Geometric Sparsity and Mimetic Discretization

Lek-Heng Lim
Program in Scientific Computing/Computational Mathematics and the Department of Mathematics, Stanford University

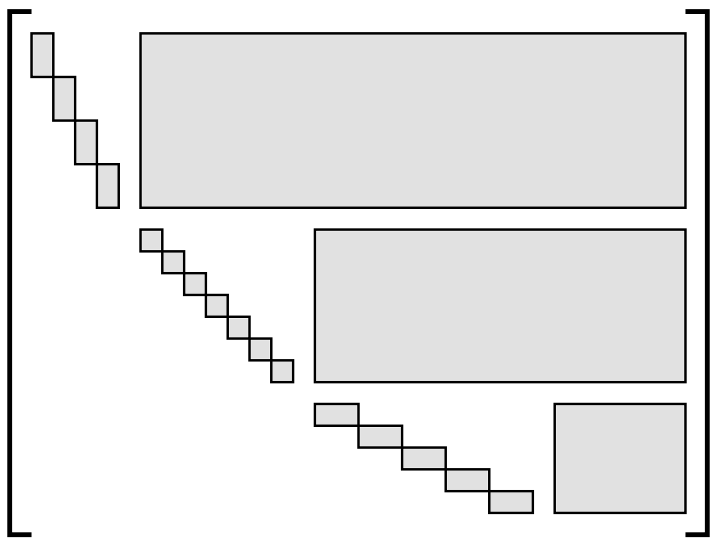
(Joint work with Gunnar Carlsson and Vin de Silva)

Thanks: Conference Organizers, NSF Grant DMS 01-01364, Gene Golub

Basic Idea I: Geometric Sparsity

Notion of geometric sparsity — accounting for sparse matrix representation of linear maps using underlying geometry of problem.

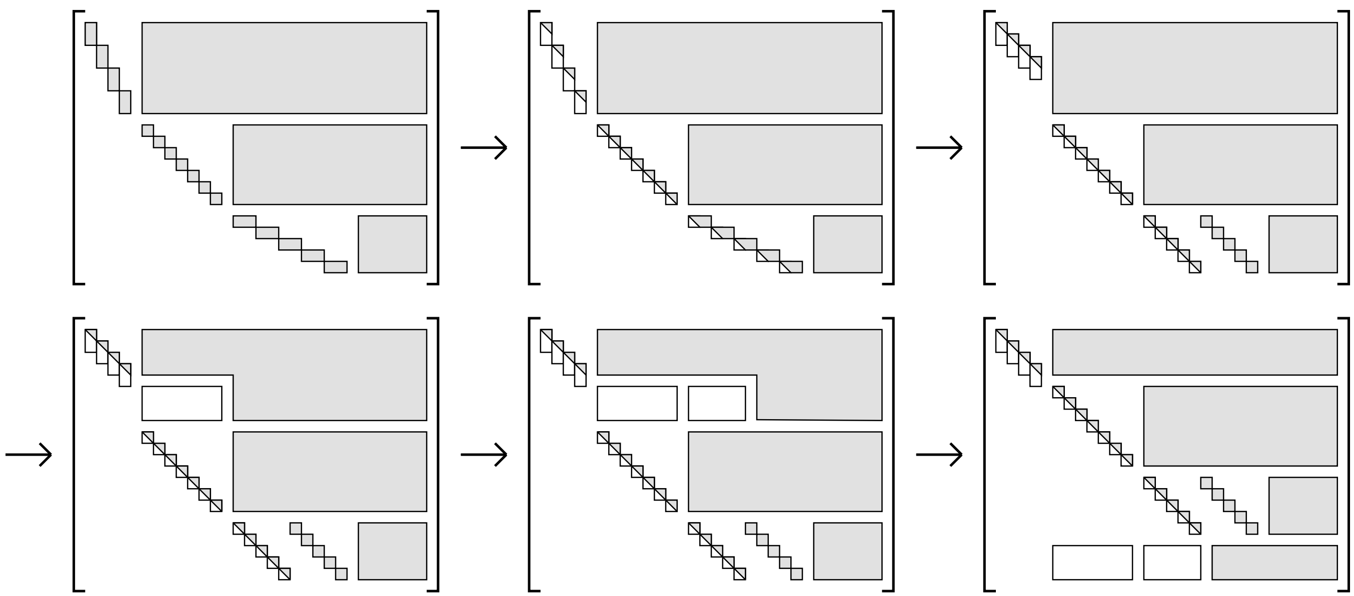
A *geometrically sparse* matrix with a scale parameter r_0 has the following form modulo row- and column-permutations:



Strongly filtered matrix: block triangular where the diagonal blocks are themselves block decomposed.

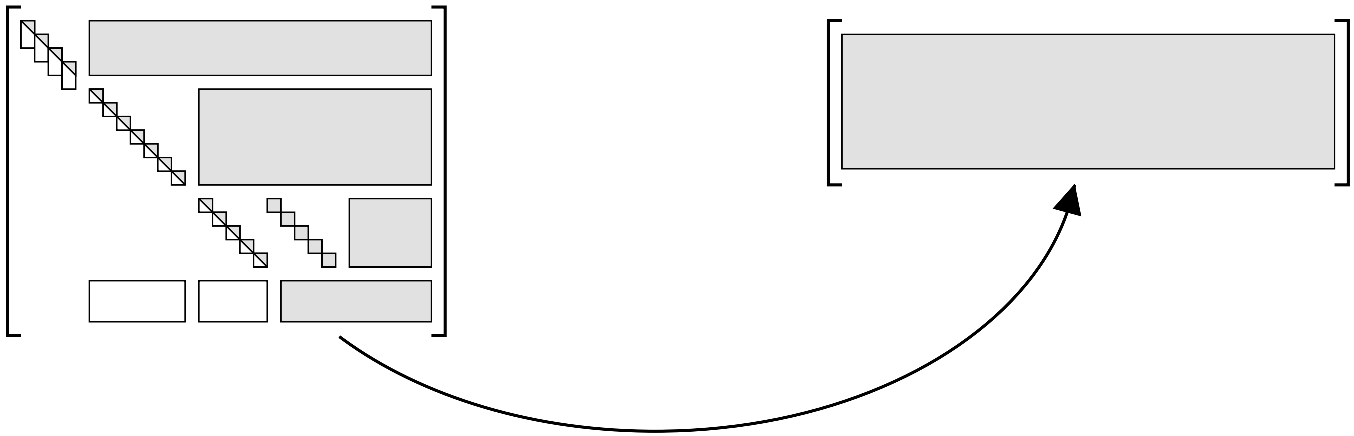
Basic Idea II: Decomposing Strongly Filtered Matrices

The block structure of strongly filtered matrices suggests a natural way of organizing the computation of LU and QR factors.



Basic Idea III: Multiscale Character

The 'remaining' submatrix on the bottom right is geometrically sparse with a larger scale parameter $r_1 > r_0$.



The same process may be repeated to this submatrix.

The process can in theory be repeated until the scale parameter is as large as the size of the bottom right submatrix.

Geometrically Sparse Matrices

'Definition' (Wilkinson). A *sparse matrix* is any matrix with enough zeros that it pays to take advantage of them.

Attempt to give a more concrete definition that accounts for how the sparseness arise.

Definition. A matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ is *geometrically sparse* with scale parameter r if there exist maps $\varphi : \{1, \ldots, m\} \to X$ and $\psi : \{1, \ldots, n\} \to X$ sending row and column indices of A into a metric space (X, d) so that $a_{ij} = 0$ whenever $d(\varphi(i), \psi(j)) > r$.

Speculation

Sparse matrices that arise from physical problems are naturally geometrically sparse or perturbations of geometrically sparse matrices.

The metric space (X, d) is suggested by the problem at hand.

Examples

$A \in \mathbb{R}^{m \times m}$ banded with bandwidth $2\ell + 1$:
$X = \mathbb{Z}$ or $\mathbb{R}$ with usual metric $|\cdot|$, $a_{ij} = 0$ if $|i - j| > \ell$.

$A = \text{diag}[A_1, \ldots, A_n] \in \mathbb{R}^{m \times m}$ block diagonal (with square blocks):
$X = \{1, \ldots, n\}$ with discrete metric δ, $a_{ij} = 0$ if $\delta(\varphi(i), \varphi(j)) > 0$.

$$A = \begin{bmatrix} \times & \times & & & \times \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \cdots & \cdots & \\ \times & & & \times & \times \end{bmatrix} \in \mathbb{R}^{m \times m}:$$

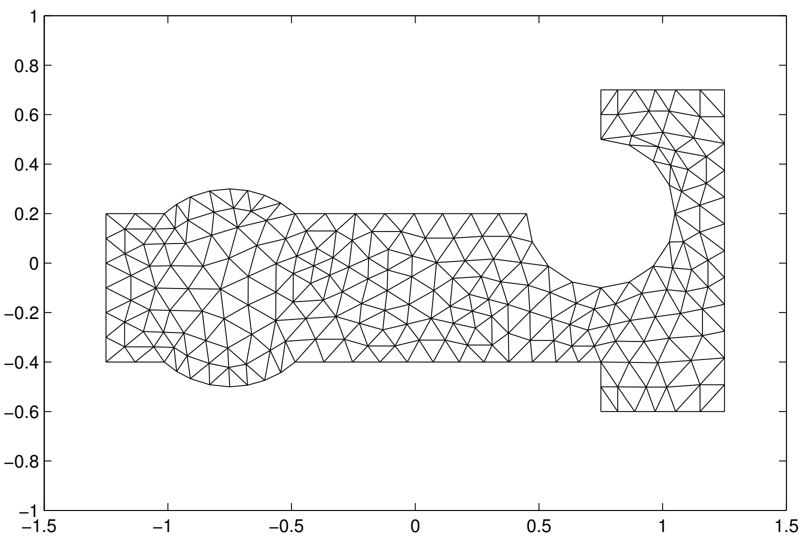
$X = S^1$ with usual (Riemannian) metric on circle d;
$\varphi : \{1, \ldots, m\} \to S^1$, $i \mapsto (\cos(2\pi i/m), \sin(2\pi i/m))$, $a_{ij} = 0$ if $d(\varphi(i), \varphi(j)) > 1$.

Examples: Numerical PDE

Finite Difference Methods: discrete approximations of partial differential operators are geometrically sparse.

$$\Delta : L^2(\mathbb{Z}) \to L^2(\mathbb{Z}), \quad \Delta f(i) = 2f(i) - [f(i-1) + f(i+1)],$$
$$\Delta : L^2(\mathbb{Z}^2) \to L^2(\mathbb{Z}^2), \quad \Delta f(i,j) = 4f(i,j) - [f(i+1,j) + f(i-1,j) + f(i,j+1) + f(i,j-1)].$$

Finite Element Methods: stiffness matrices in Galerkin's method are geometrically sparse.



Examples: Computational Topology

Σ simplicial complex embedded in $\mathbb{R}^n$. To compute $H_k(\Sigma)$, need to find null space of boundary map

$$\partial_k : C_k(\Sigma) \to C_{k-1}(\Sigma),$$
$$[v_0, \dots, v_k] \mapsto \sum_{i=0}^{k} (-1)^i [v_0, \dots, \widehat{v}_i, \dots, v_k].$$

Basis for $C_k(\Sigma)$: k-simplices;
basis for $C_{k-1}(\Sigma)$: $(k-1)$-simplices;
$X = \mathbb{R}^n$;
φ maps each $(k-1)$-simplex to its barycenter;
ψ maps each k-simplex to its barycenter;
$\ell =$ maximal diameter of any simplex of Σ.

Then the matrix representation of ∂_k is geometrically sparse with scale ℓ (likewise for the Laplacian $\Delta := \delta\partial + \partial\delta$).

An Illustration

Problem: given some domain of interest Ω; want to factorize, say, stiffness matrix A arising from an application of finite element method to an elliptic problem on Ω.

Traditional Approach: from A, construct graph $\mathcal{G}(A)$ from the non-zero pattern, sparsity is exploited in computations by examining the structure of $\mathcal{G}(A)$, e.g. graph-theoretic algorithms for minimizing fill-ins, algorithms of Dulmage-Mendelsohn and Pothen-Fan.

Our Approach: sparsity of A is a consequence of the geometry of Ω (slogan: sparsity arise because local changes have local effects), so we should use geometric information in Ω to organize our computation.

$\mathcal{G}(A)$ may capture some amount of geometric information in Ω but ought to be better to work with Ω (or an approximation of it) directly.

Čech Complex

X a space, $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ a (finite) covering of X.

Let $V(\mathcal{U}) = A$ and $\Sigma(\mathcal{U}) = \{\sigma = [\alpha_0, \ldots, \alpha_d] \mid U_{\alpha_0} \cap \cdots \cap U_{\alpha_d} \neq \varnothing\}$.

The *Čech complex*, $\check{\mathsf{C}}(\mathcal{U}) = (V(\mathcal{U}), \Sigma(\mathcal{U}))$, of the covering is an abstract simplicial complex:

$$\begin{array}{rcl} \text{vertex} & \longleftrightarrow & U_\alpha \neq \varnothing \\ \text{edge} & \longleftrightarrow & U_\alpha \cap U_\beta \neq \varnothing \\ & \vdots & \\ d\text{-simplex} & \longleftrightarrow & U_{\alpha_0} \cap \cdots \cap U_{\alpha_k} \neq \varnothing \end{array}$$

$\check{\mathsf{C}}(\mathcal{U})$ may be viewed as a topological/simplicial ‘approximation’ of X: for nice spaces X, may choose $\mathcal{U}$ so that $\check{\mathsf{C}}(\mathcal{U})$ is homotopy equivalent to X (in particular $H_*(X) \cong H_*(\check{\mathsf{C}}(\mathcal{U}))$.

Choice of Cover

$F : V \to W$ linear and geometrically sparse with parameter r.

V has basis $\{e_1, \ldots, e_n\}$ and W has basis $\{f_1, \ldots, f_m\}$ such that if

$$F(e_i) = \sum_{j=1}^{m} \alpha_{ji} f_j,$$

then $\alpha_{ji} = 0$ whenever

$$d(\varphi(i), \psi(j)) > r.$$

$\mathcal{U} = \{U_1, \ldots, U_N\}$ cover of X and $\mathcal{U}^r = \{U_1^r, \ldots, U_N^r\}$, $U_i^r = \{x \in X \mid d(x, U_i) < r\}$ 'thickened' cover.

For simplicity, choose $\mathcal{U}$ so that $\check{\mathrm{C}}(\mathcal{U}) \cong \check{\mathrm{C}}(\mathcal{U}^r)$.

We would also want to choose $\mathcal{U}$ so that $\check{\mathrm{C}}(\mathcal{U})$ correctly approximates X.

Permuting to Strongly Filtered Form

For each $\sigma = [\alpha_0, \ldots, \alpha_d] \in \check{\mathsf{C}}(\mathcal{U})$, define

$$V^\sigma := \{e_j \mid \psi(j) \in U_{\alpha_0} \cap \cdots \cap U_{\alpha_d},\ \psi(j) \notin U_{\beta_0} \cap \cdots \cap U_{\beta_{d+1}}\}$$
$$W^\sigma := \{f_i \mid \varphi(i) \in U^r_{\alpha_0} \cap \cdots \cap U^r_{\alpha_d},\ \varphi(i) \notin U^r_{\beta_0} \cap \cdots \cap U^r_{\beta_{d+1}}\}$$

and set

$$V^d := \bigoplus_{\dim(\sigma)=d} V^\sigma, \qquad W^d := \bigoplus_{\dim(\sigma)=d} W^\sigma.$$

Reordering $\{e_1, \ldots, e_n\}$ and $\{f_1, \ldots, f_m\}$ in decreasing $\dim(\sigma)$ while keeping the elements corresponding to a simplex together yields a strongly filtered matrix:

$$\begin{array}{c} \\ W^\ell \\ W^{\ell-1} \\ \vdots \\ W^1 \\ W^0 \end{array}
\begin{array}{c} \begin{array}{ccccc} V^\ell & V^{\ell-1} & V^{\ell-2} & \cdots & V^0 \end{array} \\
\begin{bmatrix} A_{\ell,\ell} & B_{\ell,\ell-1} & B_{\ell,\ell-2} & \cdots & B_{\ell,0} \\ & A_{\ell-1,\ell-1} & B_{\ell-1,\ell-2} & \cdots & B_{\ell-1,0} \\ & & \ddots & & \vdots \\ & & & A_{1,1} & B_{1,0} \\ & & & & A_{0,0} \end{bmatrix} \end{array}$$

where each $A_{d,d}$ is block diagonal:

$$\begin{array}{cc} & \begin{array}{ccc} V^{\sigma^1} & \cdots & V^{\sigma^{s(d)}} \end{array} \\ \begin{array}{c} W^{\sigma^1} \\ \vdots \\ W^{\sigma^{s(d)}} \end{array} & \left[\begin{array}{ccc} D_1^{(d)} & & \\ & \ddots & \\ & & D_{s(d)}^{(d)} \end{array} \right] \end{array},$$

$\sigma^1, \ldots, \sigma^{s(d)}$ being an exhaustive list of the d-dimensional simplices in $\check{\mathsf{C}}(\mathcal{U})$.

Organizing the Computation of LU Factors

Perform GECP on each $D_i^{(d)}$ (in parallel):

$$P_i^{(d)} D_i^{(d)} Q_i^{(d)} = \begin{bmatrix} L_i^{(d)} & 0 \\ T_i^{(d)} & I \end{bmatrix} \begin{bmatrix} U_i^{(d)} & S_i^{(d)} \\ 0 & 0 \end{bmatrix}$$

This yields the LU factors of $A_{d,d}$:

$$P_d A_{d,d} Q_d = \begin{bmatrix} L_1^{(d)} & & & 0 & & \\ & \ddots & & & \ddots & \\ & & L_{s_d}^{(d)} & & & 0 \\ T_1^{(d)} & & & I & & \\ & \ddots & & & \ddots & \\ & & T_{s_d}^{(d)} & & & I \end{bmatrix} \begin{bmatrix} U_1^{(d)} & & & S_1^{(d)} & & \\ & \ddots & & & \ddots & \\ & & U_{s_d}^{(d)} & & & S_{s_d}^{(d)} \\ 0 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & 0 & & & 0 \end{bmatrix}$$

$$=: \begin{bmatrix} L_d & 0 \\ T_d & I \end{bmatrix} \begin{bmatrix} U_d & S_d \\ 0 & 0 \end{bmatrix}$$

The off-diagonal blocks may be computed (in parallel) as

$$P_d B_{d,j} Q_j = \begin{bmatrix} L_d & 0 \\ T_d & I \end{bmatrix} \begin{bmatrix} C_{d,j}^{(1)} \\ C_{d,j}^{(0)} \end{bmatrix}$$

where

$$C_{d,j}^{(1)} = L_d^{-1} B_{d,j}^{(1)}, \qquad C_{d,j}^{(0)} = B_{d,j}^{(0)} - T_d L_d^{-1} B_{d,j}^{(1)}.$$

Further partition $C_{d,j}^{(1)}$ and $C_{d,j}^{(0)}$ into columns lying above U_d and columns lying above S_d:

$$\begin{bmatrix} C_{d,j}^{(1)} \\ C_{d,j}^{(0)} \end{bmatrix} = \begin{bmatrix} C_{d,j}^{(1,1)} & C_{d,j}^{(1,0)} \\ C_{d,j}^{(0,1)} & C_{d,j}^{(0,0)} \end{bmatrix}.$$

Doing the obvious column permutations yields

$$PAQ =$$

$$\begin{bmatrix} L_\ell & & & & & \\ & L_{\ell-1} & & & & \\ & & L_{\ell-2} & & & 0 \\ & & & \ddots & & \\ & & & & L_0 & \\ T_\ell & & & & & \\ & T_{\ell-1} & & & & \\ & & T_{\ell-2} & & & I \\ & & & \ddots & & \\ & & & & T_0 & \end{bmatrix} \begin{bmatrix} U_\ell & C^{(1,1)}_{\ell,\ell-1} & C^{(1,1)}_{\ell,\ell-2} & \cdots & C^{(1,1)}_{\ell,0} & S_\ell & C^{(1,0)}_{\ell,\ell-1} & C^{(1,0)}_{\ell,\ell-2} & \cdots & C^{(1,0)}_{\ell,0} \\ & U_{\ell-1} & C^{(1,1)}_{\ell-1,\ell-2} & \cdots & C^{(1,1)}_{\ell-1,0} & & S_{\ell-1} & C^{(1,0)}_{\ell-1,\ell-2} & \cdots & C^{(1,0)}_{\ell-1,0} \\ & & U_{\ell-2} & & \vdots & & & S_{\ell-2} & & \vdots \\ & & & \ddots & C^{(1,1)}_{1,0} & & & & \ddots & C^{(1,0)}_{1,0} \\ & & & & U_0 & & & & & S_0 \\ & C^{(0,1)}_{\ell,\ell-1} & C^{(0,1)}_{\ell,\ell-2} & \cdots & C^{(0,1)}_{\ell,0} & & C^{(0,0)}_{\ell,\ell-1} & C^{(0,0)}_{\ell,\ell-1} & \cdots & C^{(0,0)}_{\ell,0} \\ & & C^{(0,1)}_{\ell-1,\ell-2} & \cdots & C^{(0,1)}_{\ell,0} & & & C^{(0,0)}_{\ell-1,\ell-2} & \cdots & C^{(0,0)}_{\ell-1,0} \\ & & & \ddots & & & & & \ddots & \vdots \\ & & & & C^{(0,1)}_{0,0} & & & & & C^{(0,0)}_{1,0} \\ & & & & 0 & & & & & 0 \end{bmatrix}$$

Eliminate $C^{(0,1)}_*$ to get the following:

$$
PAQ = \begin{bmatrix}
L_\ell & & & & & \\
& L_{\ell-1} & & & & \\
& & L_{\ell-2} & & & 0 \\
& & & \ddots & & \\
& & & & L_0 & \\
T_\ell & C^{(0,1)}_{\ell,\ell-1} U^{-1}_{\ell-1} & C^{(0,1)}_{\ell,\ell-2} U^{-1}_{\ell-2} & & C^{(0,1)}_{\ell,0} U^{-1}_{0} & \\
& T_{\ell-1} & C^{(0,1)}_{\ell-1,\ell-2} U^{-1}_{\ell-2} & & C^{(0,1)}_{\ell-1,0} U^{-1}_{0} & \\
& & T_{\ell-2} & & & I \\
& & & \ddots & & \\
& & & & T_0 &
\end{bmatrix} \times
$$

$$
\begin{bmatrix}
U_\ell & C^{(1,1)}_{\ell,\ell-1} & C^{(1,1)}_{\ell,\ell-2} & \cdots & C^{(1,1)}_{\ell,0} & S_\ell & C^{(1,0)}_{\ell,\ell-1} & C^{(1,0)}_{\ell,\ell-2} & \cdots & C^{(1,0)}_{\ell,0} \\
& U_{\ell-1} & C^{(1,1)}_{\ell-1,\ell-2} & \cdots & C^{(1,1)}_{\ell-1,0} & & S_{\ell-1} & C^{(1,0)}_{\ell-1,\ell-2} & \cdots & C^{(1,0)}_{\ell-1,0} \\
& & U_{\ell-2} & & \vdots & & & S_{\ell-2} & & \vdots \\
& & & \ddots & C^{(1,1)}_{1,0} & & & & \ddots & C^{(1,0)}_{1,0} \\
& & & & U_0 & & & & & S_0 \\
& & & & & & G^{(0,0)}_{\ell,\ell-1} & G^{(0,0)}_{\ell,\ell-2} & \cdots & G^{(0,0)}_{\ell,0} \\
& & & & & & & G^{(0,0)}_{\ell-1,\ell-2} & \cdots & G^{(0,0)}_{\ell-1,0} \\
& & & & & & & & \ddots & \vdots \\
& & & & & & & & & G^{(0,0)}_{1,0} \\
& & & & & & & & & 0
\end{bmatrix}
$$

Multiscale Character

The lower right block

$$G^{(1)} := \begin{bmatrix} G^{(0,0)}_{\ell,\ell-1} & G^{(0,0)}_{\ell,\ell-2} & \cdots & G^{(0,0)}_{\ell,0} \\ & G^{(0,0)}_{\ell-1,\ell-2} & \cdots & G^{(0,0)}_{\ell-1,0} \\ & & \ddots & \vdots \\ & & & G^{(0,0)}_{1,0} \end{bmatrix}$$

where

$$G^{(0,0)}_{d+j,i} = C^{(0,0)}_{d+j,i} - C^{(0,1)}_{d+j,d} U_d^{-1} C^{(1,0)}_{d,i}$$

is geometrically sparse with scale parameter

$$r_1 := 2r + 4 \max_{\dim(\sigma)=d} \operatorname{diam}(U_{i_0} \cap \cdots \cap U_{i_d}).$$

The process may be repeated again. In theory, may obtain a sequence of geometrically sparse matrices of decreasing dimension $\{G^{(k)}\}$ with a corresponding increasing sequence of scale parameters:

$$r_k := 2r_{k-1} + 4 \max_{\mathrm{dim}(\sigma)=d-k+1} \mathrm{diam}(U_{i_0} \cap \cdots \cap U_{i_{d-k+1}}).$$

In practice, the geometric sparsity ceases to be useful when r_k becomes comparable in size with the dimension of $G^{(k)}$.

$\boxed{\textbf{Numerical Issues}}$

Method works fine when exact arithmetic is used. E.g. in computing $H_*(X; \mathbb{Z}/p\mathbb{Z})$, matrix decompositions over the $(\max, +)$-algebra or finite fields.

Numerically unstable when floating point operations are involved since pivots are not chosen dynamically as in GECP or GEPP.

A possible way to reduce the effect of numerical instability — Li-Demmel's static pivoting strategy.

Topological Origins of these Ideas

Geometric Sparsity — Bounded K-Theory

Organizing LU and QR Computations — Theory of Spectral Sequences

Multiscale Character — Novikov Conjecture (proof of special cases)

Divide-and-Conquer

Mayer-Vietoris. If a space X is broken up into two pieces $X = X_1 \cup X_2$, then its homology $H_*(X)$ can be calculated in terms of $H_*(X_1)$, $H_*(X_2)$ and $H_*(X_1 \cap X_2)$ via the long exact sequence

$$\cdots \to H_k(X_1 \cap X_2) \to H_k(X_1) \oplus H_k(X_2) \\ \to H_k(X) \to H_{k-1}(X_1 \cap X_2) \to \cdots$$

Spectral Sequence. Generalization of Mayer-Vietoris to $X = X_1 \cup \cdots \cup X_n$, any n.

A Spectral Sequence View

Recall previous notation. Set

$$V_d := \bigoplus_{\dim(\sigma) \geq d} V^\sigma, \qquad W_d := \bigoplus_{\dim(\sigma) \geq d} W^\sigma.$$

We have the following *filtration*:

$$V = V_0 \supseteq V_1 \supseteq \cdots \supseteq V_\ell = 0$$
$$W = W_0 \supseteq W_1 \supseteq \cdots \supseteq W_\ell = 0$$

A geometrically sparse matrix A is one that preserves this filtration, i.e.

$$A(V_d) \subseteq W_d.$$

The aforementioned technique for LU factorization is then motivated by the *spectral sequence of a filtration*.

Spectral Sequence Notation

V and W play the roles of two successive terms C_p and C_{p-1} in a filtered chain complex:

$$\cdots \xrightarrow{\partial_{p+2}} C_{p+1} \xrightarrow{\partial_{p+1}} C_p \xrightarrow{\partial_p} C_{p-1} \xrightarrow{\partial_{p-1}} \cdots$$

The filtration in question is

$$0 = C_{p,0} \subseteq C_{p,1} \subseteq \cdots \subseteq C_{p,\ell} = C_p$$

where $C_{p,q} = V_{\ell-q}$, $C_{p-1,q} = W_{\ell-q}$.

The geometrically sparse matrix A plays the role of the boundary map ∂_p.

We set

$$E^0_{p,q} := C_{p,q}/C_{p,q-1}, \qquad E^0_{p-1,q} := C_{p-1,q}/C_{p-1,q-1}.$$

Slogan

Every (geometrically) sparse matrix is a boundary map in a filtered complex that preserves the filtration:

$$\begin{array}{c} \\ E^0_{p-1,\ell} \\ E^0_{p-1,\ell-1} \\ \vdots \\ E^0_{p-1,1} \\ E^0_{p-1,0} \end{array} \begin{array}{c} \begin{array}{ccccc} E^0_{p,\ell} & E^0_{p,\ell-1} & E^0_{p,\ell-2} & \cdots & E^0_{p,0} \end{array} \\ \left[\begin{array}{ccccc} * & \times & \times & \cdots & \times \\ & * & \times & \cdots & \times \\ & & \ddots & & \vdots \\ & & & * & \times \\ & & & & * \end{array} \right] \end{array}$$

In earlier notation:

$$\begin{array}{c} \\ W_0/W_1 \\ W_1/W_2 \\ \vdots \\ W_{\ell-2}/W_{\ell-1} \\ W_{\ell-1}/W_\ell \end{array} \begin{array}{c} \begin{array}{ccccc} V_0/V_1 & V_1/V_2 & V_2/V_3 & \cdots & V_{\ell-1}/V_\ell \end{array} \\ \left[\begin{array}{ccccc} * & \times & \times & \cdots & \times \\ & * & \times & \cdots & \times \\ & & \ddots & & \vdots \\ & & & * & \times \\ & & & & * \end{array} \right] \end{array}$$

More is True

Note that

$$V_d/V_{d+1} = V^d = \bigoplus_{\dim(\sigma)=d} V^\sigma$$

$$W_d/W_{d+1} = W^d = \bigoplus_{\dim(\sigma)=d} W^\sigma$$

A geometrically sparse matrix preserves this grading, i.e.

$$A(V^\sigma) \subseteq W^\sigma.$$

That's why the diagonal blocks (marked $*$) are themselves blocked-decomposed

$$\begin{array}{cc} & \begin{array}{ccc} V^{\sigma^1} & \cdots & V^{\sigma^{s(d)}} \end{array} \\ \begin{array}{c} W^{\sigma^1} \\ \vdots \\ W^{\sigma^{s(d)}} \end{array} & \begin{bmatrix} \times & & \\ & \ddots & \\ & & \times \end{bmatrix}, \end{array}$$

References

G. Carlsson and V. de Silva, "A geometric framework for sparse matrix problems," preprint, March 11, 2002
(available from: `http://math.stanford.edu/comptop/preprints`)

V. de Silva, "Geometrically-motivated ideas for sparse matrix reduction," talk in *Conference on Algebraic Topological Methods in Computer Science*, Stanford, CA, July 30–August 3, 2001

L.-H. Lim, "A multiscale technique for decomposing sparse matrices motivated by spectral sequences," talk in *2003 SIAM Annual Meeting*, Montréal, QC, Canada, June 16–20, 2003