Chordal Graphs and Sparse Semidefinite Optimization

Lieven Vandenberghe

Electrical and Computer Engineering Department, UCLA

Joint work with Joachim Dahl, Martin S. Andersen, Yifan Sun

Workshop on Fast Iterative Methods in Optimization

Simons Institute for the Theory of Computing

October 2, 2017

Sparse semidefinite program (SDP)

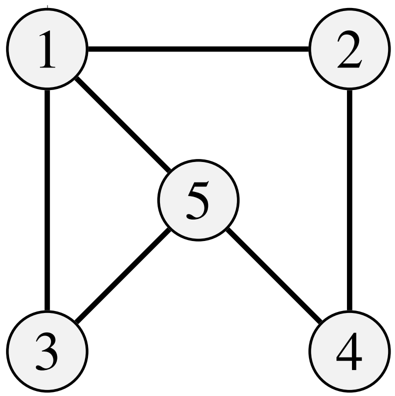
$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_iX) = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{array}$$

- variable X is $n \times n$ symmetric matrix; $X \succeq 0$ means X is positive semidefinite
- in many applications the coefficients A_i, C are sparse
- optimal X is typically dense, even for sparse A_i, C

Topic of the talk

- structure in solution X that results from sparsity in coefficients A_i, C
- results from graph and sparse matrix theory that are useful for SDP algorithms

Sparsity graph



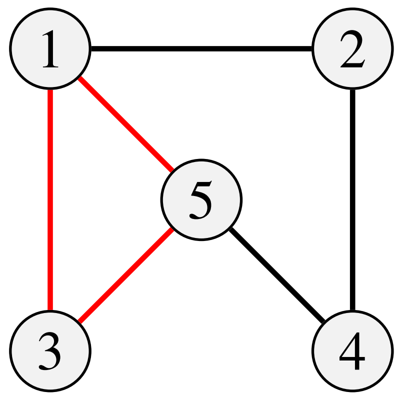
$$A = \begin{bmatrix} A_{11} & A_{21} & A_{31} & 0 & A_{51} \\ A_{21} & A_{22} & 0 & A_{42} & 0 \\ A_{31} & 0 & A_{33} & 0 & A_{53} \\ 0 & A_{42} & 0 & A_{44} & A_{54} \\ A_{51} & 0 & A_{53} & A_{54} & A_{55} \end{bmatrix}$$

- sparsity pattern of symmetric $n \times n$ matrix is set of 'nonzero' positions

$$E \subseteq \{\{i, j\} \mid i, j \in \{1, 2, \ldots, n\}\}$$

- A has sparsity pattern E if $A_{ij} = 0$ if $i \neq j$ and $\{i, j\} \notin E$
- notation: $A \in \mathbf{S}_E^n$
- represented by undirected graph (V, E) with edges E, vertices $V = \{1, \ldots, n\}$
- clique (maximal complete subgraph) forms maximal 'dense' principal submatrix

Sparsity graph



$$A = \begin{bmatrix} \mathbf{A_{11}} & A_{21} & \mathbf{A_{31}} & 0 & \mathbf{A_{51}} \\ A_{21} & A_{22} & 0 & A_{42} & 0 \\ \mathbf{A_{31}} & 0 & \mathbf{A_{33}} & 0 & \mathbf{A_{53}} \\ 0 & A_{42} & 0 & A_{44} & A_{54} \\ \mathbf{A_{51}} & 0 & \mathbf{A_{53}} & A_{54} & \mathbf{A_{55}} \end{bmatrix}$$

- sparsity pattern of symmetric $n \times n$ matrix is set of 'nonzero' positions

$$E \subseteq \{\{i, j\} \mid i, j \in \{1, 2, \dots, n\}\}$$

- A has sparsity pattern E if $A_{ij} = 0$ if $i \neq j$ and $\{i, j\} \notin E$
- notation: $A \in \mathbf{S}^n_E$
- represented by undirected graph (V, E) with edges E, vertices $V = \{1, \dots, n\}$
- clique (maximal complete subgraph) forms maximal 'dense' principal submatrix

Sparse matrix cones

we define two convex cones in $\mathbf{S}^n_E$ (symmetric $n \times n$ matrices with pattern E)

- positive semidefinite matrices

$$\mathbf{S}^n_+ \cap \mathbf{S}^n_E = \{X \in \mathbf{S}^n_E \mid X \succeq 0\}$$

- matrices with a positive semidefinite completion

$$\Pi_E(\mathbf{S}^n_+) = \{\Pi_E(X) \mid X \succeq 0\}$$

  Π_E is projection on $\mathbf{S}^n_E$

Properties

- two cones are convex
- closed, pointed, with nonempty interior (relative to $\mathbf{S}^n_E$)
- form a pair of dual cones (for the trace inner product)

Sparse semidefinite program

Standard form SDP and dual (variables $X, S \in \mathbf{S}^n$, $y \in \mathbf{R}^m$)

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_iX) = b_i, \ i = 1, \ldots, m \\ & X \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & \sum_{i=1}^m y_i A_i + S = C \\ & S \succeq 0 \end{array}$$

Equivalent pair of conic linear programs (variables $X, S \in \mathbf{S}^n_E$, $y \in \mathbf{R}^m$)

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_iX) = b_i, \ i = 1, \ldots, m \\ & X \in K \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & \sum_{i=1}^m y_i A_i + S = C \\ & S \in K^* \end{array}$$

- E is union of sparsity patterns of C, A_1, ..., A_m
- $K = \Pi_E(\mathbf{S}^n_+)$ is cone of p.s.d. completable matrices with sparsity pattern E
- $K^* = \mathbf{S}^n_+ \cap \mathbf{S}^n_E$ is cone of positive semidefinite matrices with sparsity pattern E

Outline

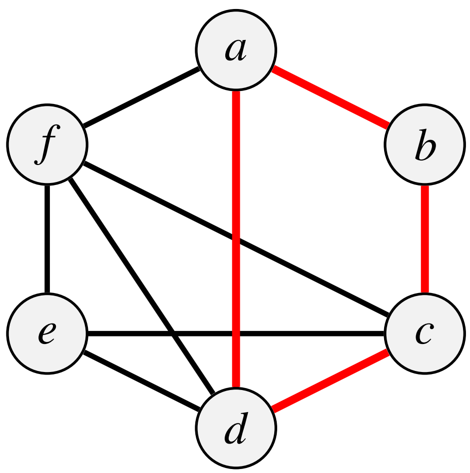
1. Chordal graphs
2. Decomposition of sparse matrix cones
3. Multifrontal algorithms for logarithmic barrier functions
4. Minimum rank positive semidefinite completion

Chordal graph

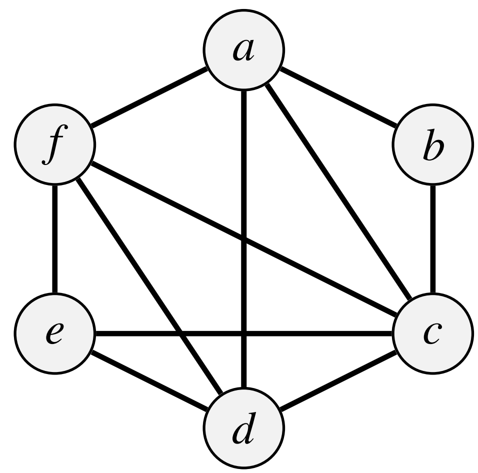
- undirected graph with vertex set V, edge set $E \subseteq \{\{v, w\} \mid v, w \in V\}$

$$G = (V, E)$$

- a **chord** of a cycle is an edge between non-consecutive vertices
- G is **chordal** if every cycle of length greater than three has a chord



not chordal



chordal

also known as triangulated, decomposable, rigid circuit graph, ...

History

chordal graphs have been studied in many disciplines since the 1960s

- combinatorial optimization (a class of *perfect* graphs)
- linear algebra (sparse factorization, completion problems)
- database theory
- machine learning (graphical models, probabilistic networks)
- nonlinear optimization (partial separability)

first used in semidefinite optimization by Fujisawa, Kojima, Nakata (1997)

Chordal sparsity and Cholesky factorization

Cholesky factorization of positive definite $A \in \mathbf{S}^n_E$:

$$PAP^T = LDL^T$$

P a permutation, L unit lower triangular, D positive diagonal

- if E is chordal, then there exists a permutation for which

$$P^T(L + L^T)P \in \mathbf{S}^n_E$$

  A has a 'zero fill' Cholesky factorization

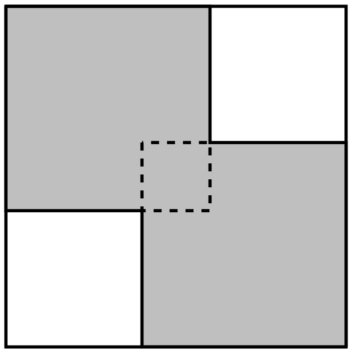
- if E is not chordal, then for every P there exist positive definite $A \in \mathbf{S}^n_E$ for which

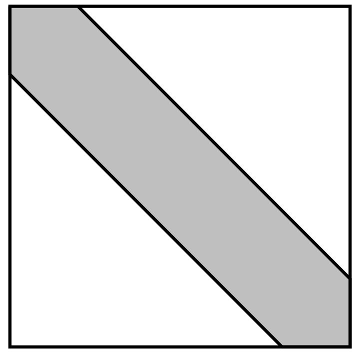
$$P^T(L + L^T)P \notin \mathbf{S}^n_E$$

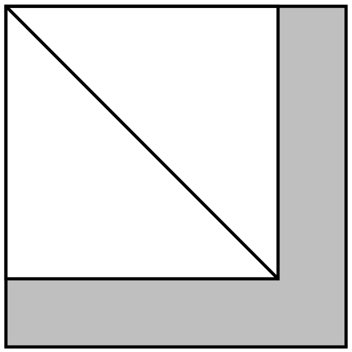
[Rose 1970]

Examples

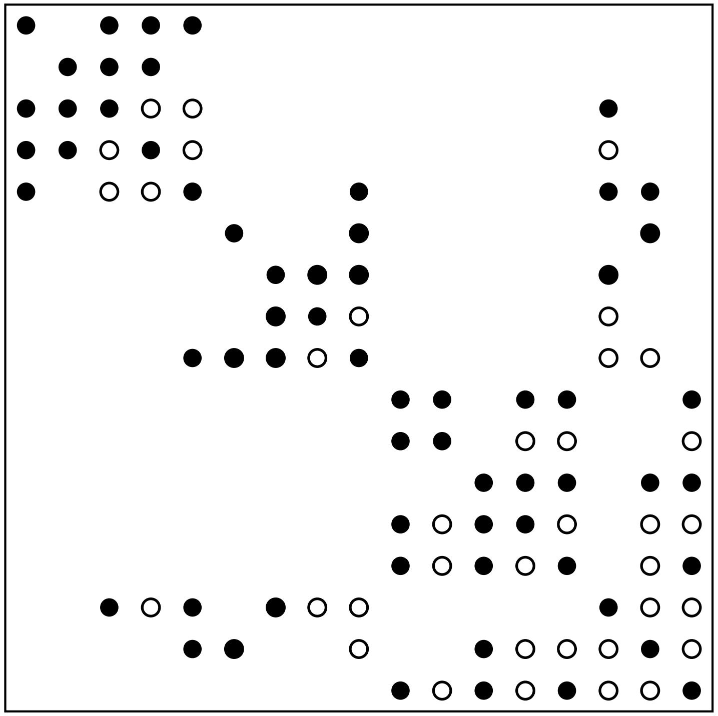
Simple patterns







Sparsity pattern of a Cholesky factor

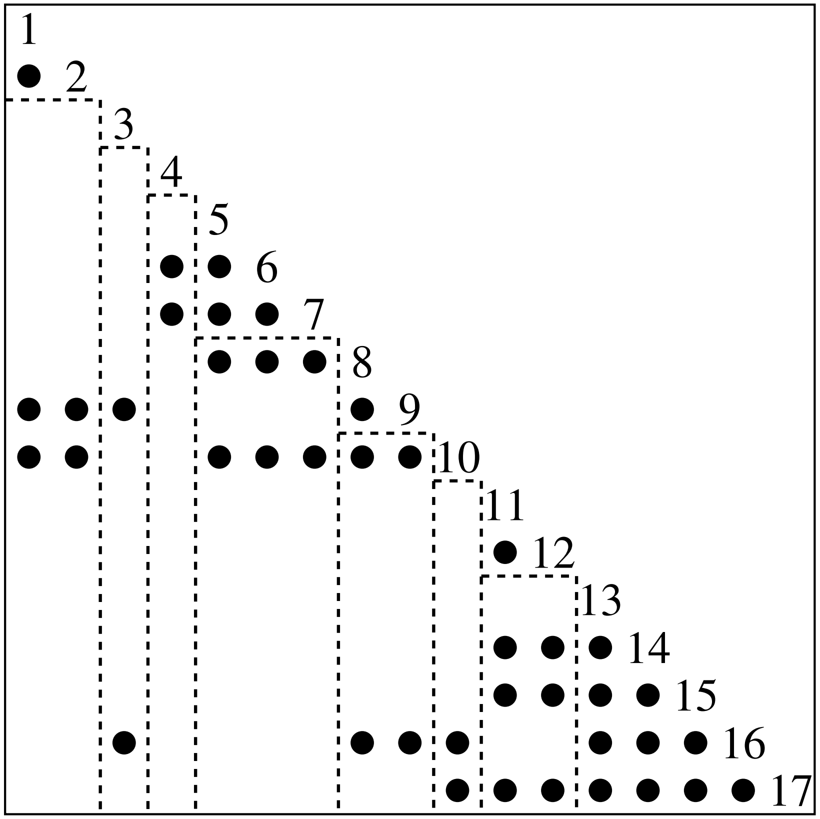


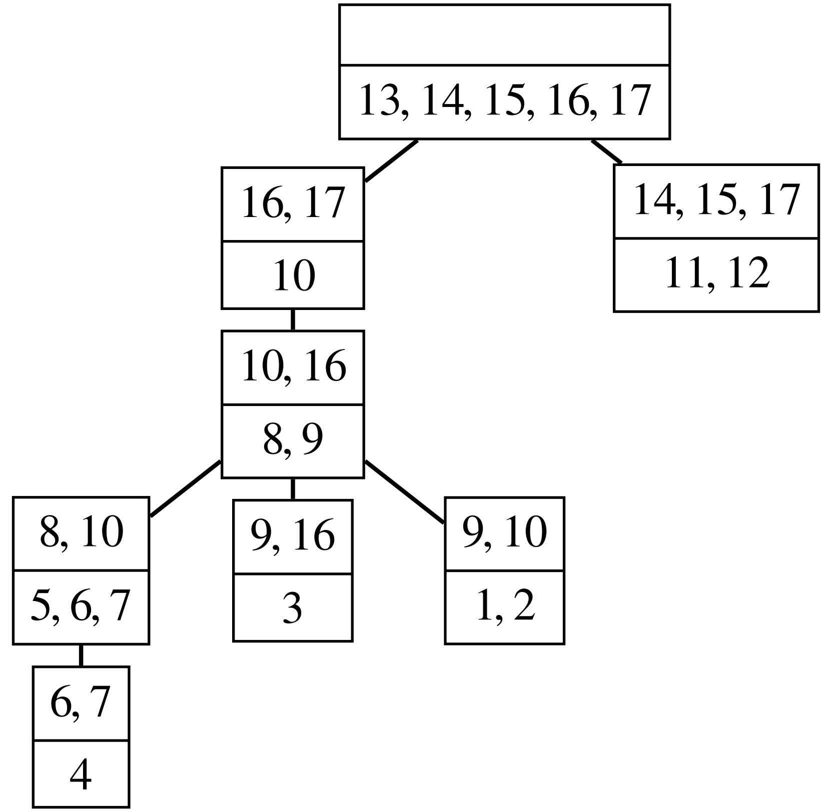
●: edges of non-chordal sparsity pattern

○: fill entries in Cholesky factorization

a chordal extension of non-chordal pattern

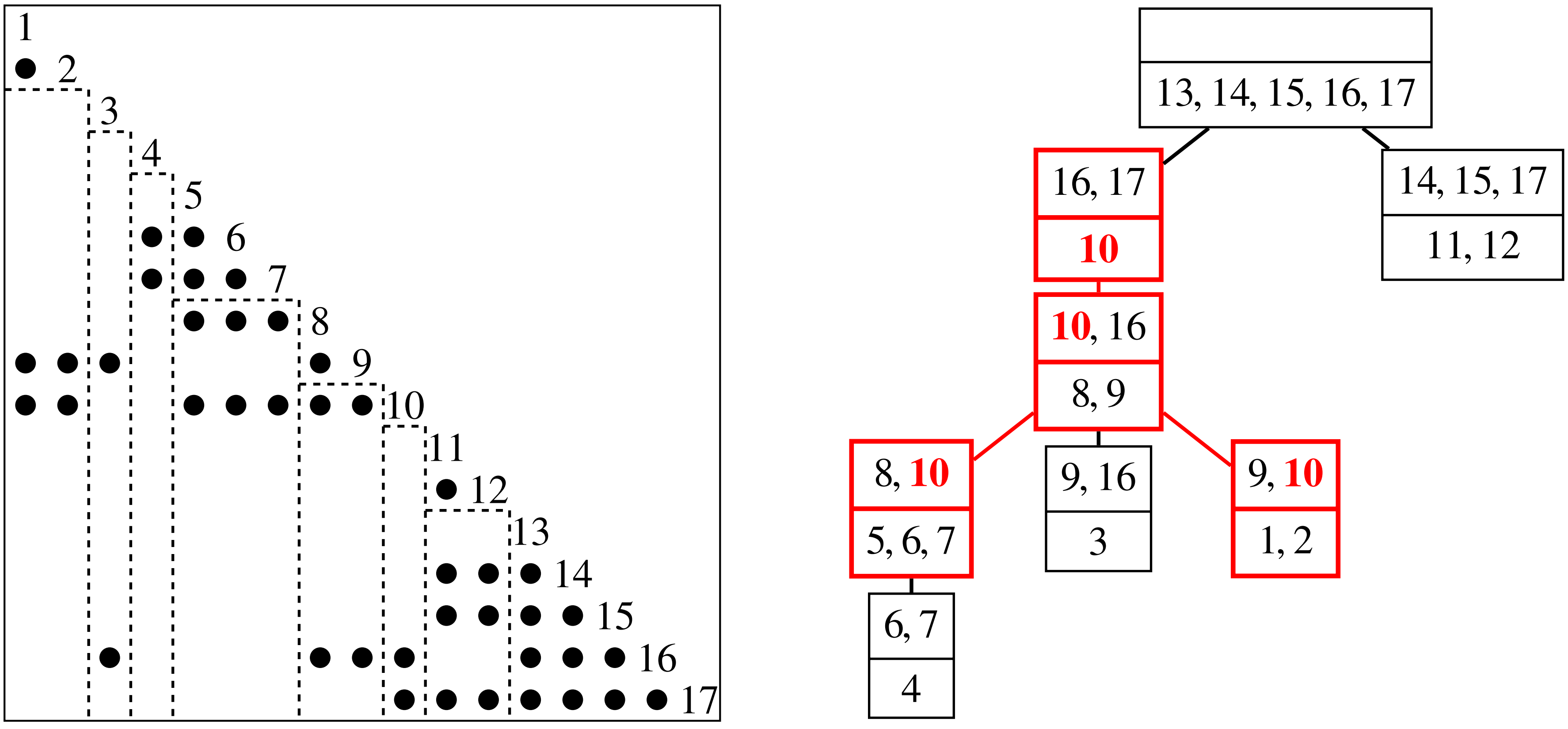
Supernodal elimination tree (clique tree)





- vertices of tree are cliques of chordal sparsity graph
- top row of each block is intersection of clique with parent clique
- bottom rows are (maximal) *supernodes*; form a partition of $\{1, 2, \ldots, n\}$
- for each v, cliques that contain v form a subtree of elimination tree

Supernodal elimination tree (clique tree)



- vertices of tree are cliques of chordal sparsity graph
- top row of each block is intersection of clique with parent clique
- bottom rows are *supernodes*; form a partition of $\{1, 2, \ldots, n\}$
- for each v, cliques that contain v form a subtree of elimination tree

Outline

1. Chordal graphs

2. **Decomposition of sparse matrix cones**

3. Multifrontal algorithms for logarithmic barrier functions

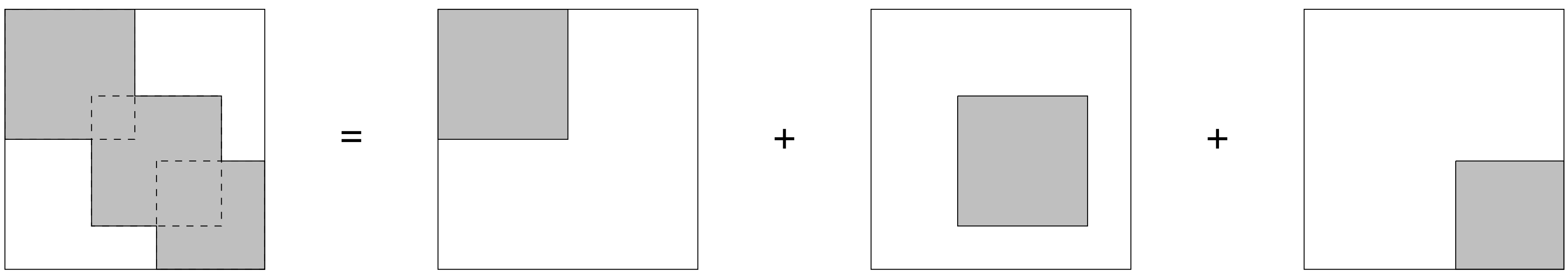
4. Minimum rank positive semidefinite completion

Positive semidefinite matrices with chordal sparsity pattern

$S \in \mathbf{S}^n_E$ is positive semidefinite if and only if it can be expressed as

$$S = \sum_{\text{cliques } \gamma_i} P_{\gamma_i}^T H_i P_{\gamma_i} \qquad \text{with } H_i \succeq 0$$

(for an index set β, P_β is 0-1 matrix of size $|\beta| \times n$ with $P_\beta x = x_\beta$ for all x)

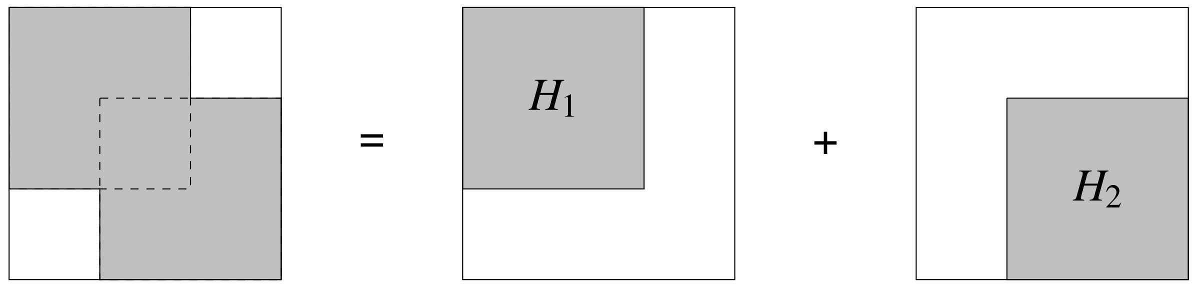


$$S \succeq 0 \qquad P_{\gamma_1}^T H_1 P_{\gamma_1} \succeq 0 \qquad P_{\gamma_2}^T H_2 P_{\gamma_2} \succeq 0 \qquad P_{\gamma_3}^T H_3 P_{\gamma_3} \succeq 0$$

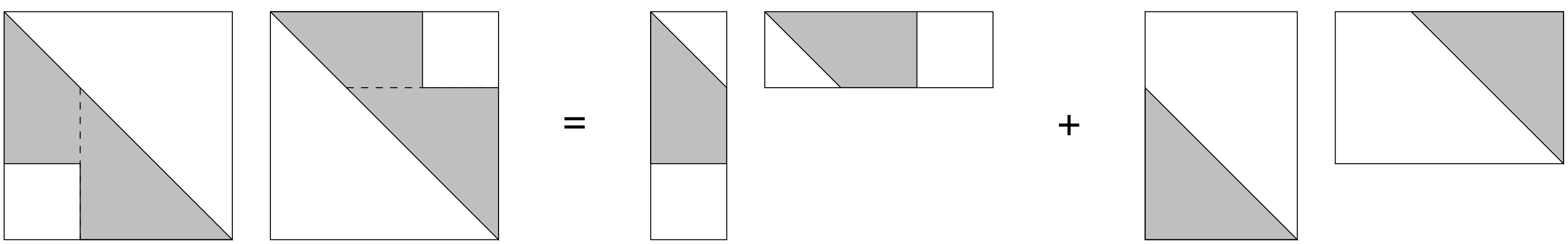
[Griewank and Toint 1984], [Agler, Helton, McCullough, Rodman 1988]

Decomposition from Cholesky factorization

- example with two cliques:



H_1 and H_2 follow by combining columns in Cholesky factorization



- readily computed from update matrices in multifrontal Cholesky factorization

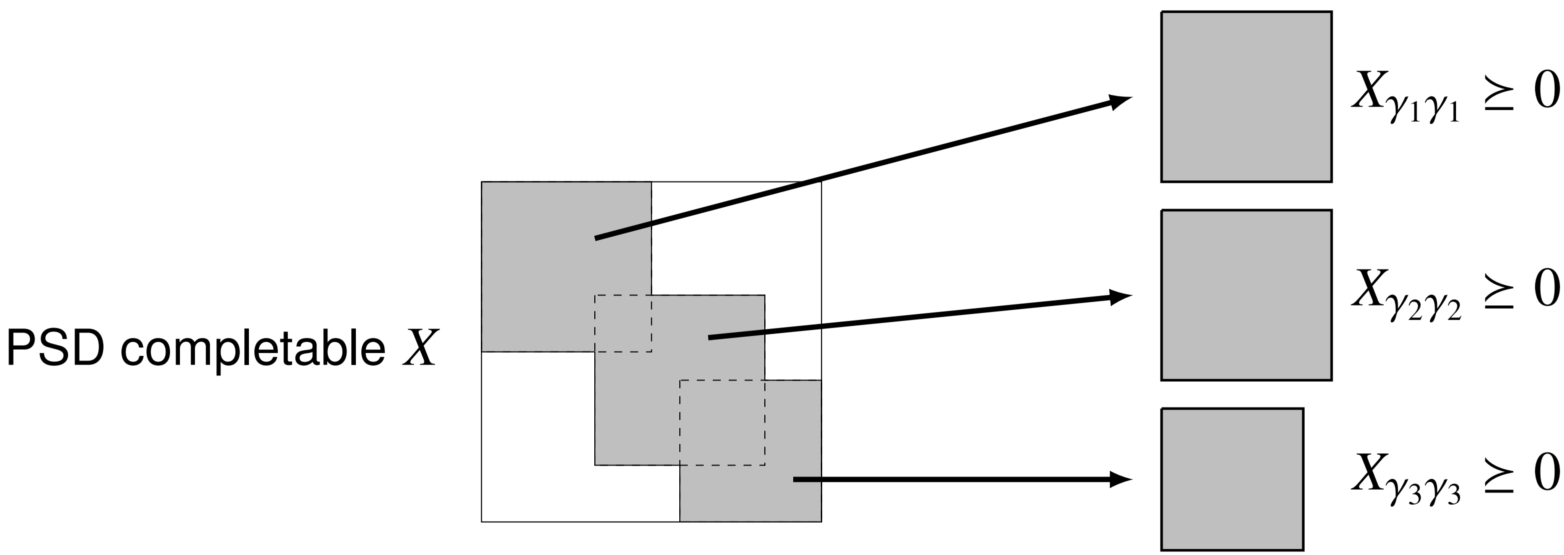
PSD completable matrices with chordal sparsity

$X \in \mathbf{S}_E^n$ has a positive semidefinite completion if and only if

$$X_{\gamma_i \gamma_i} \succeq 0 \quad \text{for all cliques } \gamma_i$$

follows from duality and clique decomposition of positive semidefinite cone

Example (three cliques γ_1, γ_2, γ_3)



[Grone, Johnson, Sá, Wolkowicz, 1984]

Sparse semidefinite optimization

$$
\begin{array}{ll}
\text{minimize} & \mathbf{tr}(CX) \\
\text{subject to} & \mathbf{tr}(A_i X) = b_i, \quad i = 1, \ldots, m \\
& X \in K
\end{array}
$$

- E is union of sparsity patterns of C, A_1, ..., A_m
- $K = \Pi_E(\mathbf{S}^n_+)$ is cone of p.s.d. completable matrices
- without loss of generality, can assume E is chordal

Decomposition algorithms

- cone K is intersection of simple cones ($X_{\gamma_i \gamma_i} \succeq 0$ for all cliques γ_i)
- first used in interior-point methods [Fukuda *et al.* 2000], [Nakata *et al.* 2003]
- first-order, splitting, and dual decomposition methods
  [Lu, Nemirovski, Monteiro 2007], [Lam, Zhang, Tse 2011], [Sun *et al.* 2014, 2015],
  [Pakazad *et al.* 2017], [Zheng, Fantuzzi, Papachristodoulou, Goulart, Wynn 2017], ...

Outline

1. Chordal graphs

2. Decomposition of sparse matrix cones

3. **Multifrontal algorithms for logarithmic barrier functions**

4. Minimum rank positive semidefinite completion

Sparse SDP as nonsymmetric conic linear program

Standard form SDP

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_iX) = b_i, \ i = 1, \ldots, m \\ & X \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & \sum_{i=1}^m y_i A_i + S = C \\ & S \succeq 0 \end{array}$$

Equivalent conic linear program

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_iX) = b_i, \ i = 1, \ldots, m \\ & X \in K \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & \sum_{i=1}^m y_i A_i + S = C \\ & S \in K^* \end{array}$$

- $K \in \Pi_E(\mathbf{S}^n_+)$ is cone of p.s.d. completable matrices with pattern E
- $K^* \in \mathbf{S}^n_+ \cap \mathbf{S}^n_E$ is cone of p.s.d. matrices with pattern E
- optimization problem in a lower-dimensional space $\mathbf{S}^n_E$
- K is not self-dual; no symmetric primal-dual interior-point methods

Barrier function for positive semidefinite cone

$$\phi(S) = -\log\det S, \qquad \mathrm{dom}\,\phi = \{S \in \mathbf{S}^n_E \mid S \succ 0\}$$

- gradient (negative projected inverse)

$$\nabla\phi(S) = -\Pi_E(S^{-1})$$

requires entries of dense inverse S^{-1} on diagonal and for $\{i, j\} \in E$

- Hessian applied to sparse $Y \in \mathbf{S}^n_E$:

$$\nabla^2\phi(S)[Y] = \frac{d}{dt}\nabla\phi(S + tY)\bigg|_{t=0} = \Pi_E\left(S^{-1}YS^{-1}\right)$$

requires projection of dense product $S^{-1}YS^{-1}$

Multifrontal algorithms

assume E is a chordal sparsity pattern (or chordal extension)

Cholesky factorization [Duff and Reid 1983]

- factorization $S = LDL^T$ gives function value of barrier: $\phi(S) = -\sum_i \log D_{ii}$
- computed by a recursion on elimination tree in topological order

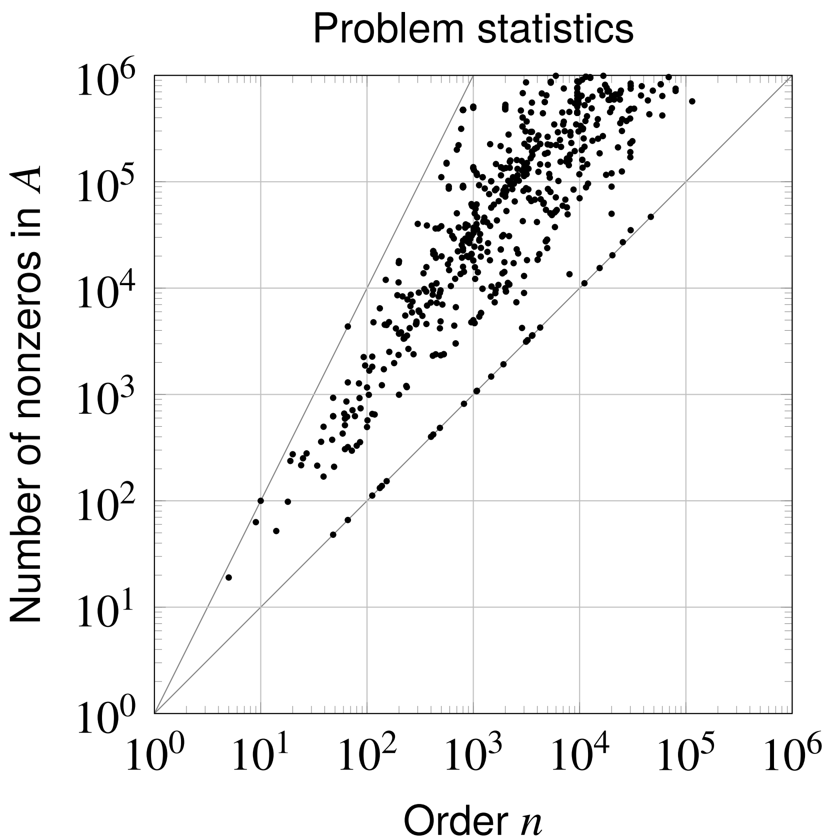
Gradient [Campbell and Davis 1995], [Andersen *et al.* 2013]

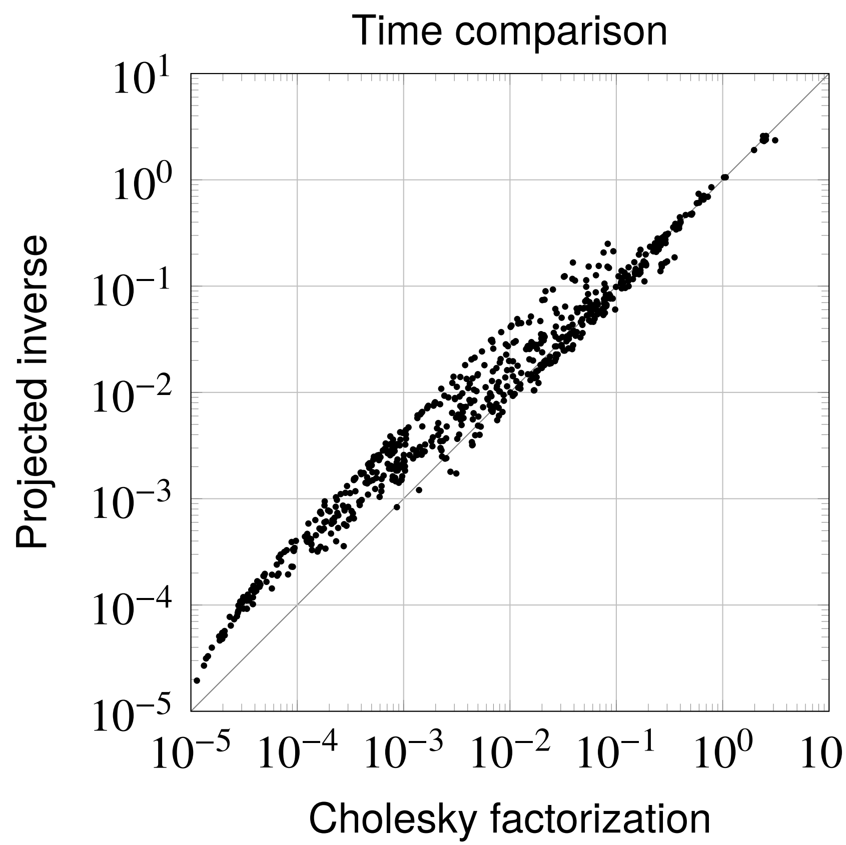
- compute $\nabla\phi(S) = -\Pi_E(S^{-1})$ from equation $S^{-1}L = L^{-T}D^{-1}$
- recursion on elimination tree in inverse topological order

Hessian

- compute $\nabla^2\phi(S)[Y] = \Pi_E(S^{-1}YS^{-1})$ by linearizing recursion for gradient
- two recursions on elimination tree (topological and inverse topological order)

Projected inverse versus Cholesky factorization





- 667 patterns from University of Florida Sparse Matrix Collection
- time in seconds for projected inverse and Cholesky factorization
- code at `github.com/cvxopt/chompack`

Barrier for positive semidefinite completable cone

$$\phi_*(X) = \sup_S \left(-\,\mathbf{tr}(XS) - \phi(S)\right), \qquad \mathrm{dom}\,\phi_* = \{X = \Pi_E(Y) \mid Y \succ 0\}$$

- this is the conjugate of the barrier $\phi(S) = -\log\det S$ for the sparse p.s.d. cone
- inverse $Z = \widehat{S}^{-1}$ of optimal $\widehat{S}$ is maximum determinant PD completion of X:

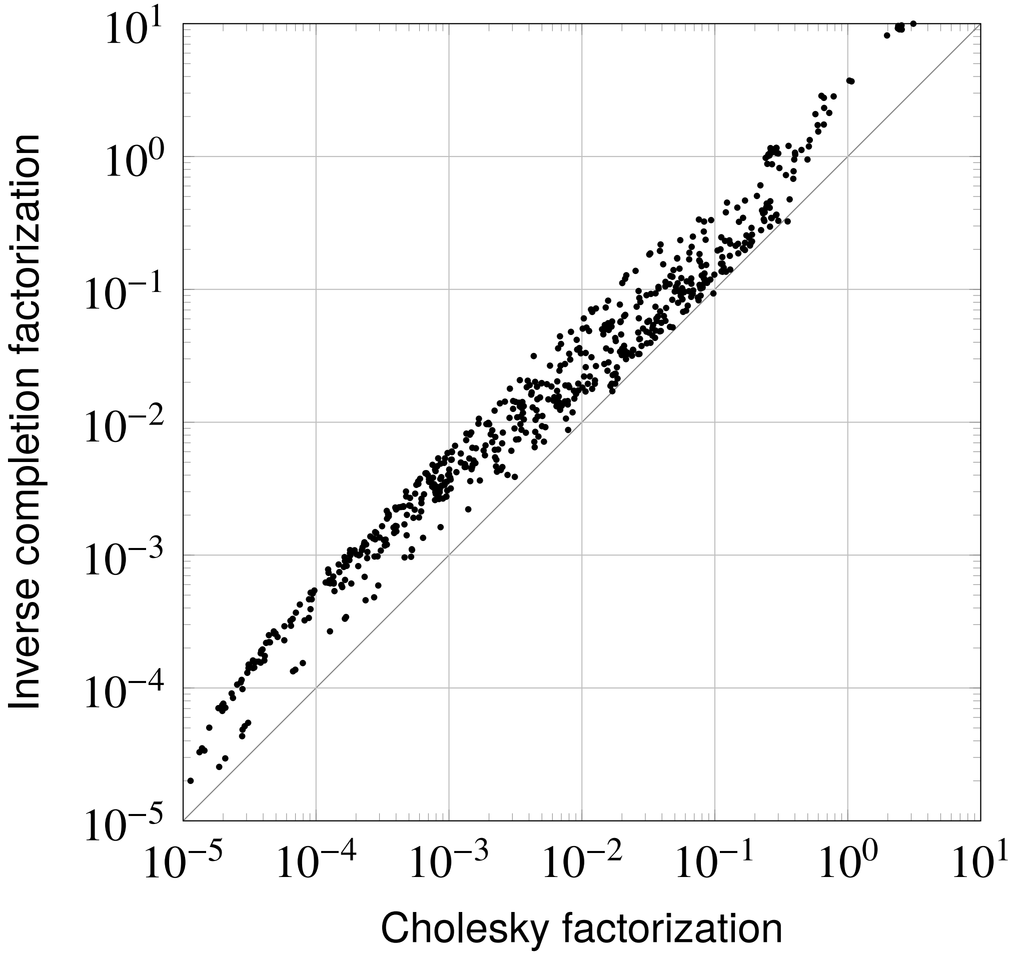
$$\begin{array}{ll} \text{maximize} & \log\det Z \\ \text{subject to} & \Pi_E(Z) = X \end{array}$$

- gradient and Hessian of ϕ_* at X are

$$\nabla\phi_*(X) = -\widehat{S}, \qquad \nabla^2\phi_*(X) = \nabla^2\phi(\widehat{S})^{-1}$$

for chordal E, efficient ‘multifrontal’ algorithms for Cholesky factors of $\hat{S}$, given X

Inverse completion versus Cholesky factorization



time for Cholesky factorization of inverse of maximum determinant PD completion

Nonsymmetric interior-point methods

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_i X) = b_i, \quad i = 1, \ldots, m \\ & X \in \Pi_E(\mathbf{S}^n_+) \end{array}$$

- can be solved by nonsymmetric primal or dual barrier methods
- logarithmic barriers for cone $\Pi_E(\mathbf{S}^n_+)$ and its dual cone $\mathbf{S}^n_+ \cap \mathbf{S}^n_E$:

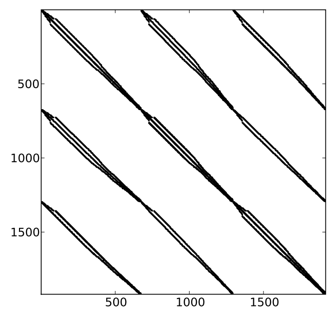
$$\phi_*(X) = \sup_S \left(-\mathbf{tr}(XS) + \log\det S\right), \qquad \phi(S) = -\log\det S$$

- fast evaluation of barrier values and derivatives if pattern is chordal
- examples: linear complexity per iteration for band or arrow pattern
- code and numerical results at `github.com/cvxopt/smcp`

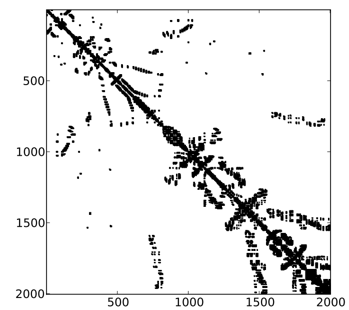
[Fukuda *et al.* 2000], [Burer 2003], [Srijungtongsiri and Vavasis 2004], [Andersen *et al.* 2010]

Sparsity patterns

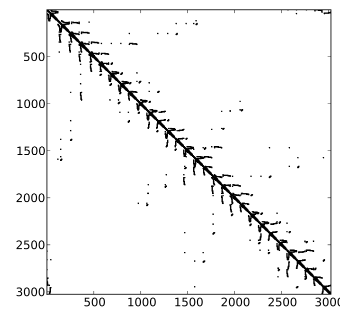
- sparsity patterns from University of Florida Sparse Matrix Collection
- $m = 200$ constraints
- randomly generated data with 0.05% nonzeros in A_i relative to $|E|$



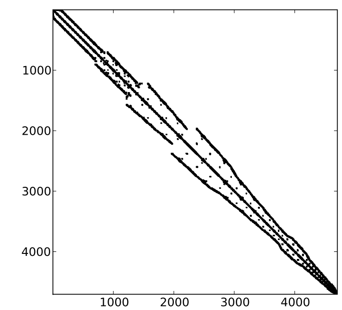
rs228
$n = 1{,}919$



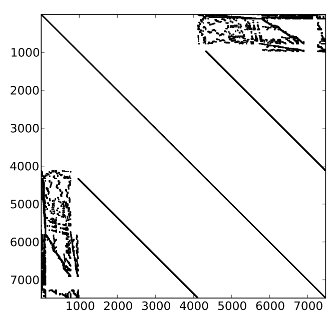
rs35
$n = 2{,}003$



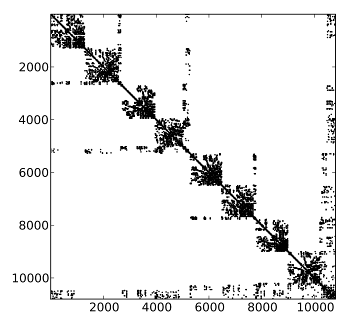
rs200
$n = 3{,}025$



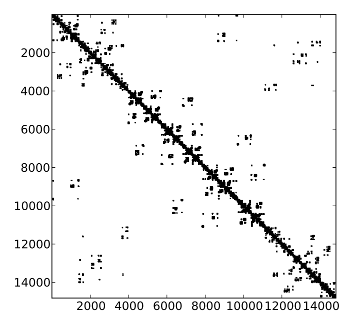
rs365
$n = 4{,}704$



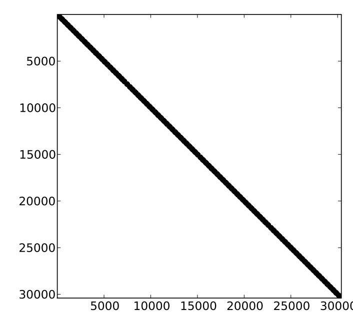
rs1555
$n = 7{,}479$



rs828
$n = 10{,}800$



rs1184
$n = 14{,}822$



rs1288
$n = 30{,}401$

Results

n	DSDP	SDPA	SDPA-C	SDPT3	SeDuMi	SMCP
1919	1.4	30.7	5.7	10.7	511.2	2.3
2003	4.0	34.4	41.5	13.0	521.1	15.3
3025	2.9	128.3	6.0	33.0	1856.9	2.2
4704	15.2	407.0	58.8	99.6	4347.0	18.6

n	DSDP	SDPA-C	SMCP
7479	22.1	23.1	9.5
10800	482.1	1812.8	311.2
14822	791.0	2925.4	463.8
30401	mem	2070.2	320.4

- average time per iteration for different solvers
- SMCP uses nonsymmetric matrix cone approach [Andersen *et al.* 2010]

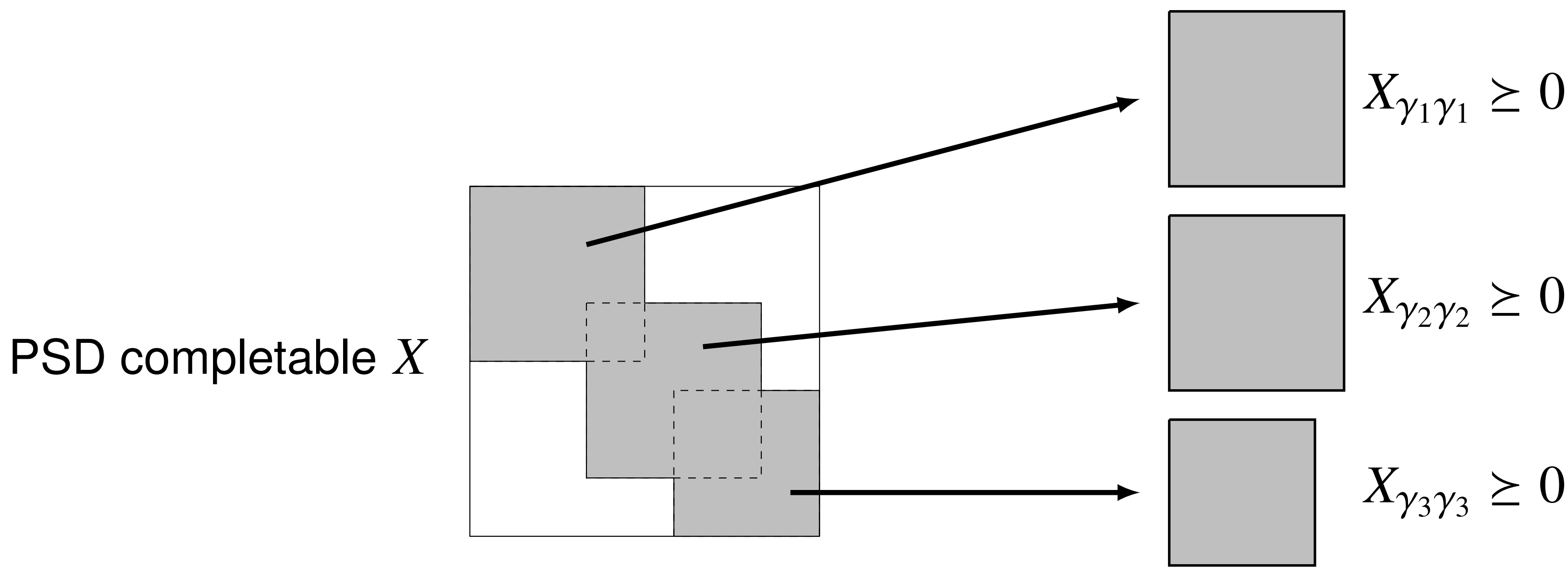
Outline

1. Chordal graphs
2. Decomposition of sparse matrix cones
3. Multifrontal algorithms for logarithmic logarithmic barriers
4. **Minimum rank positive semidefinite completion**

Minimum rank PSD completion with chordal sparsity

recall that $X \in \mathbf{S}^n_E$ has a positive semidefinite completion if and only if

$$X_{\gamma_i\gamma_i} \succeq 0 \quad \text{for all cliques } \gamma_i$$



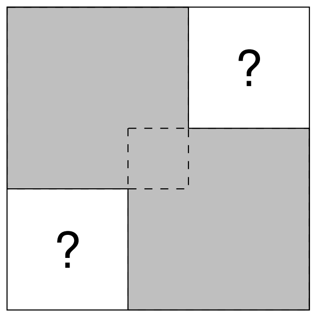
the minimum rank PSD completion has rank equal to

$$\max_{\text{cliques } \gamma_i} \operatorname{rank}(X_{\gamma_i\gamma_i})$$

[Dancis 1992]

Two-block completion problem

we consider the simple two-block completion problem



$$X = \begin{bmatrix} X_{11} & X_{12} & 0 \\ X_{21} & X_{22} & X_{23} \\ 0 & X_{32} & X_{33} \end{bmatrix}$$

- a completion exists if and only if

$$C_1 = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \succeq 0, \qquad C_2 = \begin{bmatrix} X_{22} & X_{23} \\ X_{32} & X_{33} \end{bmatrix} \succeq 0$$

- we construct a positive semidefinite completion of rank

$$r = \max\{\operatorname{rank}(C_1), \operatorname{rank}(C_2)\}$$

Two-block completion algorithm

- compute matrices U, V, $\tilde{V}$, W of column dimension r such that

$$\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix}^T, \qquad \begin{bmatrix} X_{22} & X_{23} \\ X_{32} & X_{33} \end{bmatrix} = \begin{bmatrix} \tilde{V} \\ W \end{bmatrix} \begin{bmatrix} \tilde{V} \\ W \end{bmatrix}^T$$

- since $VV^T = \tilde{V}\tilde{V}^T$, there exists an orthogonal $r \times r$ matrix Q such that

$$V = \tilde{V}Q$$

(computed from SVDs: take $Q = Q_2 Q_1^T$ where $V = P\Sigma Q_1^T$ and $\tilde{V} = P\Sigma Q_2^T$)

- a completion of rank r is given by

$$\begin{bmatrix} UQ^T \\ \tilde{V} \\ W \end{bmatrix} \begin{bmatrix} UQ^T \\ \tilde{V} \\ W \end{bmatrix}^T = \begin{bmatrix} X_{11} & X_{12} & UQ^T W^T \\ X_{21} & X_{22} & X_{23} \\ WQU^T & X_{32} & X_{33} \end{bmatrix}$$

Sparse semidefinite optimization

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_i X) = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{array}$$

- any feasible X can be replaced by a PSD completion of $\Pi_E(X)$:

$$\tilde{X} \succeq 0, \qquad \Pi_E(\tilde{X}) = \Pi_E(X)$$

- for chordal E, can take $\tilde{X} = YY^T$ with rank bounded by largest clique size
- proves exactness of some simple SDP relaxations
- useful for rounding solution of SDP relaxations to minimum rank solution

Summary

Sparse matrix theory: PSD and PSD-completable matrices with chordal pattern

- decomposition of sparse matrix cones as sum or intersection of simple cones
- fast algorithms for evaluating barrier functions and derivatives
- simple algorithms for maximum determinant and minimum rank completion

Applications in SDP algorithms

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_iX) = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{array}$$

- decomposition and splitting methods
- nonsymmetric interior-point methods