Implementation of nonsymmetric interior-point methods for linear optimization over sparse matrix cones

1401111495 苏奇

1100010676 林意

1401111493 段高鹏

2014.12.29

北京大學

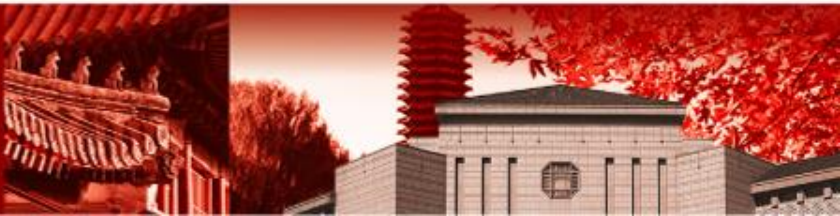
问题的提出

- Semidefinite program (SDP)

$$\begin{aligned} \text{minimize } & C \bullet X \\ \text{subject to } & A_i \bullet X = b_i, \quad i = 1, \ldots, m, \\ & X \succeq 0 \end{aligned} \tag{1a}$$

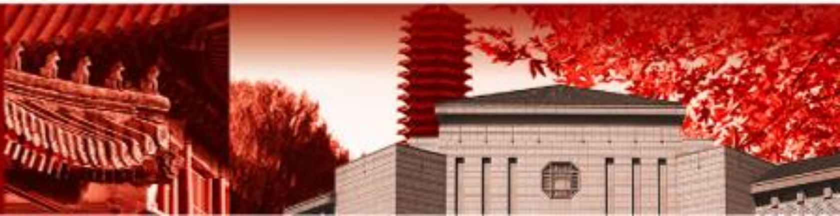
- dual problem

$$\begin{aligned} \text{maximize } & b^T y \\ \text{subject to } & \sum_{i=1}^{m} y_i A_i + S = C \\ & S \succeq 0 \end{aligned} \tag{1b}$$



一些重要的符号说明

- S_V^n： the symmetric matrices of order n with sparsity pattern V
- $X \geq_c 0$： the sparse matrix X has a positive semidefinite completion
- $P_V(X)$： the projection on S_V^n of positive semidefinite matrix X



问题转化

The pair of SDPs (1a)-(1b) is equivalent to

$$\text{P:} \quad \begin{array}{ll} \text{minimize} & C \bullet X \\ \text{subject to} & A_i \bullet X = b_i, \quad i = 1, \ldots, m \\ & X \succeq_c 0 \end{array} \qquad \text{D:} \quad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & \sum_{i=1}^{m} y_i A_i + S = C \\ & S \succeq 0, \end{array} \tag{2}$$

assume that $C, A_1, \ldots, A_m \in S_V^n$ with $y \in R^m, S \in S_V^n$.

一些概念的介绍

A sparsity pattern of a symmetric matrix is definited by the set of positions (i,j) where the matrix is allowed to be nonzero,i.e.$X \in S_V^n$ has sparsity pattern V if $X_{ij} = X_{ji} = 0$ for $(i, j) \notin V$.

$Y = P_V(X), i.e., Y_{ij} = X_{ij}$ if $(i, j) \in V$ and otherwise $Y_{ij} = 0$.

Chordal matrix cones

The sparsity pattern is chordal if the graph G_V is chordal, i.e., every cycle of length greater than three has a chord (an edge joining nonconsecutive nodes of the cycle).

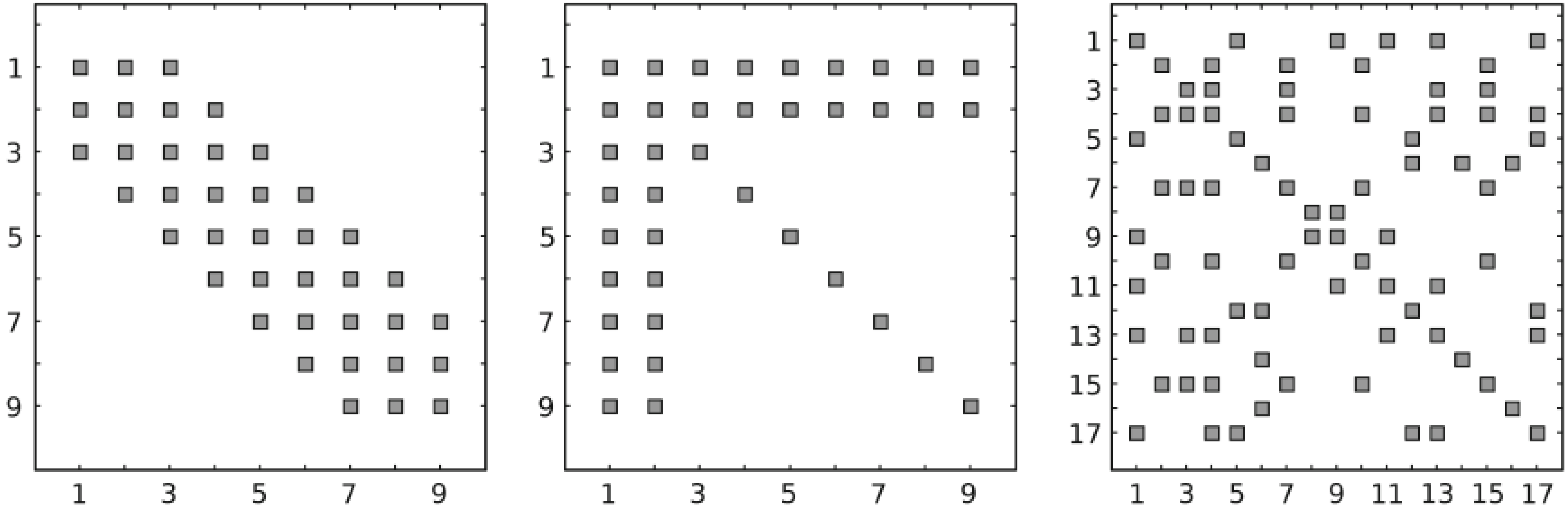


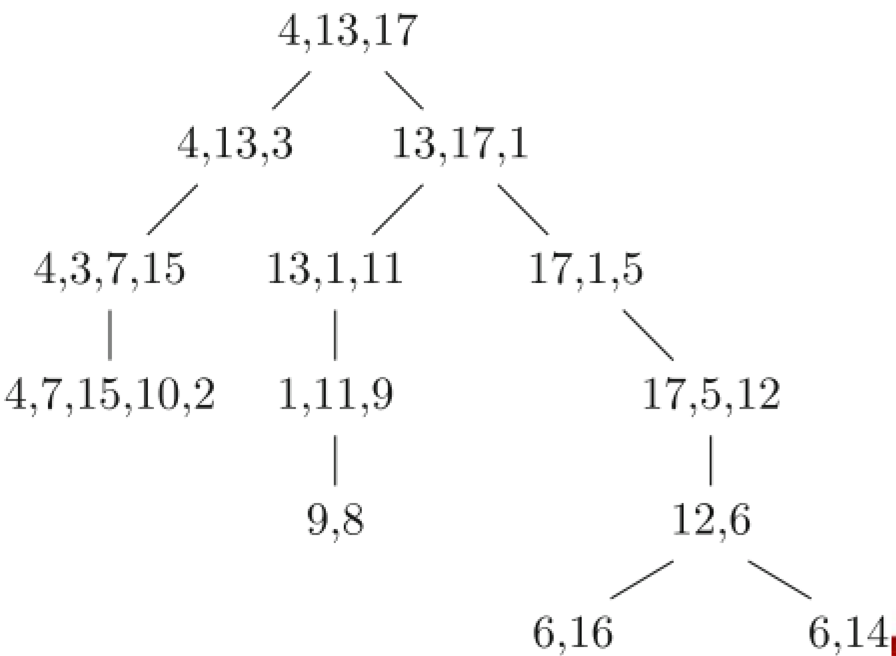
Fig. 1 Examples of chordal sparsity patterns

clique

A clique in G_V is a set of nodes that defines a maximal complete subgraph.

A clique tree is a maximum weight spanning tree of the clique graph.

Fig. 2 Clique tree of the third chordal sparsity pattern in Fig. 1



Properties

The useful properties of chordal sparsity patterns follow from a basic property known as the running intersection property.

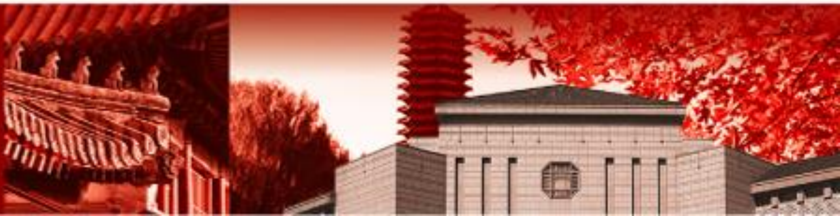
Suppose G_V has l cliques W_1, W_2 ... W_l,define
$U_1 = \emptyset$, $V_1 = W_1$, and, for $i = 2, \ldots, l$,

$$U_i = W_i \cap (W_1 \cup W_2 \cup \cdots \cup W_{i-1}), \quad V_i = W_i \backslash (W_1 \cup W_2 \cup \cdots \cup W_{i-1}).$$

Then the property states that

$$U_i = W_i \cap W_{\text{par}(i)}, \quad V_i = W_i \backslash W_{\text{par}(i)}.$$

where $W_{\text{par}(i)}$ is the parent of W_i in the clique tree.



Cholesky factorization

If $S \in S^n_{V,++}$,then there exists a permutation matrix P and a lower triangular matrix L such that

$$P^T SP = LL^T$$

and $L + L^T$ has the sparsity pattern V.

Value and gradient of dual barrier

logarithmic barrier function for $S^n_{V,+}$, defined

$$\phi : \mathbf{S}^n_V \to \mathbf{R}, \quad \phi(S) = -\log\det S, \quad \mathbf{dom}\,\phi = \mathbf{S}^n_{V,++}$$

computer Cholesky factorization

$$\phi(S) = -2\sum_{i=1}^{n} \log L_{ii}.$$

The gradient of ϕ is given by

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

Hessian and inverse Hessian of dual barrier

The Hessian of ϕ at S,applied to a matrix $Y \in S_V^n$,

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

The expression can be evaluated from the Cholesky factorization of S and the projected inverse $P_V(S^{-1})$ via a pair of adjoint linear operators

$$\nabla^2 \phi(S)[Y] = \mathcal{L}_{\mathrm{adj}}(\mathcal{L}(Y)),$$

$$\nabla^2 \phi(S)^{-1}[Y] = \mathcal{L}^{-1}(\mathcal{L}_{\mathrm{adj}}^{-1}(Y))$$

Maximum determinant positive definite completion

Given a matrix $X \in S_V^n$,find the positive definite solution Z

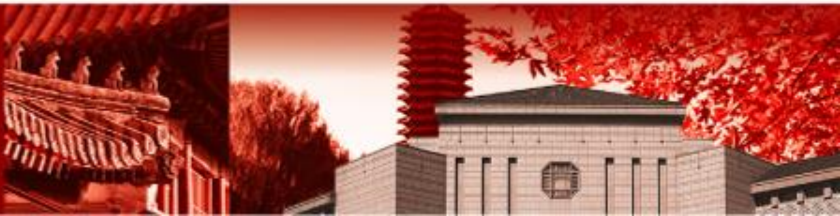
$$\begin{array}{ll} \text{maximize} & \log\det Z \\ \text{subject to} & P_V(Z) = X. \end{array}$$

If V is chordal, the solution can be computerd from X [4,18, 25,31,38,53].

Equivalent computing the Cholesky factor of $W = Z^{-1}$ [16].

$$P_V(W^{-1}) = X.$$

where$W \in S_V^n$.



Value and gradient of primal barrier

$$\mathbf{S}^n_{V,\mathrm{c}+} = \{X \in \mathbf{S}^n_V \mid X \succeq_\mathrm{c} 0\} = (\mathbf{S}^n_{V,+})^*$$

Use the Legendre transform of the barrie ϕ of S^n_V[39].

For $X >_c 0$,the barrier function is defined

$$\phi_\mathrm{c}(X) = \sup_{S \succ 0} \left(-X \bullet S - \phi(S)\right).$$

If the sparsity pattern is chordal,the solution $S \in S^n_V$ satisfies

$$P_V(S^{-1}) = X,$$

Evaluating ϕ_c:

Computer the Cholesky factorization $\hat{S} = LL^T$,$\hat{S}$ is the solution of $P_V(S^{-1}) = X$, then compute

$$\phi_c(X) = \log\det \hat{S} - n = 2\sum_{i=1}^{n} \log L_{ii} - n.$$

Follows from properties of Legendre transforms that

$$\nabla\phi_c(X) = -\hat{S},$$

and

$$X \bullet \nabla\phi_c(X) = -n.$$

Hessian and inverse Hessian of primal barrier

the Hessian of the primal barrier function is

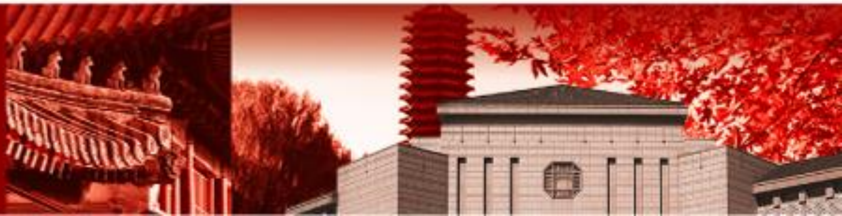
$$\nabla^2 \phi_c(X) = \nabla^2 \phi(\hat{S})^{-1},$$

Compute $\nabla^2 \phi_c(X)^{-1}[Y]$ using the algorithm for evaluating the dual Hessian.

Primal and dual Newton Systems

$$A_i \bullet X = b_i, \quad i = 1, \ldots, m, \qquad \sum_{i=1}^{m} y_i A_i + S = C, \qquad S = -\mu \nabla \phi_c(X)$$

$$A_i \bullet X = b_i, \quad i = 1, \ldots, m, \qquad \sum_{i=1}^{m} y_i A_i + S = C, \qquad X = -\mu \nabla \phi(S)$$



Primal and dual Newton Systems

$$A_i \bullet X = b_i, \quad i = 1, \ldots, m, \qquad \sum_{i=1}^{m} y_i A_i + S = C, \qquad S = -\mu \nabla \phi_c(X)$$

$$A_i \bullet X = b_i, \quad i = 1, \ldots, m, \qquad \sum_{i=1}^{m} y_i A_i + S = C, \qquad X = -\mu \nabla \phi(S)$$

Replacing X, y, S with $X+\Delta X$, $y+\Delta y$, $S+\Delta S$

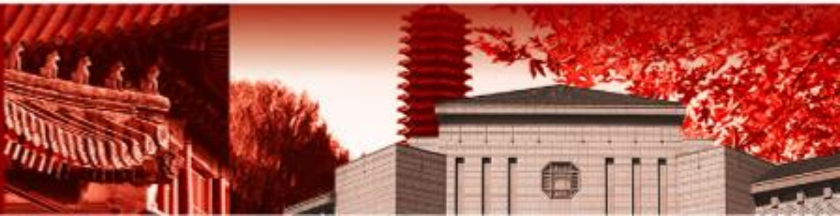
$$A_i \bullet \Delta X = r_i, \quad i = 1, \ldots, m, \qquad \sum_{i=1}^{m} \Delta y_i A_i - \mu \nabla^2 \phi_c(X)[\Delta X] = R$$

where $r_i = b_i - A_i \bullet X$ and

$$R = C - \sum_{i=1}^{m} y_i A_i + \mu \nabla \phi_c(X).$$

$$\nabla \phi_c(X) = -\hat{S}$$

$$P_V(\hat{S}^{-1}) = X$$



$$H \Delta y = g$$

$$H_{ij} = A_i \bullet \left(\nabla^2 \phi_c(X)^{-1}[A_j] \right), \quad i, j = 1, \ldots, m$$

$$g_i = \mu r_i + A_i \bullet (\nabla^2 \phi_c(X)^{-1}[R])$$

$$\nabla^2 \phi_c(X) = \nabla^2 \phi(\hat{S})^{-1} \qquad P_V(\hat{S}^{-1}) = X$$

$$H_{ij} = A_i \bullet (\nabla^2 \phi(\hat{S})[A_j]), \quad i, j = 1, \ldots, m$$

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

Method 1: Cholesky factorization

$$T1: \quad U := \nabla^2 \phi_c(X)^{-1}[A_j]$$

$$\textbf{for } i = j \text{ to } m \textbf{ do}$$

$$\quad H_{ij} := A_i \bullet U$$

$$\textbf{end for}$$

$$A_j = \sum_{(p,q)\in I_j} (A_j)_{pq} e_p e_q^T$$

$$H_{ij} = \sum_{(p,q)\in I_j} (A_j)_{pq} \left(A_i \bullet (\nabla^2 \phi_c(X)^{-1}[e_p e_q^T] \right)$$

$$= \sum_{(p,q)\in I_j} (A_j)_{pq} \left(A_i \bullet P_V(\hat{S}^{-1} e_p e_q^T \hat{S}^{-1}) \right)$$

$$= \sum_{(p,q)\in I_j} (A_j)_{pq} \left(A_i \bullet (L^{-T} L^{-1} e_p e_q^T L^{-T} L^{-1}) \right)$$

$T2:$ Solve $LL^T u_k = e_k$ for $k \in \{i \mid A_j e_i \neq 0\}$

for $i = j$ to m **do**

$$H_{ij} := \sum_{(p,q)\in I_j} (A_j)_{pq} u_q^T A_i u_p$$

end for

Method 2: QR factorization

$$H_{ij} = A_i \bullet \left(\nabla^2 \phi_c(X)^{-1}[A_j]\right) = \mathcal{L}(A_i) \bullet \mathcal{L}(A_j)$$

$$\boxed{\nabla^2 \phi_c(X)^{-1} = \nabla^2 \phi(\hat{S}) = \mathcal{L}_{\mathrm{adj}} \circ \mathcal{L}}$$

$$H = \tilde{A}^T \tilde{A}$$

$$\begin{bmatrix} -D & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

For sparse A , taking a sparse LDL^T factorization

For dense A, taking a QR decomposition of $\tilde{A} = D^{-1/2} A^T$

Section 5: Implementation

- A feasible-start path-following algorithm
- Feasible-start primal scaling method

Algorithm Outline

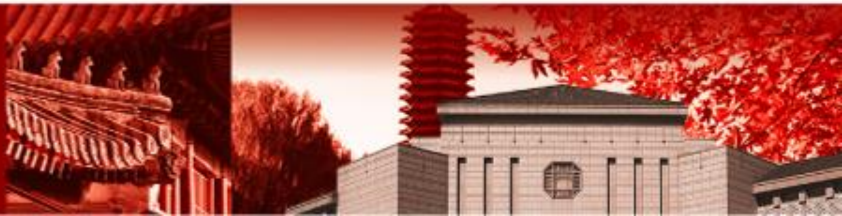
- 1. Primal centering.
- 2. Prediction step.
- 3. Stopping criteria

Equations Solved in Each Iterations

$$A_i \cdot \Delta X = 0, \qquad i = 1, \dots, m$$

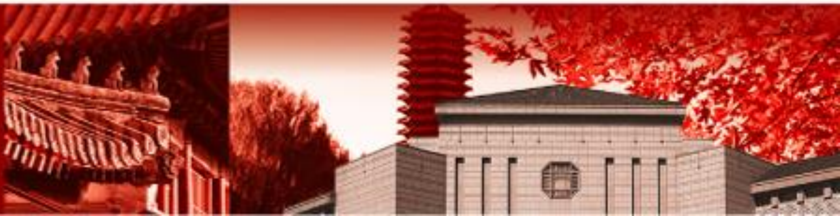
$$\sum_{i=1}^{m} \Delta y_i A_i + \Delta S = 0\,,$$

$$\mu \nabla^2 \phi_c(X)[\Delta X] + \Delta S = -\mathrm{R}.$$



Assumptions

- A_i are linearly independent.
- A chordal embedding is available. The clique tree has been computed.
- Requires a strictly feasible starting point X and an initial value of positive parameter μ.
- Depends on parameters:
  - $\delta \in (0,1)$
  - $\gamma \in (0,0.5)$
  - $\beta \in (0,1)$
  - Tolerances ϵ_{abs} and ϵ_{rel}.



Step 1: Primal Centering

- Solve equations with $R = C + \mu\nabla\phi_c(X)$. Solution: ΔX_{cnt} , Δy_{cnt}, ΔS_{cnt}
- Evaluate Newton decrement

$$\lambda = (\Delta X_{cnt} \cdot \nabla^2\phi_c(X)[\Delta X_{cnt}])^{1/2}$$

- If$\lambda \leq \delta$, set

$$S := C + \Delta S_{cnt}, \qquad y := \Delta y_{cnt}$$

go to step 2

Step 1 (continued)

- Else update $X := X + \alpha\Delta X_{cnt}$, repeat Step1

  α is a step size from a backtracking line search and satisfies $X + \alpha\Delta X_{cnt} \succ_c 0$ and Armijo conditon

$$\frac{1}{\mu}\left(C \cdot (X + \alpha\Delta X_{cnt})\right) + \phi_c(X + \alpha\Delta X_{cnt})$$
$$\leq \frac{1}{\mu}(C \cdot X) + \phi_c(X) - \alpha\gamma\lambda^2$$

Step 2: Prediction Step

- Solve equations with $R = S$

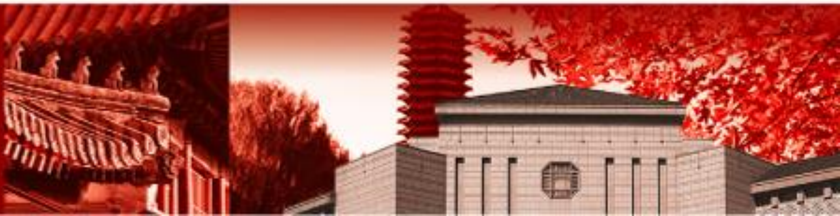
  Solution: ΔX_{at} , Δy_{at}, ΔS_{at}

- $\tilde{X} := X - \Delta X_{cnt}$

$$\alpha = 0.98 \sup\{\alpha \in [0,1) | \tilde{X} + \alpha \Delta X_{at} \succ_c 0, \quad S + \alpha \Delta S_{at} \succ_c 0\}$$

$$\hat{\mu} := \frac{(\tilde{X} + \alpha \Delta X_{at}) \cdot (S + \alpha \Delta S_{at})}{n}$$

Update $\mu = \hat{\mu}$



Step 2(continued)

- Solve equations with $\mu = \hat{\mu}$ and
$$R = S + \hat{\mu}\nabla\phi_c(X).$$
- Conduct a backtracking line search to find a primal step size α_p satisfies

  $X + \alpha\Delta X \succ_c 0$ and Armijo condition, to find a dual step size $\alpha_d = max_k\{\beta^k | S + \beta^k \Delta S \succ 0\}$
- Update
  - $X := X + \alpha_p \Delta X,$
  - $y := y + \alpha_d \Delta y,$
  - $S := S + \alpha_d \Delta S.$

Step 3: Stopping Criteria

- Terminate if

$$X \cdot S \leq \epsilon_{abs}$$

or

$$\min\{C \cdot X, -b^T y\} < 0$$

And

$$\frac{X \cdot S}{-\min\{C \cdot X, -b^T y\}} \leq \epsilon_{rel}$$

Otherwise, set $\mu := (X \cdot S)/n$, and go to step 1.

Remarks

- Finding the feasible-start point: Solve the least-norm problem

$$\min ||X||_F^{\,2}$$
$$s.t.\ A_i \cdot X = b_i$$

If the solution $X_{ln} \nsucceq_c 0$, solve phase Ⅰ problem

$$\min s$$
$$s.t. A_i \cdot X = b_i, tr(X) \leq M, X + (s - \epsilon) \succcurlyeq_c 0$$

Remarks

- Non-chordal sparsity patterns can be embedded in a chordal sparsity pattern.

Embedding could enhance the performance.

- Numerical Stability

Table 1 DIMACS error measures for control6 from SDPLIB

Solver	ϵ_1	ϵ_3	ϵ_5	ϵ_6
M1	1.63e−07	0.00e+00	1.04e−05	6.79e−06
M2	9.97e−14	0.00e+00	4.30e−10	3.63e−10
CSDP	5.67e−08	9.41e−09	3.66e−08	1.42e−08
SDPA	4.17e−07	1.81e−09	1.15e−06	1.03e−06
SDPT3	3.50e−07	1.80e−09	7.80e−07	7.40e−07
SEDUMI	1.45e−06	0.00e+00	−2.92e−08	3.28e−06

谢 谢!