Sparsity and decomposition in semidefinite optimization

Lieven Vandenberghe ECE Department, UCLA

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

DYSCO PAI/IUAP Network Study Day Leuven, November 28, 2017

Semidefinite program (SDP)

minimize
$$\mathbf{tr}(CX)$$

subject to $\mathbf{tr}(A_iX) = b_i, \quad i = 1, \dots, m$
 $X \ge 0$

variable *X* is $n \times n$ symmetric matrix; $X \ge 0$ means *X* is positive semidefinite

- matrix inequalities arise naturally in many areas (for example, control, statistics)
- used in convex modeling systems (CVX, YALMIP, CVXPY, ...)
- relaxations of nonconvex quadratic and polynomial optimization

Algorithms

- primal-dual interior-point algorithms (used in SeDuMi, SDPT3, MOSEK)
- nonlinear programming methods based on parameterization $X = YY^T$
- first order methods

This talk: structure in solution X that results from sparsity in coefficients A_i , C

cost of solving SDP with banded matrices (bandwidth 11, 100 constraints)



- for bandwidth 1 (linear program), cost/iteration is linear in *n*
- for bandwidth > 1, cost grows as n^2 or faster

[Andersen, Dahl, Vandenberghe 2010]

Power flow optimization

an optimization problem with non-convex quadratic constraints

Variables

- complex voltage v_i at each node (bus) of the network
- complex power flow s_{ij} entering the link (line) from node *i* to node *j*

Non-convex constraints

• (lower) bounds on voltage magnitudes

$$v_{\min} \le |v_i| \le v_{\max}$$

• flow balance equations:

$$\xrightarrow{S_{ij}} \underbrace{g_{ij}}_{\text{bus } i} \underbrace{g_{ij}}_{\text{bus } j} \underbrace{s_{ij}}_{\text{bus } j} = \overline{g}_{ij} |v_i - v_j|^2$$

 g_{ij} is admittance of line from node *i* to *j*

Semidefinite relaxation of optimal power flow problem

- introduce matrix variable $X = \operatorname{Re}(vv^H)$, *i.e.*, with elements $X_{ij} = \operatorname{Re}(v_i \bar{v}_j)$
- voltage bounds and flow balance equations are convex in *X*:

$$v_{\min} \le |v_i| \le v_{\max} \longrightarrow v_{\min}^2 \le X_{ii} \le v_{\max}^2$$
$$s_{ij} + s_{ji} = \bar{g}_{ij} |v_i - v_j|^2 \longrightarrow s_{ij} + s_{ji} = \bar{g}_{ij} (X_{ii} + X_{jj} - 2X_{ij})$$

- replace constraint $X = \operatorname{Re}(vv^H)$ with weaker constraint $X \ge 0$
- relaxation is exact if optimal X has rank two

Sparsity in SDP relaxation:

off-diagonal X_{ij} appears in constraints only if there is a line between buses i and j

[Jabr 2006] [Bai et al. 2008] [Lavaei and Low 2012], [Molzahn et al. 2013], ...



• 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: $A \in \mathbf{S}_E^n$
- represented by undirected graph (V, E) with edges E, vertices $V = \{1, ..., n\}$
- clique (maximal complete subgraph) forms maximal 'dense' principal submatrix



• 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: $A \in \mathbf{S}_E^n$
- represented by undirected graph (V, E) with edges E, vertices $V = \{1, ..., n\}$
- clique (maximal complete subgraph) forms maximal 'dense' principal submatrix

Sparse matrix cones

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

• positive semidefinite matrices

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

• matrices with a positive semidefinite completion

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

 Π_E is projection on \mathbf{S}_E^n

Properties

- two cones are convex
- closed, pointed, with nonempty interior (relative to S_F^n)
- form a pair of dual cones (for the trace inner product)

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

minimizetr(CX)maximize $b^T y$ subject to $tr(A_iX) = b_i, i = 1, ..., m$ subject to $\sum_{i=1}^m y_i A_i + S = C$ $X \ge 0$ $S \ge 0$

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

minimizetr(CX)maximize $b^T y$ subject to $tr(A_iX) = b_i, i = 1, ..., m$ subject to $\sum_{i=1}^m y_i A_i + S = C$ $X \in K$ $S \in K^*$

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

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

• 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



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

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 $A \in \mathbf{S}_{E}^{n}$:

$$PAP^T = LDL^T$$

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

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

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

A has a 'zero fill' Cholesky factorization

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

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

Simple patterns







Sparsity pattern of a Cholesky factor

- •: edges of non-chordal sparsity pattern
- o: 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, ..., *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 $\{1, 2, ..., n\}$
- for each *v*, cliques that contain *v* form a subtree of elimination tree

- 1. Sparse semidefinite programs
- 2. Chordal graphs

3. Decomposition of sparse matrix cones

- 4. Multifrontal algorithms for logarithmic barrier functions
- 5. Minimum rank positive semidefinite completion

Positive semidefinite matrices with chordal sparsity pattern

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

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

(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)



[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

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

 $X_{\gamma_i\gamma_i} \geq 0$ for all cliques γ_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

minimize $\mathbf{tr}(CX)$ subject to $\mathbf{tr}(A_iX) = b_i, \quad i = 1, \dots, m$ $X \in K$

- *E* is union of sparsity patterns of *C*, A_1, \ldots, A_m
- $K = \prod_{E} (\mathbf{S}_{+}^{n})$ 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 $(X_{\gamma_i\gamma_i} \ge 0 \text{ for all cliques } \gamma_i)$
- 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]
 ...

Example: sparse nearest matrix problems

• find nearest sparse PSD-completable matrix with given sparsity pattern

minimize $||X - A||_F^2$ subject to $X \in \Pi_E(\mathbf{S}^n_+)$

• find nearest sparse PSD matrix with given sparsity pattern

minimize $||S + A||_F^2$ subject to $S \in \mathbf{S}^n_+ \cap \mathbf{S}^n_E$

these two problems are duals:



Decomposition methods

from the decomposition theorems, the two problems can be written

| Primal: | minimize | $\ X - A\ _F^2$ |
|---------|------------|---|
| | subject to | $X_{\gamma_i\gamma_i} \ge 0$ for all cliques γ_i |
| Dual: | minimize | $\ A + \sum_i P_{\gamma_i}^T H_i P_{\gamma_i}\ _F^2$ |
| | subject to | $H_i \geq 0$ for all cliques γ_i |

Algorithms

- Dykstra's algorithm (dual block coordinate ascent)
- (fast) dual projected gradient algorithm (FISTA)
- Douglas-Rachford splitting, ADMM

sequence of projections on PSD cones of order $|\gamma_i|$ (eigenvalue decomposition)

Results

sparsity patterns from University of Florida Sparse Matrix Collection

| п | density | #cliques | avg. clique size | max. clique |
|--------|---------|----------|------------------|-------------|
| 20141 | 2.80e-3 | 1098 | 35.7 | 168 |
| 38434 | 1.25e-3 | 2365 | 28.1 | 188 |
| 57975 | 9.04e-4 | 8875 | 14.9 | 132 |
| 79841 | 9.71e-4 | 4247 | 44.4 | 337 |
| 114599 | 2.02e-4 | 7035 | 18.9 | 58 |

| | total runtime (sec) | | | time/i | time/iteration (sec) | | |
|--------|---------------------|---------|-------|--------|----------------------|-----|--|
| n | FISTA | Dykstra | DR | FISTA | Dykstra | DR | |
| 20141 | 2.5e2 | 3.9e1 | 3.8e1 | 1.0 | 1.6 | 1.5 | |
| 38434 | 4.7e2 | 4.7e1 | 6.2e1 | 2.1 | 1.9 | 2.5 | |
| 57975 | > 4hr | 1.4e2 | 1.1e3 | 3.5 | 5.7 | 6.4 | |
| 79841 | 2.4e3 | 3.0e2 | 2.4e2 | 6.3 | 7.6 | 9.7 | |
| 114599 | 5.3e2 | 5.5e1 | 1.0e2 | 2.6 | 2.2 | 4.0 | |

[Sun and Vandenberghe 2015]

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

Standard form SDP

minimize $\mathbf{tr}(CX)$ subject to $\mathbf{tr}(A_iX) = b_i, i = 1, \dots, m$ $X \ge 0$

maximize
$$b^T y$$

subject to $\sum_{i=1}^m y_i A_i + S = C$
 $S \ge 0$

Equivalent conic linear program

minimize
$$tr(CX)$$
maximize $b^T y$ subject to $tr(A_iX) = b_i, i = 1, ..., m$ subject to $\sum_{i=1}^m y_i A_i + S = C$ $X \in K$ $S \in K^*$

- $K \in \prod_{E}(\mathbf{S}^{n}_{+})$ 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}_E^n
- *K* is not self-dual; no symmetric primal-dual interior-point methods

Barrier function for positive semidefinite cone

$$\phi(S) = -\log \det S, \qquad \operatorname{dom} \phi = \{S \in \mathbf{S}_E^n \mid S > 0\}$$

• 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 $Y \in \mathbf{S}_E^n$:

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

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 $\nabla^2 \phi(S)[Y] = \prod_E (S^{-1}YS^{-1})$ 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

$$\phi_*(X) = \sup_S (-\operatorname{tr}(XS) - \phi(S)), \quad \operatorname{dom} \phi_* = \{X = \Pi_E(Y) \mid Y > 0\}$$

- 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:

maximize
$$\log \det Z$$

subject to $\Pi_E(Z) = X$

• 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

minimize $\mathbf{tr}(CX)$ subject to $\mathbf{tr}(A_iX) = b_i, \quad i = 1, \dots, m$ $X \in \prod_E(\mathbf{S}^n_+)$

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

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

- 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|



Results

| п | 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 |

| п | 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]

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

Minimum rank PSD completion with chordal sparsity

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

 $X_{\gamma_i\gamma_i} \geq 0$ for all cliques γ_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



• a completion exists if and only if

$$C_{1} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \ge 0, \qquad C_{2} = \begin{bmatrix} X_{22} & X_{23} \\ X_{32} & X_{33} \end{bmatrix} \ge 0$$

• 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

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

minimize
$$\operatorname{tr}(CX)$$

subject to $\operatorname{tr}(A_iX) = b_i, \quad i = 1, \dots, m$
 $X \ge 0$

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

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

• for chordal *E*, can take $\tilde{X} = YY^T$ with rank bounded by largest clique size

minimize
$$\mathbf{tr}(Y^T C Y)$$

subject to $\mathbf{tr}(Y^T A_i Y) = b_i, \quad i = 1, ..., m$

- proves exactness of some simple SDP relaxations
- useful for rounding solution of SDP relaxations to minimum rank solution

| | | | MOSEK 8 | | SeDuMi v1.05 | | SDPT3 v4.0 | |
|----------|------|----------------|----------------------------------|------------------------------------|----------------------------------|------------------------------------|----------------------------------|------------------------------------|
| | n | max. clique | $\operatorname{rank}(X^{\star})$ | $\operatorname{rank}(X^{\bullet})$ | $\operatorname{rank}(X^{\star})$ | $\operatorname{rank}(X^{\bullet})$ | $\operatorname{rank}(X^{\star})$ | $\operatorname{rank}(X^{\bullet})$ |
| IEEE-118 | 118 | 20 | 1 | 1 | 1 | 1 | 1 | 1 |
| IEEE-300 | 300 | 17 | 5 | 1 | 5 | 1 | 5 | 1 |
| 2383wp | 2383 | 31 | 17 | 1 | 17 | 1 | 17 | 1 |
| 2736sp | 2736 | 30 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2737sop | 2737 | 29 | 44 | 1 | 43 | 1 | 43 | 1 |
| 2746wop | 2746 | 30 | 32 | 1 | 32 | 1 | 32 | 1 |
| 2746wp | 2746 | 31 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3012wp | 3012 | 32 | 346 | 13 | 346 | 13 | 337 | 17 |
| 3120sp | 3120 | 32 | 514 | 27 | 572 | 32 | 519 | 27 |
| 3375wp | 3375 | 33 | 451 | 19 | 451 | 19 | 454 | 21 |

- benchmark problems from MATPOWER package
- rank is number of eigenvalues greater than $10^{-5}\sqrt{n}\lambda_{max}$
- X^{\star} is the (Hermitian) solution of the relaxation computed by SDP solver
- X^{\bullet} is minimum rank PSD completion of $\Pi_E(X^{\star})$



 X^{\star} is computed by SeDuMi; X^{\bullet} is minimum rank completion of $\Pi_{E}(X^{\star})$

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

minimize
$$\mathbf{tr}(CX)$$

subject to $\mathbf{tr}(A_iX) = b_i, \quad i = 1, \dots, m$
 $X \ge 0$

- decomposition and splitting methods
- nonsymmetric interior-point methods