The exact fine structure of the inverse of discrete elliptic operators

S. Chandrasekaran — ECE, UCSB — *Email:* `shiv@ucsb.edu`

N. Govindarajan — Mech, UCSB

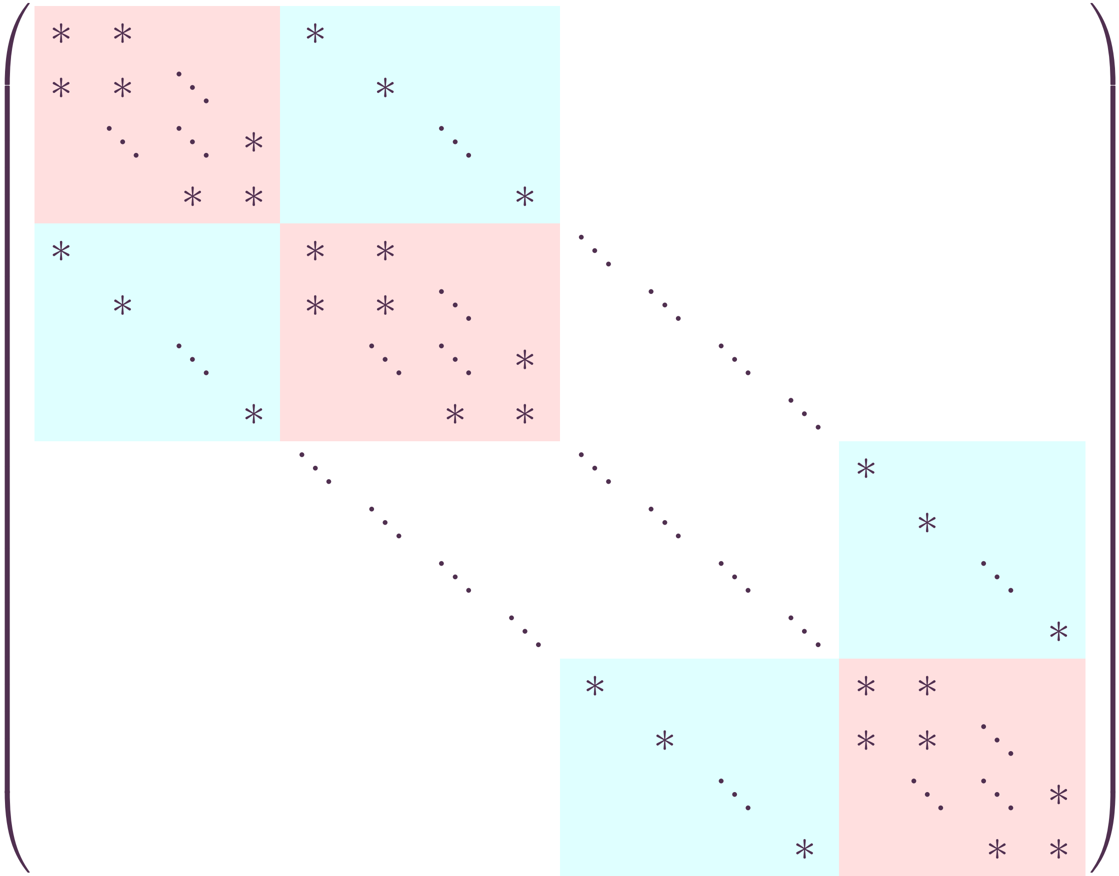
May 22, 2018

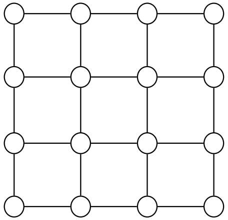
May 1, 2018

SIAM ALA '18 — Hong Kong

The problem

- What is the **exact** structure of the **inverse** of the following matrix?



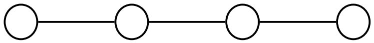


SSS representation

- The answer is well-known in the 1D case:

$$\begin{pmatrix} * & * & \\ * & \ddots & \ddots \\ & \ddots & * \end{pmatrix}^{-1} = \begin{pmatrix} D_1 & P_1Q_2^T & P_1R_2Q_3^T & \cdots \\ U_2V_1^T & D_2 & P_2Q_3^T & \cdots \\ U_3W_2V_1^T & U_3V_2^T & \ddots & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

- The associated graph is:



- The inverse of SSS is also SSS with **same block sizes**:

$$\begin{pmatrix} D_1 & P_1Q_2^T & P_1R_2Q_3^T & \cdots \\ U_2V_1^T & D_2 & P_2Q_3^T & \cdots \\ U_3W_2V_1^T & U_3V_2^T & \ddots & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}^{-1} = \begin{pmatrix} \hat{D}_1 & \hat{P}_1\hat{Q}_2^T & \hat{P}_1\hat{R}_2\hat{Q}_3^T & \cdots \\ \hat{U}_2\hat{V}_1^T & \hat{D}_2 & \hat{P}_2\hat{Q}_3^T & \cdots \\ \hat{U}_3\hat{W}_2\hat{V}_1^T & \hat{U}_3\hat{V}_2^T & \ddots & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

Hankel blocks

- Key concept is the Hankel block low-rank factorization:

$$\begin{pmatrix} P_1R_2R_3Q_4^T & P_1R_2R_3R_4Q_5^T & \cdots \\ P_2R_3Q_4^T & P_2R_3R_4Q_5^T & \cdots \\ P_3Q_4^T & P_3R_4Q_5^T & \cdots \end{pmatrix} = \begin{pmatrix} P_1R_2R_3 \\ P_2R_3 \\ P_3 \end{pmatrix} \begin{pmatrix} Q_4^T & R_4Q_5^T & R_4R_5Q_6^T & \cdots \end{pmatrix}$$

- Hankel blocks (H and G) are off diagonal blocks:

$$\begin{pmatrix} A_{m\times m} & H_{m\times n} \\ G_{n\times m} & B_{n\times n} \end{pmatrix}^{-1} = \begin{pmatrix} * & *H_{m\times n}* \\ *G_{n\times m}* & * \end{pmatrix}$$

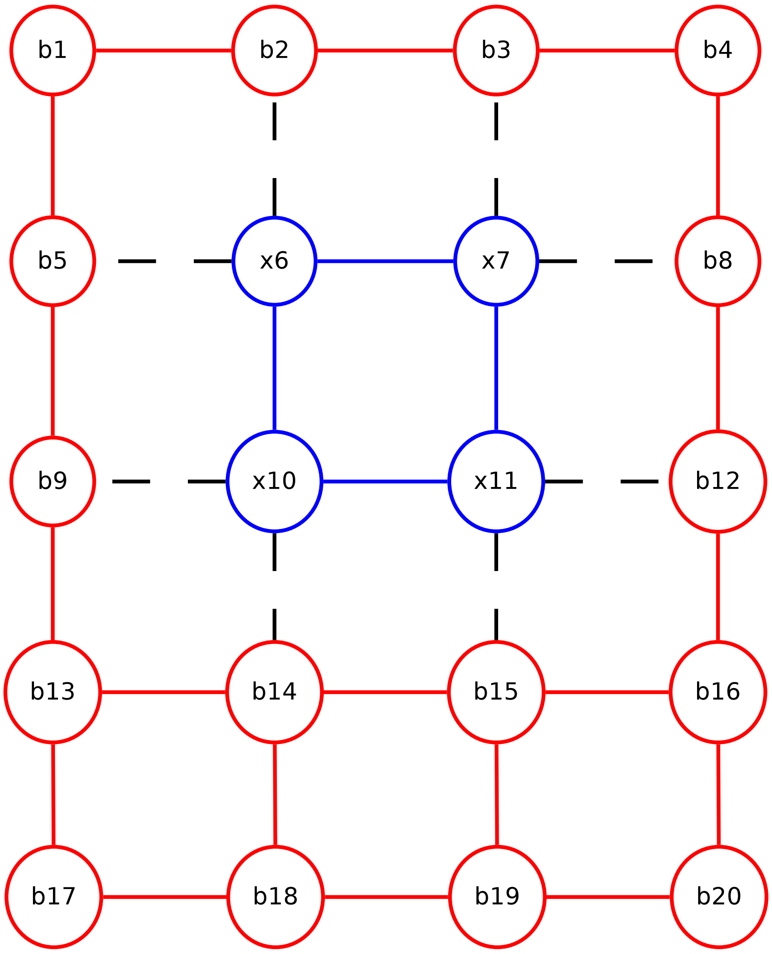
- So Hankel blocks in matrix and its inverse have same rank
- What is the right algebraic generalization if there is one?

- Associate a graph to a matrix:
  - Sparse matrices can be associated to their incidence graphs
  - SSS matrices will be associated with the linear graph
  - HSS matrices will be associated with the binary tree with edges between siblings
  - FMM matrices will be associated with their signal flow graph on the partition tree
- Let A be a matrix and $\mathbb{G}$ its associated graph:
  - A specific partition of A goes along with $\mathbb{G}$:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n21} & \cdots & A_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

    Each pair (x_i, b_i) is associated uniquely with a node of $\mathbb{G}$, and conversely.
  - Every **induced** sub-graph $\mathbb{H}$ of $\mathbb{G}$ induces a Hankel block in $P_1 A P_2$, where P_i are permutation matrices

- Example of $\mathbb{G}$ with $\mathbb{H}$ nodes in blue and $\overline{\mathbb{H}}$ nodes in red:



Graph induced rank structure

- For the previous graph $\mathbb{G}$, induced sub-graph $\mathbb{H}$ and its induced complement $\overline{\mathbb{H}}$, we get the induced ordering and partition of A:

$$\begin{array}{cc} & \begin{array}{cc} x_{\overline{\mathbb{H}}} & x_{\mathbb{H}} \end{array} \\ \begin{array}{c} b_{\overline{\mathbb{H}}} \\ b_{\mathbb{H}} \end{array} & \begin{pmatrix} A_{\overline{\mathbb{H}},\overline{\mathbb{H}}} & A_{\overline{\mathbb{H}},\mathbb{H}} \\ A_{\mathbb{H},\overline{\mathbb{H}}} & A_{\mathbb{H},\mathbb{H}} \end{pmatrix} \end{array} = P_1 A P_2$$

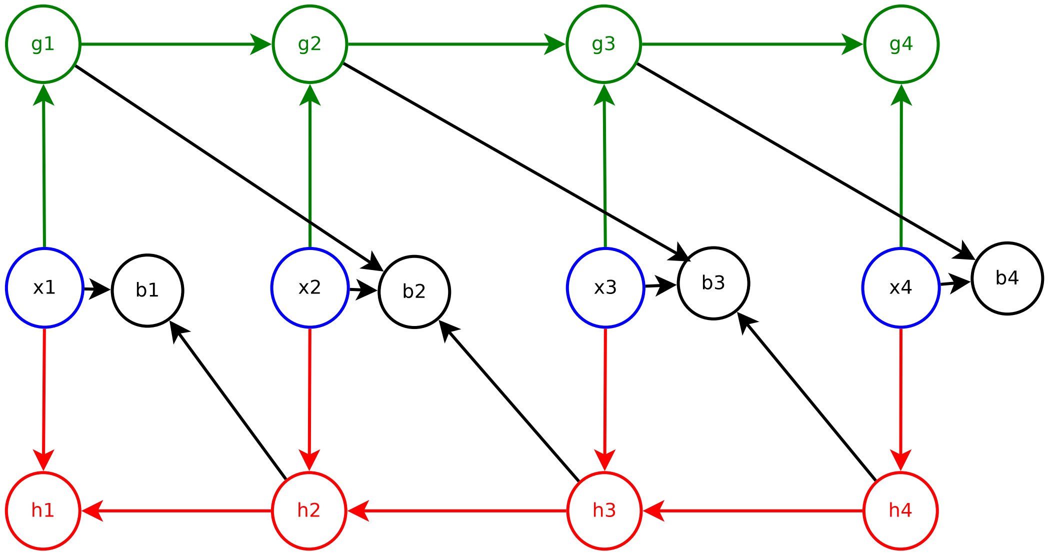
- We call $A_{\overline{\mathbb{H}},\mathbb{H}}$ the Hankel block induced by $\mathbb{H}$.
- A node N in $\mathbb{H}$ is called a **border node** if there is an edge in $\mathbb{G}$ from N to a node in $\overline{\mathbb{H}}$.
- The **border rank** of $\mathbb{H}$ is defined to be the number of border nodes in $\mathbb{H}$. We will denote this as $\rho(\mathbb{H})$.
- The pair $(A, \mathbb{G})$ is said have a **graph induced rank structure** if there is a constant c such that

$$\mathrm{rank}(A_{\overline{\mathbb{H}},\mathbb{H}}) \leqslant c\rho(\mathbb{H}) < M$$

for *all* induced sub-graphs $\mathbb{H}$ of $\mathbb{G}$, where M is the size of A.

- **Every sparse matrix** has the graph induced rank structure (GIRS) property with its incidence graph.
  - Proof is trivial from the non-zero structure
- The **inverse of every sparse matrix** has the GIRS property with the same incidence graph even though it is usually a dense matrix.
  - Proof is trivial from the Hankel block property
- The **product** of 2 GIRS matrices (using the same graph $\mathbb{G}$) is GIRS on $\mathbb{G}$
  - Proof is via Hankel blocks, but ranks add up
- The **SSS** representation is a special case of GIRS with a **simple linear graph**.
  - Proof follows from realizing that the signal flow graph is the associated graph in this case
  - The ranks of the induced Hankel blocks for the induced subgraph $\mathbb{H}$ are determined by the state-space variables in the border nodes of $\mathbb{H}$ and $\overline{\mathbb{H}}$

- The signal flow graph is essentially the same as the incidence graph for banded matrices



-

$$
\begin{aligned}
g_i &= V_i^T x_i + W_i g_{i-1} \\
h_i &= Q_i^T x_i + R_i h_{i+1} \\
b_i &= D_i x_i + U_i g_{i-1} + P_i h_{i+1}
\end{aligned}
$$

Diagonal representation of SSS

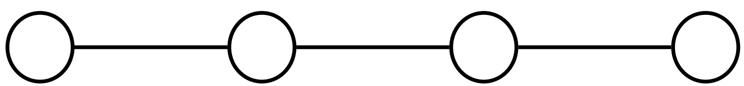
- The standard sparse Schur complement representation of SSS:

$$D = \mathrm{diag}\{D_i\}, \qquad U = \mathrm{diag}\{U_i\}, \qquad W = \mathrm{diag}\{W_i\}, \qquad V = \mathrm{diag}\{V_i\}, \qquad \text{etc.},$$

  then with Z as the shift down matrix:

$$\begin{pmatrix} I - WZ & & V^T \\ & I - RZ^T & Q^T \\ -U & -P & D \end{pmatrix} \begin{pmatrix} g \\ h \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ b \end{pmatrix},$$

  we get a sparse matrix with incidence graph (after **rearrangement**):



- The above matrix is for computing x from b. In the other direction:

$$\begin{pmatrix} I - WZ & & 0 \\ & I - RZ^T & 0 \\ -U & -P & -I \end{pmatrix} \begin{pmatrix} g \\ h \\ b \end{pmatrix} = \begin{pmatrix} -V^T x \\ -Q^T x \\ -x \end{pmatrix}.$$

- Whether we need to compute x from b or b from x, we just call a **sparse Gaussian elimination** code with the **right ordering** of the sparse matrix.
- Resulting complexity is **linear** in both cases as the incidence graph is linear, which has no fill-in.

- Is there an SSS like representation for arbitrary GIRS matrices?
- We introduce the **implicit** Dewilde representations:
  - To each node i of the graph $\mathbb{G}$ add the state space variable $g_i \in \mathbb{R}^{n_i}$ with n_i to be determined
  - Let $E(i)$ denote the edges of $\mathbb{G}$ that share an edge with node i
  - Define the state-space constraints

$$g_i = V_i^T x_i + \sum_{j \in E(i)} W_{i,j} g_j$$

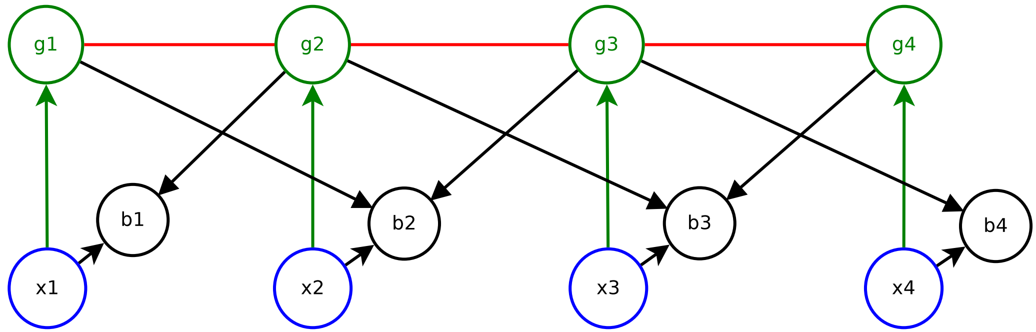
    where the weight matrices $W_{i,j}$ have to be determined
  - Define the outputs as (note that the **sums** are **identical**)

$$b_i = D_i x_i + \sum_{j \in E(i)} U_{i,j} g_j$$

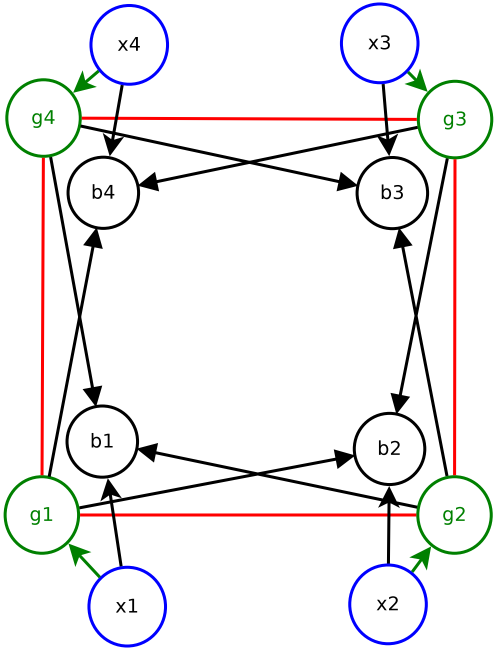
- Note that there is **no** explicit flow to compute the state-space variables from the inputs x_i

Examples of implicit Dewilde representations

- Line graph



- Circle graph



Sparse matrix for implicit Dewilde representation

- The recursions in diagonal form.
- Define the matrix-valued map $Z[\cdot]$ such that TFE:

$$
\begin{aligned}
g_i &= V_i^T x_i + \sum_{j \in E(i)} W_{i,j} g_j \\
g &= V^T x + Z[W]\, g
\end{aligned}
$$

- Then the sparse matrix representation is:

$$b = (D + Z[U](I - Z[W])^{-1} V^T) x$$

- Given an implicit Dewilde representation there are **linear** time algorithms to compute the implicit Dewilde representations of
  - products
  - inverses
  - sums
- Proof: Follows classical Dewilde and van der Veen theory when $Z[W] = WZ$ with Z as the shift down matrix

Inverse scattering

- So did we find an efficient representation for GIRS matrices?
  - Not yet
  - **Claim**: Implicit Dewilde does satisfy GIRS [not obvious at all]
- Computing an implicit Dewilde representation
  - Minimal representations are (too) easy for sparse matrices and hence their inverses etc.
  - Difficult to get **minimal** representations for general matrices (e.g. 2D integral operator and a mesh graph)
    - $\rightarrow$ Equivalent to the inverse scattering problem for a distributed dynamical system under equilibrium observations
    - $\rightarrow$ Unlike classical physics problems the internal state space dimensions are unknown
- **Claim**: Implicit Dewilde representation are universal (like SSS/HSS/FMM) as long as $\mathbb{G}$ is vertex-disjoint path connected
  - Proof is via explicit Dewilde representations which are easily universal as they include SSS as a special case
- **Question**: Does *every* GIRS matrix have an implicit Dewilde representation with edge ranks at most $c\gamma$, with γ a non-trivial constant?
- *Partial* answer: Sparse matrices and their inverses, etc., do.

Gauss's demon

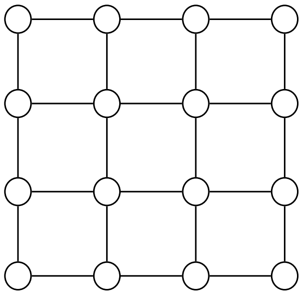
- Entering the club is difficult (unknown complexity in general)
- Exiting the club requires paying (the same) Gauss's price
  - Multiplying with a *dense* matrix requires **sparse Gaussian elimination** on $\mathbb{G}$
  - Multiplying *inverse* with a *dense* matrix requires **sparse Gaussian elimination** on $\mathbb{G}$
- It would be nice to pay Gauss's price only once
  - In SSS, multiplication is cheaper than inversion

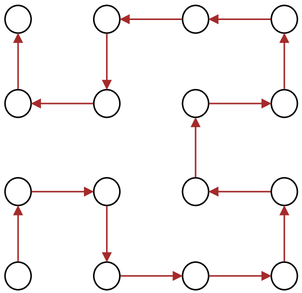
Explicit Dewilde representations

- Following Dewilde and van der Veen we seek explicit representations instead.
- What is an explicit representation?
  - The state-space variables can be partially ordered into a **causal** graph assuming input variables x_i are known
    - $\rightarrow$ So no Gaussian elimination needed for multiplication with a dense matrix
- Highly non-unique
- Desirable properties
  - The inverse scattering problem must be tractable
  - Implicit Dewilde representations must have short explicit Dewilde representations
  - Closure of explicit Dewilde representations under multiplication, inversion and additions (already true for implicit Dewilde representations)
- How to design one?
  - Canonical example is SSS

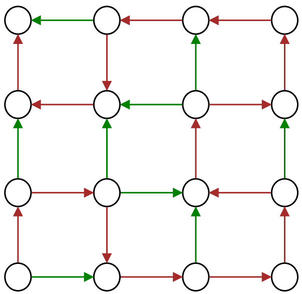
Example

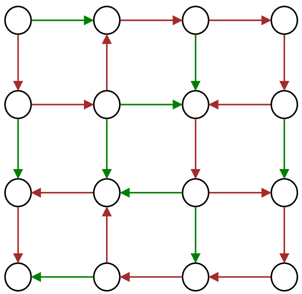
- Mesh graph $\rightarrow$ Directed vertex-disjoint path cover (1)





- Respect the induced order of the nodes for the remaining edges





- The dual graph is obtained by reversing the edges

- To each node of the first directed graph assign the state-space variables g_i
- To each node of the second dual graph assign the state-space variables h_i
- Let $P(i)$ denote the nodes of a directed graph that have a directed edge to node i
- Define the state-space dynamics

$$\begin{aligned}
g_i &= V_i^T x_i + \sum_{j \in P(i)} W_{i,j}\, g_j \\
h_i &= Q_i^T x_i + \sum_{j \in P(i)} R_{i,j}\, h_j \\
b_i &= D_i\, x_i + \sum_{j \in P(i)} U_{i,j}\, g_j + \sum_{j \in P(i)} P_{i,j}\, h_j
\end{aligned}$$

- As before define the matrix-valued operator $Z[\cdot]$:

$$\begin{aligned}
\sum_{j \in P(i)} W_{i,j}\, g_j &\equiv Z[W]g \\
\sum_{j \in P(i)} R_{i,j}\, h_j &\equiv Z^T[R]\, h
\end{aligned}$$

Sparse matrix for explicit Dewilde representation

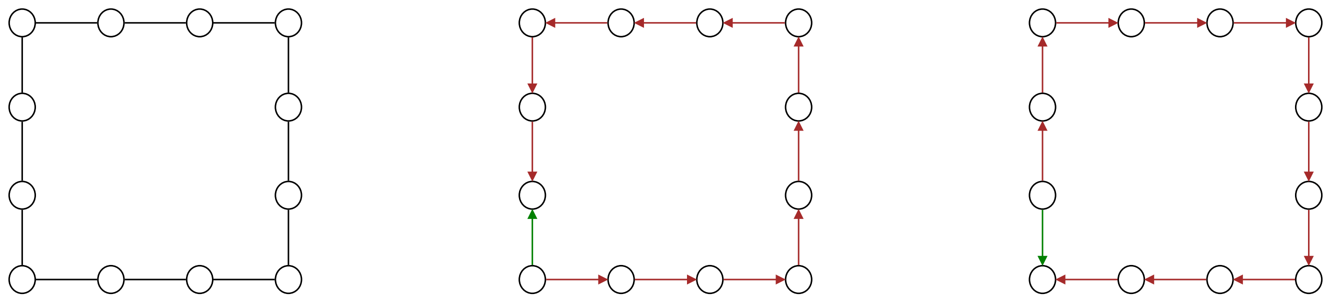
- We can write the equations as

$$b = (D + Z[U](I - Z[W])^{-1}V^T + Z^T[P](I - Z^T[R])^{-1}Q^T)x$$

- We can compute b given x in linear time using the causal graph structure
  - The second term is strictly lower triangular
  - The third term is strictly upper triangular
- We can compute x given b by **sparse Gaussian elimination**
  - Complexity determines on elimination ordering for $\mathbb{G}$
- Identifying the representation from the dense matrix (inverse scattering) seems tractbale based on the causal structure (nested low-rank completion plus SSS)
- **Claim**: Short explicit Dewilde representation implies existence of short implicit Dewilde representation
  - Proof is trivial
  - Note that $Z_{\text{implicit}}[] \equiv Z_{\text{explicit}}[] + Z^T_{\text{explicit}}[]$
- Question: Is the converse true?
  - Time for an example

Circular SSS

- Circle graph



- Associated canonical sparse matrix and its inverse

$$\begin{pmatrix} * & * & 0 & \cdots & 0 & {\color{red}*} \\ * & * & * & \ddots & & 0 \\ 0 & * & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & * & 0 \\ 0 & & \ddots & * & * & * \\ {\color{blue}*} & 0 & \cdots & 0 & * & * \end{pmatrix}^{-1}$$

- Applications: periodic boundary conditions

- Consider an inverse of the following type:

$$\left(\begin{pmatrix} * & & & \vdots & \\ & * & \cdots & pq^T & \cdots \\ & \vdots & * & \vdots & \\ \cdots & uv^T & \cdots & * & \\ & \vdots & & & * \end{pmatrix} + \begin{pmatrix} 0 & \cdots & \cdots & 0 & A \\ \vdots & \ddots & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \ddots & \vdots \\ B & 0 & \cdots & \cdots & 0 \end{pmatrix} \right)^{-1}$$

  where p, q, u, v are column vectors and A, B are block matrices.
- This is the inverse of an SSS plus block-rank-2 matrix
- It has the following type of structure

$$\begin{pmatrix} F & L & \boldsymbol{H} \\ L & F & L \\ \boldsymbol{H} & L & F \end{pmatrix}$$

  with the legend
  - F — full rank
  - L — low rank (but not rank 1)
  - $\boldsymbol{H}$ — comparable to rank of A and B

Is explicit Dewilde useful?

- Compared to the implicit Dewilde representation the explicit Dewilde representation will have higher ranks
- But the ranks are still low and worth exploiting in this example
- Does it hold generally?
- Can we construct the explicit Dewilde representation in this case?
  - We follow an approach similar to SSS

Circle SSS

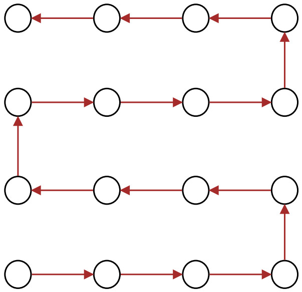
- Key formula

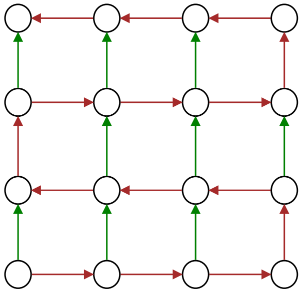
$$Z[U](I - Z[W])^{-1}V^T =$$

$$\begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & 0 \\ U_1V_0^T & 0 & \ddots & \cdots & \cdots & 0 \\ U_2W_1V_0^T & U_2V_1^T & \ddots & \ddots & & \vdots \\ U_3W_2W_1V_0^T & U_3W_2V_1^T & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & 0 \\ U_nW_{n-1}W_{n-2}\cdots W_2W_1V_0^T + U_0V_0^T & \cdots & * & U_nW_{n-1}V_{n-2}^T & U_nV_{n-1}^T & 0 \end{pmatrix}$$

- Note that other than the bottom left corner this is just the SSS representation
- Note that U_0 occurs uniquely and linearly
- Choose U_0 such that it minimizes the ranks of *all* Hankel blocks (using a suitable measure)
  - This is *not* the standard low-rank matrix completion problem
- Then identify the rest of the representation using the standard SSS technique
- This is reminiscent of the layer peeling approach in inverse scattering

- Consider the mesh:





- Note that if we set all weight matrices (translation operators) to 0 on the green edges then we are left with an SSS problem (which would be non-minimal)
- Therefore find a sequence of green edges such that their weights can be chosen to minimize the ranks of all other Hankel blocks.
  - You have to start at the end of the path and work backward; otherwise you will not get anything close to a minimal representation
  - Picking U_i and V_i to be column basis and row basis for block row and column Hankel blocks will not yield a representation (unlike FMM representation)
- Consider the 2×3 mesh explicitly:

- The strictly lower triangular part of the explicit Dewilde representation:

$$\begin{pmatrix} 0 & 0 \\ U_{2,1}V_1^T & 0 \\ U_{3,2}W_{2,1}V_1^T & U_{3,2}V_2^T \\ U_{4,3}W_{3,2}W_{2,1}V_1^T & U_{4,3}W_{3,2}V_2^T \\ ((U_{5,4}W_{4,3}W_{3,2}+U_{5,2})W_{2,1})V_1^T & (U_{5,4}W_{4,3}W_{3,2}+U_{5,2})V_2^T \\ (U_{6,5}(W_{5,4}W_{4,3}W_{3,2}+W_{5,2})W_{2,1}+U_{6,1})V_1^T & U_{6,5}(W_{5,4}W_{4,3}W_{3,2}+W_{5,2})V_2^T \end{pmatrix} \rightarrow$$

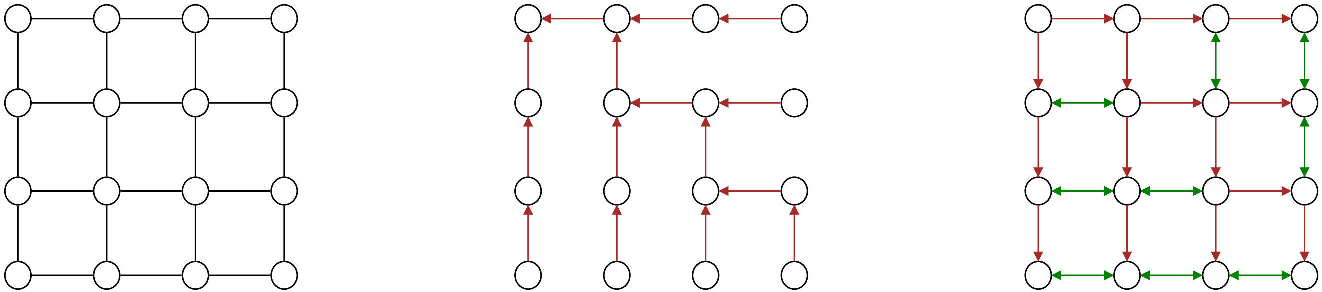
$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ U_{4,3}V_3^T & 0 & 0 \\ U_{5,4}W_{4,3}V_3^T & U_{5,4}V_4^T & 0 \\ U_{6,5}W_{5,4}W_{4,3}V_3^T & U_{6,5}W_{5,4}V_4^T & U_{6,5}V_5^T \end{pmatrix}$$

- First choose $U_{6,1}$ to minimize ranks of all lower Hankel blocks, freezing $U_{6,5}$ and V_1 to be full rank basis for the last block row and first block column respectively
- Next choose $W_{5,2}$, freezing $W_{2,1}$ and V_2 so as to obtain a full rank row basis for the join of the first two block columns
- Next choose $U_{5,2}$

Repeat

- The assumptions on $U_{6,5}$ and V_1 are sub-optimal
- So we can sweep through again with the same algorithm trying to get better compression
- Does the method converge?

- Rather than vertex-disjoint path covers we can also use tree covers



- Let $C(i)$ denote the set of child nodes of node i
- Let $S(i)$ denote the set of neighbors of node i, such that the corresponding edges are not in the tree
- The state-space dynamics (FMM recursions) are

$$\begin{aligned} g_i &= V_i^T x_i + \sum_{j \in C(i)} W_{i,j}\, g_j \\ h_i &= R_{C^{-1}(i)}\, h_{C^{-1}(i)} + \sum_{j \in S(i)} B_{i,j}\, g_j \\ b_i &= D_i\, x_i + U_i\, h_i \end{aligned}$$

- Define matrix valued maps $Z[\cdot]$ and $T[\cdot]$ such that TFE:

$$\begin{aligned} g_i &= V_i^T x_i + \sum_{j \in C(i)} W_{i,j}\, g_j \\ g &= V^T x + Z^T[W]\, g \\ h_i &= R_{C^{-1}(i)}\, h_{C^{-1}(i)} + \sum_{j \in S(i)} B_{i,j}\, g_j \\ h &= Z[R]\, h + T[B]\, g \end{aligned}$$

- Then

$$\begin{pmatrix} I - Z[W] & 0 & -V^T \\ -T[B] & I - Z[R] & 0 \\ 0 & U & D \end{pmatrix} \begin{pmatrix} g \\ h \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ b \end{pmatrix}$$

- So

$$b = (D + U\,(I - Z[R])^{-1} T[B] (I - Z^T[W])^{-1} V^T) x$$

- Note that for both Dewilde and Rokhlin representations, the Gaussian price is the *same* and determined by $\mathbb{G}$, except for changes in the edge ranks.

Diagonal representation

- In the HSS case $Z[\cdot]$ and $T[\cdot]$ are simple enough that a theory that matches the Dewilde and van der Veen approach can be developed
- However the algebra is harder to do in diagonal form as we have to deal with the commutators of $Z[\cdot]$, $Z^T[\cdot]$ and $T[\cdot]$
- The identification in the general case is harder than the standard FMM case as we again have additive terms and they have to be chosen to minimize ranks of many overlapping row and column Hankel blocks
- Such hierarchical representations with additive terms have been used before (Beylkin et. al.)

Open questions

- Does GIRS imply short implicit Dewilde representations?
  - Converse is true
  - Is there a (efficient) model reduction algorithm?
- Are there efficient inverse scattering algorithms for these representations?
- Does short implicit Dewilde imply short explicit Dewilde?
  - Product of 2 "circular tri-diagonal" matrices gives pause
- Product of 2 explicit Dewilde representations might be easier on $\mathbb{G}^2$ (already done for FMM). What does it imply for the inverse?
- The inverse scattering algorithm could be needed implicity in all the other algorithms
- Is it possible to have an efficient direct solver for sparse matrices that ignores GIRS property?
- Should surface integral operators in 3D problems use a 2D (toroidal) mesh Dewilde representation?