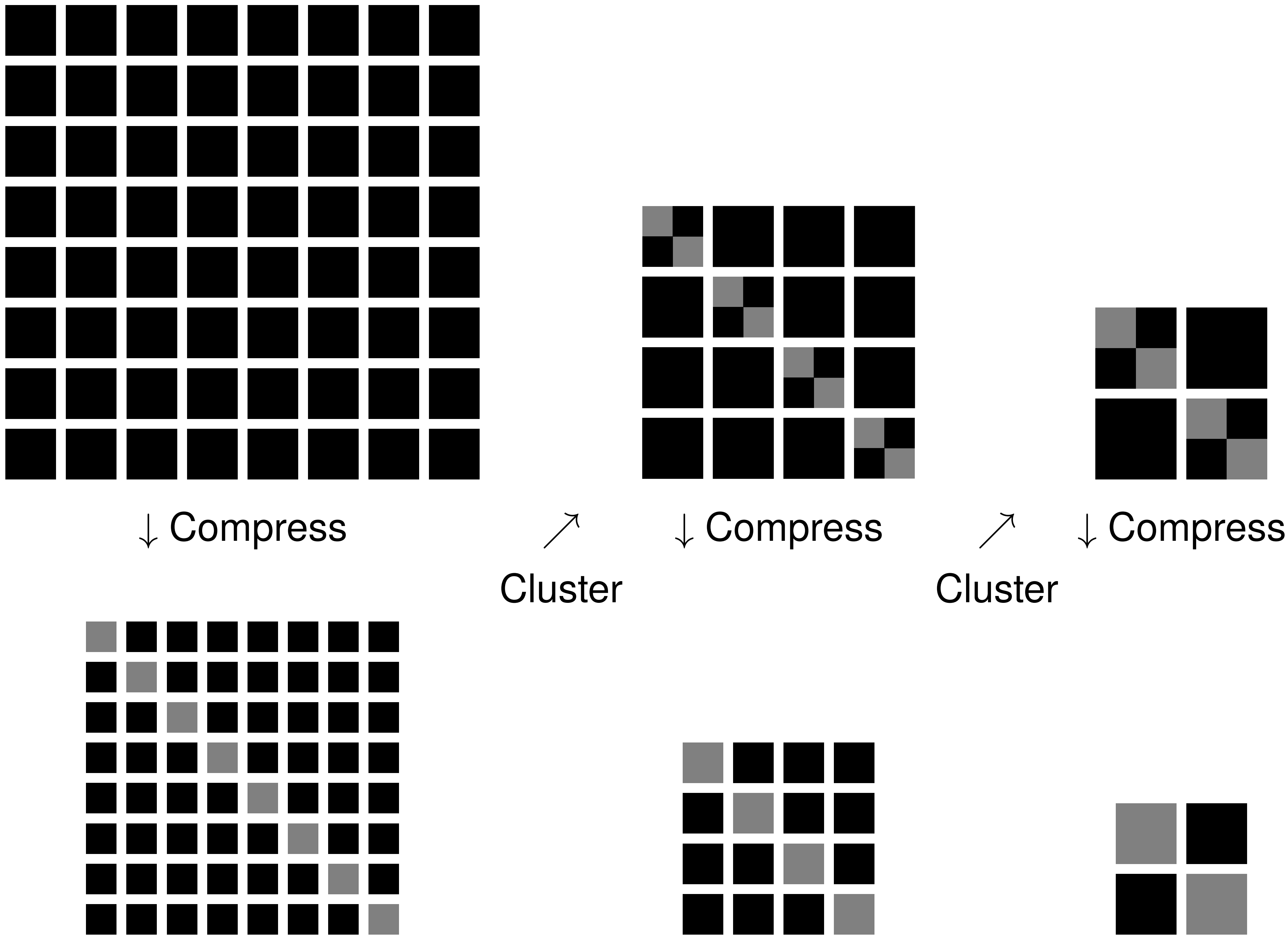


The computational gain is $(k/n)^3$. Good, but not earth-shattering.

Question: How do we get to $O(N)$?

Answer: It turns out that the reduced matrix is itself compressible. Recurse!

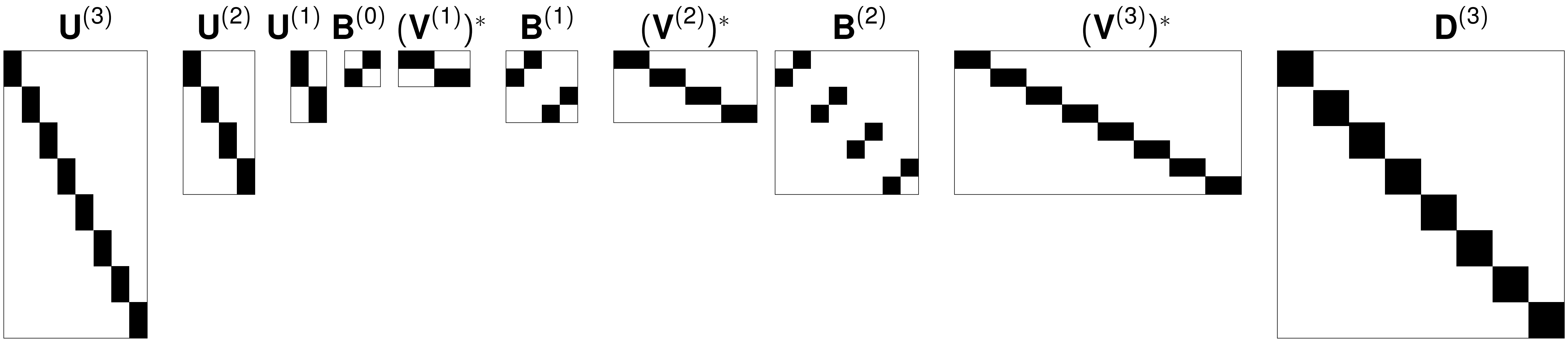
A globally $O(N)$ algorithm is obtained by hierarchically repeating the process:



Formally, one can view this as a telescoping factorization of **A**:

$$\mathbf{A} = \mathbf{U}^{(3)}\big(\mathbf{U}^{(2)}\big(\mathbf{U}^{(1)}\,\mathbf{B}^{(0)}\,(\mathbf{V}^{(1)})^{*} + \mathbf{B}^{(1)}\big)(\mathbf{V}^{(2)})^{*} + \mathbf{B}^{(2)}\big)(\mathbf{V}^{(3)})^{*} + \mathbf{D}^{(3)}.$$

Expressed pictorially, the factorization takes the form



The *inverse of A* then takes the form

$$\mathbf{A}^{-1} = \mathbf{E}^{(3)}\big(\mathbf{E}^{(2)}\big(\mathbf{E}^{(1)}\,\hat{\mathbf{D}}^{(0)}\,(\mathbf{F}^{(1)})^{*} + \hat{\mathbf{D}}^{(1)}\big)(\mathbf{F}^{(2)})^{*} + \hat{\mathbf{D}}^{(2)}\big)(\mathbf{V}^{(3)})^{*} + \hat{\mathbf{D}}^{(3)}.$$

All matrices are block diagonal except $\hat{\mathbf{D}}^{(0)}$*, which is small.*

Formal definition of an HBS matrix

Let us first recall the concept of a binary tree on the index vector:

Let $\mathbf{A}$ be an $N \times N$ matrix.

Suppose $\mathcal{T}$ is a binary tree on the index vector $I = [1,\, 2,\, 3,\, \dots,\, N]$.

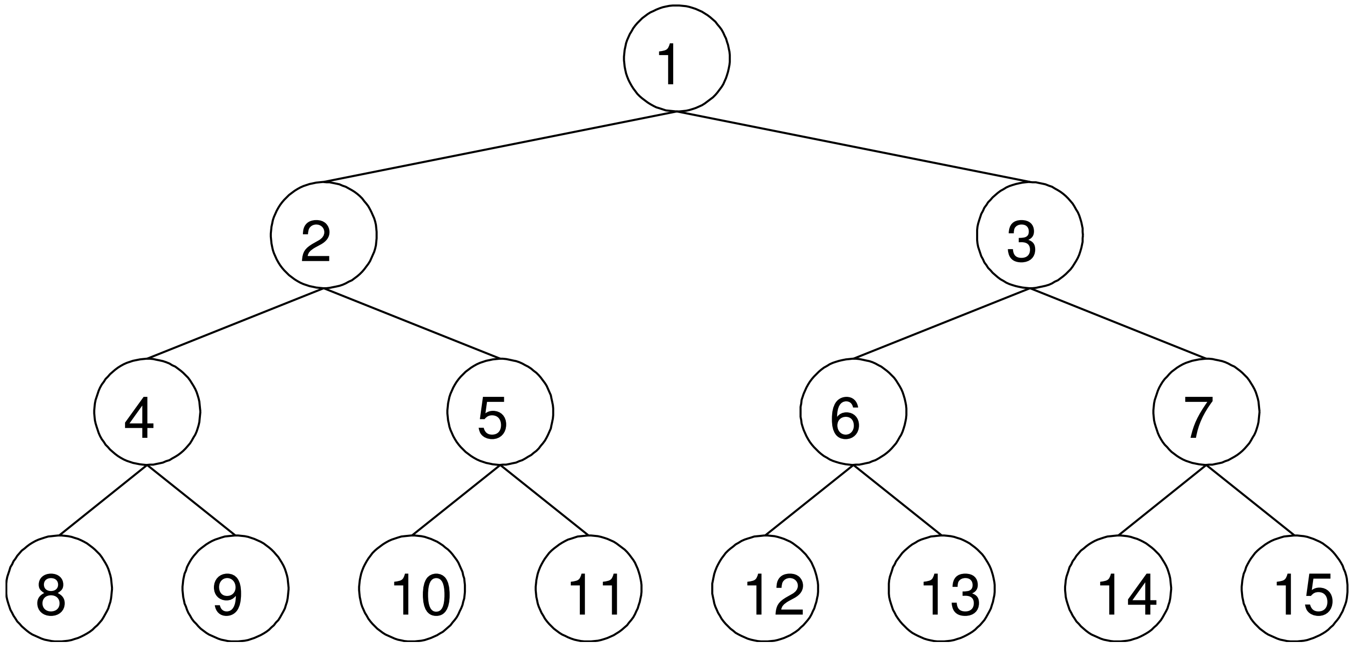
For a node τ in the tree, let I_τ denote the corresponding index vector.

Level 0

Level 1

Level 2

Level 3



$I_1 = [1,\, 2,\, \dots,\, 400]$

$I_2 = [1,\, 2,\, \dots,\, 200]$, $I_3 = [201,\, 202,\, \dots,\, 400]$

$I_4 = [1,\, 2,\, \dots,\, 100]$, $I_5 = [101,\, 102,\, \dots,\, 200]$, $\dots$

$I_8 = [1,\, 2,\, \dots,\, 50]$, $I_9 = [51,\, 52,\, \dots,\, 100]$, $\dots$

For nodes σ and τ on the same level, set $\mathbf{A}_{\sigma,\tau} = \mathbf{A}(I_\sigma, I_\tau)$.

Formal definition of an HBS matrix

Suppose $\mathcal{T}$ is a binary tree.

For a node τ in the tree, let I_τ denote the corresponding index vector.

For leaves σ and τ, set $\mathbf{A}_{\sigma,\tau} = \mathbf{A}(I_\sigma, I_\tau)$ and suppose that all off-diagonal blocks satisfy

$$\begin{array}{cccccc} \mathbf{A}_{\sigma,\tau} & = & \mathbf{U}_\sigma & \tilde{\mathbf{A}}_{\sigma,\tau} & \mathbf{V}_\tau^* & \qquad \sigma \neq \tau \\ n \times n & & n \times k & k \times k & k \times n & \end{array}$$

For non-leaves σ and τ, let $\{\sigma_1, \sigma_2\}$ denote the children of σ, and let $\{\tau_1, \tau_2\}$ denote the children of τ. Set

$$\mathbf{A}_{\sigma,\tau} = \begin{bmatrix} \tilde{\mathbf{A}}_{\sigma_1,\tau_1} & \tilde{\mathbf{A}}_{\sigma_1,\tau_2} \\ \tilde{\mathbf{A}}_{\sigma_2,\tau_1} & \tilde{\mathbf{A}}_{\sigma_2,\tau_2} \end{bmatrix}$$

Then suppose that the off-diagonal blocks satisfy

$$\begin{array}{cccccc} \mathbf{A}_{\sigma,\tau} & = & \mathbf{U}_\sigma & \tilde{\mathbf{A}}_{\sigma,\tau} & \mathbf{V}_\tau^* & \qquad \sigma \neq \tau \\ 2k \times 2k & & 2k \times k & k \times k & k \times 2k & \end{array}$$

An HBS matrix **A** *associated with a tree* $\mathcal{T}$ *is specified by the following factors:*

	Name:	Size:	Function:
For each leaf node τ:	$\mathbf{D}_\tau$	$n \times n$	The diagonal block $\mathbf{A}(I_\tau, I_\tau)$.
	$\mathbf{U}_\tau$	$n \times k$	Basis for the columns in the blocks in row τ.
	$\mathbf{V}_\tau$	$n \times k$	Basis for the rows in the blocks in column τ.
For each parent node τ:	$\mathbf{B}_\tau$	$2k \times 2k$	Interactions between the children of τ.
	$\mathbf{U}_\tau$	$2k \times k$	Basis for the columns in the (reduced) blocks in row τ.
	$\mathbf{V}_\tau$	$2k \times k$	Basis for the rows in the (reduced) blocks in column τ.

Inversion of an HBS matrix

loop over all levels, finer to coarser, $\ell = L, L-1, \dots, 1$

loop over all boxes τ on level ℓ,

if τ is a leaf node

$\mathbf{X} = \mathbf{D}_\tau$

else

Let σ_1 and σ_2 denote the children of τ.

$$\mathbf{X} = \begin{bmatrix} \mathbf{D}_{\sigma_1} & \mathbf{B}_{\sigma_1,\sigma_2} \\ \mathbf{B}_{\sigma_2,\sigma_1} & \mathbf{D}_{\sigma_2} \end{bmatrix}$$

end if

$\mathbf{D}_\tau = \left(\mathbf{V}_\tau^* \, \mathbf{X}^{-1} \, \mathbf{U}_\tau\right)^{-1}$.

$\mathbf{E}_\tau = \mathbf{X}^{-1} \, \mathbf{U}_\tau \, \mathbf{D}_\tau$.

$\mathbf{F}_\tau^* = \mathbf{D}_\tau \, \mathbf{V}_\tau^* \, \mathbf{X}^{-1}$.

$\mathbf{G}_\tau = \mathbf{X}^{-1} - \mathbf{X}^{-1} \, \mathbf{U}_\tau \, \mathbf{D}_\tau \, \mathbf{V}_\tau^* \, \mathbf{X}^{-1}$.

end loop

end loop

$$\mathbf{G}_1 = \begin{bmatrix} \mathbf{D}_2 & \mathbf{B}_{2,3} \\ \mathbf{B}_{3,2} & \mathbf{D}_3 \end{bmatrix}^{-1}.$$

```
function EFG = OMNI_invert_HBS_nsym(NODES)
nboxes = size(NODES,2);
EFG = cell(3,nboxes);
ATD_VEC = cell(1,nboxes);
% Loop over all nodes, from finest to coarser.
for ibox = nboxes:(-1):2
  % Assemble the diagonal matrix.
  if (NODES{5,ibox}==0) % ibox is a leaf.
    AD = NODES{40,ibox};
  elseif (NODES{5,ibox}==2) % ibox has precisely two children
    ison1 = NODES{4,ibox}(1);
    ison2 = NODES{4,ibox}(2);
    AD = [ATD_VEC{ison1},NODES{46,ison1};NODES{46,ison2},ATD_VEC{ison2}];
  end
  % Extract the matrices U and V.
  U = NODES{38,ibox};
  V = NODES{39,ibox};
  % Construct the various projection maps.
  ADinv = inv(AD);
  ATD = inv(V'*ADinv*U);
  ATD_VEC{ibox} = ATD;
  EFG{1,ibox} = ADinv*U*ATD;
  EFG{2,ibox} = ATD*(V')*ADinv;
  EFG{3,ibox} = ADinv - EFG{1,ibox}*(V'*ADinv);
end
% Assemble the "top matrix" and invert it:
AT = [ATD_VEC{2},NODES{46,2};NODES{46,3},ATD_VEC{3}];
EFG{3,1} = inv(AT);
return
```

Now let us return to the question of how to compute a block-separable factorization of a matrix $\mathbf{A}$, where the low-rank factorization is based on an *interpolative decomposition.*

Example: Consider an $N \times N$ matrix $\mathbf{A}$, and a partitioning of the index vector

$$I = \{1, 2, 3, \dots, N\} = I_4 \cup I_5 \cup I_6 \cup I_7.$$

We then seek to determine matrices $\{\mathbf{U}_\tau, \mathbf{V}_\tau\}_{\tau=4}^{7}$ and index vectors $\tilde{I}_\kappa \subset I_\kappa$ such that

$$\mathbf{A}(I_\tau, I_\sigma) = \mathbf{U}_\tau \tilde{\mathbf{A}}_{\tau,\sigma} \mathbf{V}_\sigma^*, \qquad \sigma \neq \tau,$$

where $\tilde{\mathbf{A}}_{\tau,\sigma} = \mathbf{A}(\tilde{I}_\tau, \tilde{I}_\sigma)$ is a submatrix of $\mathbf{A}_{\tau,\sigma}$.

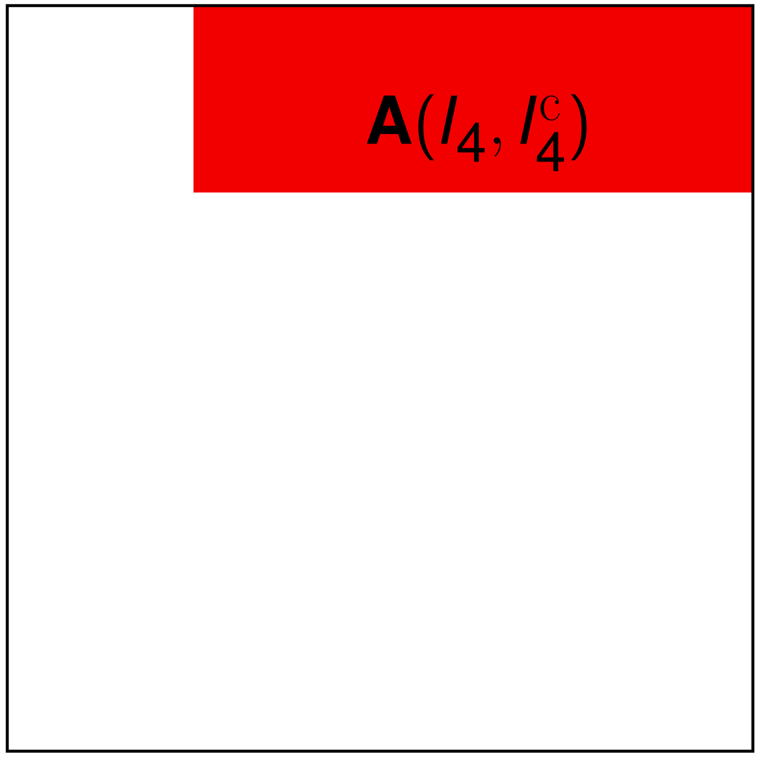
In other words, we seek a factorization

$$\mathbf{A} = \underbrace{\begin{bmatrix} \mathbf{U}_4 & & & \\ & \mathbf{U}_5 & & \\ & & \mathbf{U}_6 & \\ & & & \mathbf{U}_7 \end{bmatrix}}_{=\mathbf{U}} \underbrace{\begin{bmatrix} \mathbf{0} & \tilde{\mathbf{A}}_{45} & \tilde{\mathbf{A}}_{46} & \tilde{\mathbf{A}}_{47} \\ \tilde{\mathbf{A}}_{54} & \mathbf{0} & \tilde{\mathbf{A}}_{56} & \tilde{\mathbf{A}}_{57} \\ \tilde{\mathbf{A}}_{64} & \tilde{\mathbf{A}}_{65} & \mathbf{0} & \tilde{\mathbf{A}}_{67} \\ \tilde{\mathbf{A}}_{74} & \tilde{\mathbf{A}}_{75} & \tilde{\mathbf{A}}_{76} & \mathbf{0} \end{bmatrix}}_{=\tilde{\mathbf{A}}} \underbrace{\begin{bmatrix} \mathbf{V}_4^* & & & \\ & \mathbf{V}_5^* & & \\ & & \mathbf{V}_6^* & \\ & & & \mathbf{V}_7^* \end{bmatrix}}_{=\mathbf{V}^*} + \underbrace{\begin{bmatrix} \mathbf{D}_4 & & & \\ & \mathbf{D}_5 & & \\ & & \mathbf{D}_6 & \\ & & & \mathbf{D}_7 \end{bmatrix}}_{=\mathbf{D}}.$$

What is the role of the basis matrices $\mathbf{U}_\tau$ and $\mathbf{V}_\tau$?

Recall our toy example:
$$\mathbf{A} = \begin{bmatrix} \mathbf{D}_4 & \mathbf{U}_4\,\tilde{\mathbf{A}}_{45}\,\mathbf{V}_5^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{46}\,\mathbf{V}_6^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{47}\,\mathbf{V}_7^* \\ \mathbf{U}_5\,\tilde{\mathbf{A}}_{54}\,\mathbf{V}_4^* & \mathbf{D}_5 & \mathbf{U}_5\,\tilde{\mathbf{A}}_{56}\,\mathbf{V}_6^* & \mathbf{U}_5\,\tilde{\mathbf{A}}_{57}\,\mathbf{V}_7^* \\ \mathbf{U}_6\,\tilde{\mathbf{A}}_{64}\,\mathbf{V}_4^* & \mathbf{U}_6\,\tilde{\mathbf{A}}_{65}\,\mathbf{V}_5^* & \mathbf{D}_6 & \mathbf{U}_6\,\tilde{\mathbf{A}}_{67}\,\mathbf{V}_7^* \\ \mathbf{U}_7\,\tilde{\mathbf{A}}_{74}\,\mathbf{V}_4^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{75}\,\mathbf{V}_5^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{76}\,\mathbf{V}_6^* & \mathbf{D}_7 \end{bmatrix}.$$

We see that the columns of $\mathbf{U}_4$ must span the column space of the matrix $\mathbf{A}(I_4, I_4^{\mathrm{c}})$ where I_4 is the index vector for the first block and $I_4^{\mathrm{c}} = I\backslash I_4$.

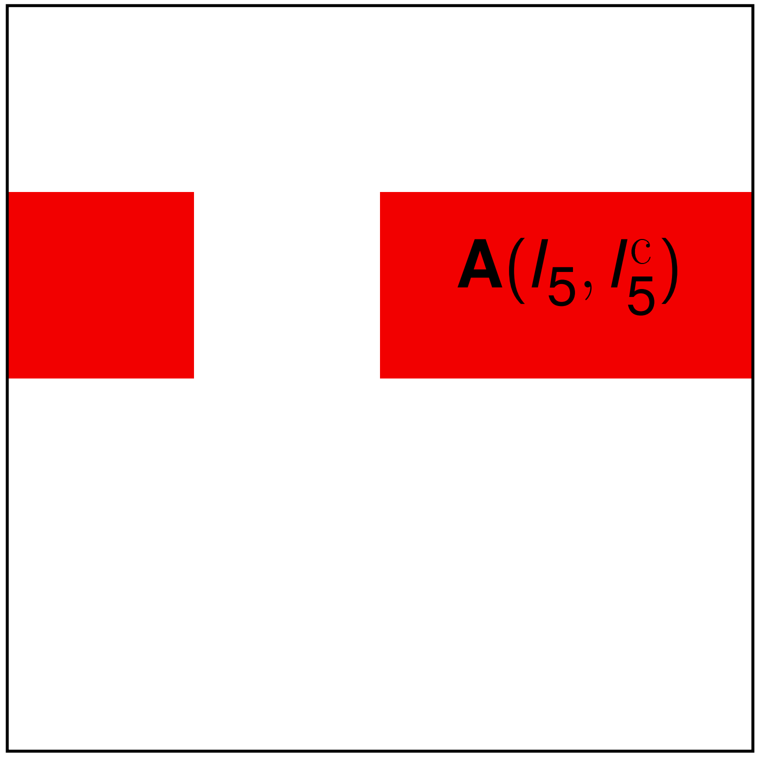


The matrix $\mathbf{A}$

What is the role of the basis matrices $\mathbf{U}_\tau$ and $\mathbf{V}_\tau$?

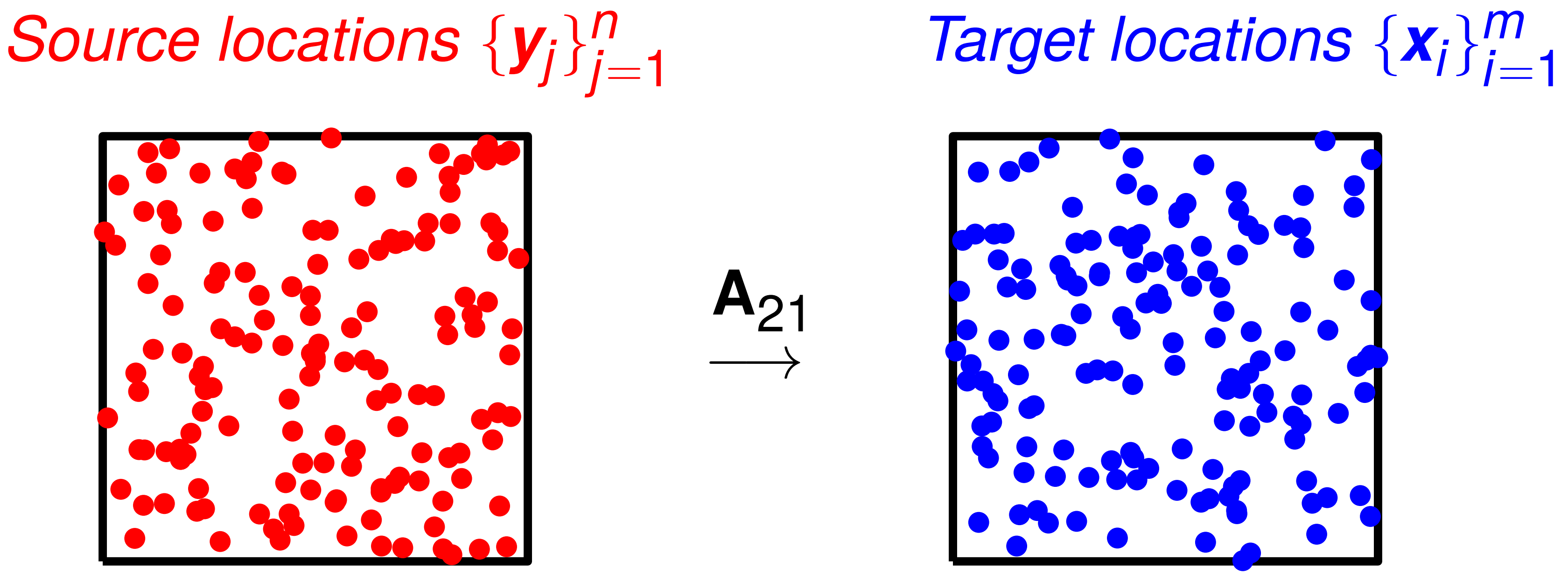
Recall our toy example:
$$\mathbf{A} = \begin{bmatrix} \mathbf{D}_4 & \mathbf{U}_4\,\tilde{\mathbf{A}}_{45}\,\mathbf{V}_5^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{46}\,\mathbf{V}_6^* & \mathbf{U}_4\,\tilde{\mathbf{A}}_{47}\,\mathbf{V}_7^* \\ \mathbf{U}_5\,\tilde{\mathbf{A}}_{54}\,\mathbf{V}_4^* & \mathbf{D}_5 & \mathbf{U}_5\,\tilde{\mathbf{A}}_{56}\,\mathbf{V}_6^* & \mathbf{U}_5\,\tilde{\mathbf{A}}_{57}\,\mathbf{V}_7^* \\ \mathbf{U}_6\,\tilde{\mathbf{A}}_{64}\,\mathbf{V}_4^* & \mathbf{U}_6\,\tilde{\mathbf{A}}_{65}\,\mathbf{V}_5^* & \mathbf{D}_6 & \mathbf{U}_6\,\tilde{\mathbf{A}}_{67}\,\mathbf{V}_7^* \\ \mathbf{U}_7\,\tilde{\mathbf{A}}_{74}\,\mathbf{V}_4^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{75}\,\mathbf{V}_5^* & \mathbf{U}_7\,\tilde{\mathbf{A}}_{76}\,\mathbf{V}_6^* & \mathbf{D}_7 \end{bmatrix}.$$

We see that the columns of $\mathbf{U}_5$ must span the column space of the matrix $\mathbf{A}(I_5, I_5^{\rm c})$ where I_5 is the index vector for the first block and $I_5^{\rm c} = I \backslash I_5$.



The matrix **A**

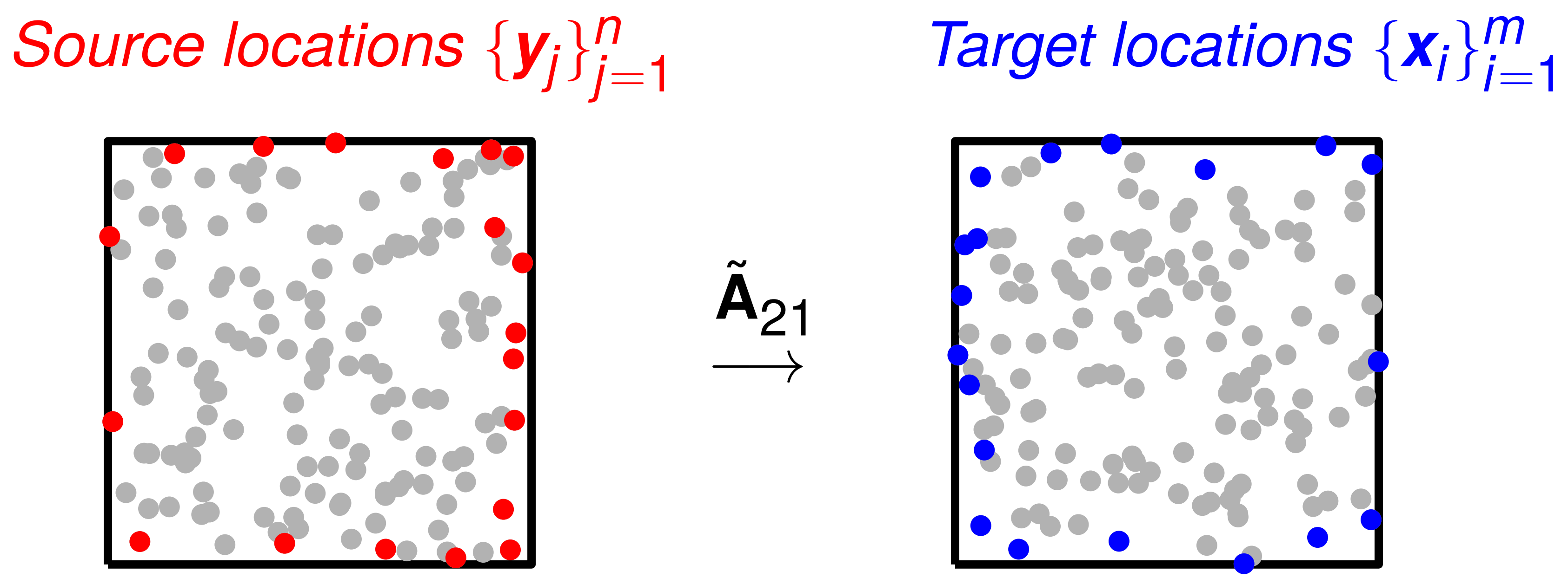
As mentioned earlier, it is handy to use the *interpolative decomposition (ID)*, in which $\mathbf{U}_\tau$ and $\mathbf{V}_\tau$ contain identity matrices. To review how this works, consider a situation with n sources in a domain Ω_1 inducing m potentials in a different domain Ω_2.



Let $\mathbf{A}_{21}$ denote the $m \times n$ matrix with entries $\mathbf{A}_{21}(i,j) = \log|\boldsymbol{x}_i - \boldsymbol{y}_j|$. Then

$$\underset{m \times 1}{\mathbf{f}} = \underset{m \times n}{\mathbf{A}_{21}} \; \underset{n \times 1}{\mathbf{q}}$$

As mentioned earlier, it is handy to use the *interpolative decomposition (ID)*, in which $\mathbf{U}_\tau$ and $\mathbf{V}_\tau$ contain identity matrices. To review how this works, consider a situation with n sources in a domain Ω_1 inducing m potentials in a different domain Ω_2.



Let $\mathbf{A}_{21}$ denote the $m \times n$ matrix with entries $\mathbf{A}_{21}(i,j) = \log|\boldsymbol{x}_i - \boldsymbol{y}_j|$. Then

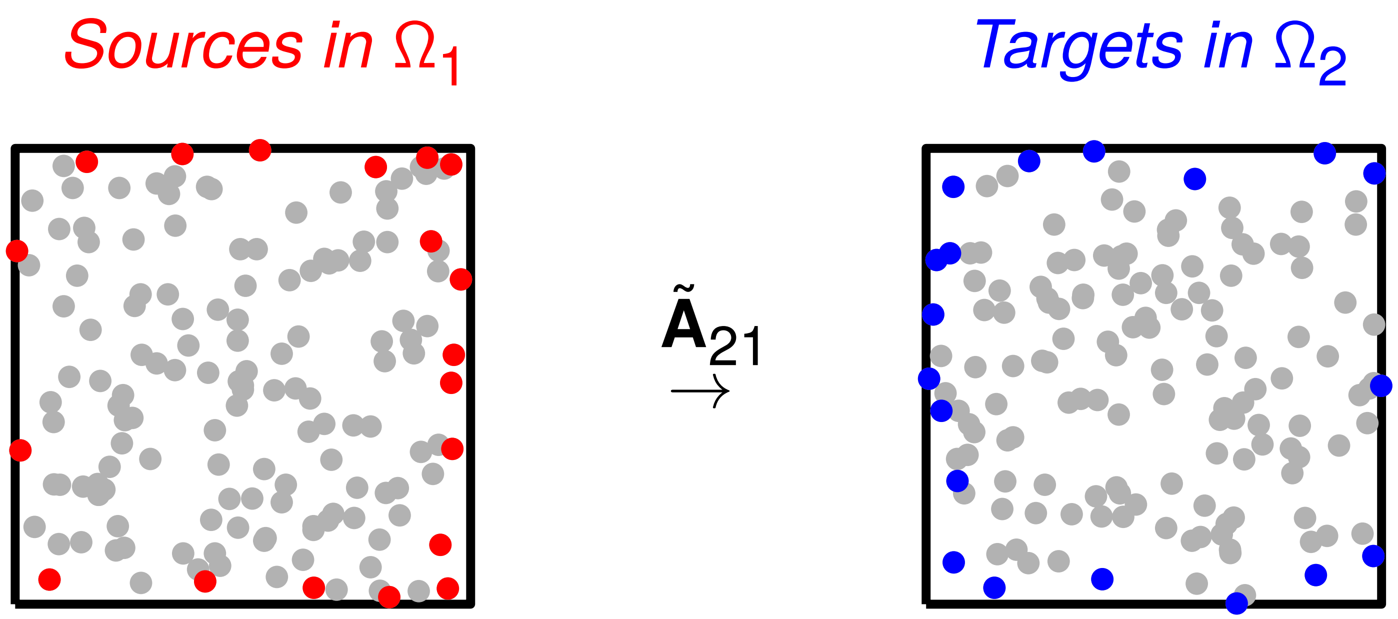
$$\underset{m\times 1}{\mathbf{f}} = \underset{m\times n}{\mathbf{A}_{21}}\ \underset{n\times 1}{\mathbf{q}} = \underset{m\times k}{\mathbf{U}_2}\ \underset{k\times k}{\tilde{\mathbf{A}}_{21}}\ \underset{k\times n}{\mathbf{V}_1^*}\ \underset{n\times 1}{\mathbf{q}}$$

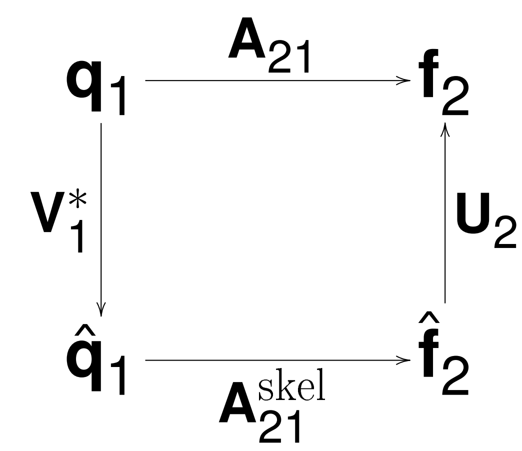
where $\tilde{\mathbf{A}}_{21} = \mathbf{A}_{21}(\tilde{I}_2, \tilde{I}_1)$ is a $k \times k$ submatrix of $\mathbf{A}$.

The index vector $\tilde{I}_1 \subseteq \{1, 2, \dots, n\}$ marks the chosen *skeleton source locations.*

The index vector $\tilde{I}_2 \subseteq \{1, 2, \dots, m\}$ marks the chosen *skeleton target locations.*

Review of ID: Consider a rank-k factorization of an $m \times n$ matrix: $\mathbf{A}_{21} = \mathbf{U}_2\, \tilde{\mathbf{A}}_{21}\, \mathbf{V}_1^*$





To precision 10^{-10}*, the rank is* 19*.*

Advantages of the ID:

- The rank k is typically close to optimal.
- Applying $\mathbf{V}_1^*$ and $\mathbf{U}_2$ is cheap — they both contain $k \times k$ identity matrices.
- The matrices $\mathbf{V}_1^*$ and $\mathbf{U}_2$ are well-conditioned.
- Finding the k points is cheap — simply use Gaussian elimination.
- The map $\tilde{\mathbf{A}}_{12}$ is simply a restriction of the original map $\mathbf{A}_{12}$.
  (We loosely say that "the physics of the problem is preserved".)
- Interaction between **adjacent** boxes can be compressed (no buffering required).

When the ID is used to compress the off-diagonal blocks, then all “black” blocks in the graphic below are *unchanged* compared to the original matrix. All you do is extract sub-blocks of the original off-diagonal blocks!