Semidefinite Programming and Matrix Completions

Henry Wolkowicz

Department of Combinatorics & Optimization
University of Waterloo

OUTLINE

1. Short intro. to SDP
2. SDP and positive definite matrix completions
3. Euclidean distance matrix (EDM) completions
4. New characterization for EDM;
   solving large sparse problems

(Advantages of using $X - \mu Z^{-1} = 0$ form of perturbed complementary slackness.)

SDP BACKGROUND and NOTATION

Semidefinite Programming
looks just like
Linear Programming

$$\begin{array}{lllll} & p^* = & \max & \operatorname{trace} CX & (\langle C, X\rangle) \\ (\mathbf{PSDP}) & & \text{s.t.} & \mathcal{A}X = b & (\text{linear}) \\ & & & X \succeq 0, \quad (X \in \mathcal{P}) & (\text{nonneg}) \end{array}$$

$\preceq$ denotes the Löwner partial order

$A \preceq B$ if $B - A \succeq 0$

$\mathcal{S}^n$ denotes $n \times n$ symmetric matrices

$$\mathcal{A} : \mathcal{S}^n \to \Re^m$$

$$(\mathcal{A}X)_i = \operatorname{trace}(A_i X), \text{ for given } A_i \in \mathcal{S}^n$$

$\mathcal{P}$ - cone of positive semidefinite matrices

replaces

$\Re^n_+$ - nonnegative orthant

DUALITY

payoff function, player Y to player X (Lagrangian)

$$L(X,y) := \operatorname{trace}(CX) + y^t(b - \mathcal{A}X)$$

Optimal (worst case) strategy for player X:

$$p^* = \max_{X \succeq 0} \min_{y} L(X,y)$$

Using the *hidden constraint* $b - \mathcal{A}X = 0$, recovers primal problem.

$$
\begin{aligned}
L(X,y) &= \operatorname{trace}(CX) + y^t(b - \mathcal{A}X) \\
&= b^t y + \operatorname{trace}\,(C - \mathcal{A}^* y)\, X
\end{aligned}
$$

adjoint operator, $\mathcal{A}^* y = \sum_i y_i A_i$

$$
\langle \mathcal{A}^* y, X \rangle = \langle y, \mathcal{A}X \rangle\,, \quad \forall X, y
$$

$$
p^* = \max_{X \succeq 0} \min_{y} L(X,y) \leq d^* := \min_{y} \max_{X \succeq 0} L(X,y)
$$

The *hidden constraint* $C - \mathcal{A}^* y \preceq 0$

$$p^* = \max_{X \succeq 0} \min_y L(X,y) \leq d^* := \min_y \max_{X \succeq 0} L(X,y)$$

dual obtained from optimal strategy of competing player Y;
use *hidden constraint* $C - \mathcal{A}^* y \preceq 0$

$$(\mathbf{DSDP}) \quad \begin{array}{rcc} d^* = & \min & b^t y \\ & \text{s.t.} & \mathcal{A}^* y \succeq C \end{array}$$

for the primal

$$(\mathbf{PSDP}) \quad \begin{array}{rcc} p^* = & \max & \operatorname{trace} CX \\ & \text{s.t.} & \mathcal{A}X = b \\ & & X \succeq 0 \end{array}$$

Characterization of optimality for the
dual pair X, y (slack $Z \succeq 0$)

$$\begin{aligned} \mathcal{A}^* y - Z = C & \qquad \text{dual feasibility} \\ AX = b & \qquad \text{primal feasibility} \\ ZX = 0 & \qquad \text{complementary slackness} \end{aligned}$$

$$ZX = \mu I \qquad \text{perturbed}$$

Forms the basis for:

interior point methods
(primal simplex method, dual simplex method)

Positive Definite Completions of Partial Hermitian Matrices

- $\mathcal{G}(V, E)$ finite undirected graph
- $A(\mathcal{G})$ is a $\mathcal{G}$-partial matrix (a_{ij} defined iff $\{i, j\} \in E$)
- $A(\mathcal{G})$ is a $\mathcal{G}$-partial positive matrix if $a_{ij} = \overline{a_{ji}}, \forall \{i, j\} \in E$ and all existing principal minors are positive.
- with $\mathcal{J} = (V, \bar{E}), E \subset \bar{E}$ a $\mathcal{J}$-partial matrix $B(\mathcal{J})$ extends the $\mathcal{G}$-partial matrix $A(\mathcal{G})$ if $b_{ij} = a_{ij}, \forall \{i, j\} \in E$
- $\mathcal{G}$ is positive completable if every $\mathcal{G}$-partial positive matrix can be extended to a positive definite matrix.

$\mathcal{G}$ is **chordal** if there are no minimal cycles of length ≥ 4. (every cycle of length ≥ 4 has a chord)

THEOREM (Grone, Johnson, Sa, Wolkowicz)
$\mathcal{G}$ is positive completable iff $\mathcal{G}$ is chordal. $\blacksquare$

equivalently - strict feasibility for SDP:

$$\begin{aligned} \operatorname{trace} E_{ij} P &= a_{ij}, \quad \forall \{i,j\} \in E \\ P &\succ 0 \end{aligned}$$

where $E_{ij} = e_i e_j^t + e_j e_k^t$

Approximate Positive Semidefinite Completions

(with Charlie Johnson
and Brenda Kroschel)

given:

$H = H^t \geq 0$ a real, nonnegative (elementwise) **symmetric matrix of weights**, with positive diagonal elements $H_{ii} > 0, \ \forall i$;
and $A = A^*$ the **given partial Hermitian matrix**
(i.e. some elements approximately fixed; others free; for notational purposes, assume free elements set to 0 if not specified.)

$||A||_F = \sqrt{\operatorname{trace} A^* A}$ *Frobenius norm*, $\circ$ denotes *Hadamard product.*

$$f(P) := ||H \circ (A - P)||_F^2$$

weighted, best approximate, completion problem

$$(AC) \qquad \begin{array}{lll} \mu^* := & \min & f(P) \\ & \text{subject to} & KP = b \\ & & P \succeq 0, \end{array}$$

where $K : \mathcal{H}^n \to \mathcal{C}^m$ linear operator

Lagrangian:

$$L(P, y, \Lambda) = f(P) + \langle y, b - KP\rangle - \operatorname{trace} \Lambda P$$

Dual problem:

$$(DAC) \quad \begin{array}{ll} \max & f(P) + \langle y, b - KP\rangle - \operatorname{trace} \Lambda P \\ \text{subject to} & \nabla f(P) - K^* y - \Lambda = 0 \\ & \Lambda \succeq 0. \end{array}$$

THEOREM The matrix $\bar{P} \succeq 0$ and vector-matrix $\bar{y}, \bar{\Lambda} \succeq 0$ solve AC and DAC if and only if

$$
\begin{array}{ll}
K\bar{P} = b & \text{primal feas.} \\
2H^{(2)} \circ (\bar{P} - A) - K^*\bar{y} - \bar{\Lambda} = 0 & \text{dual feas.} \\
\operatorname{trace} \bar{\Lambda}\bar{P} = 0 & \text{compl. slack.}
\end{array}
$$

■

For simplicity and sparsity, discard linear operator K and replace with appropriate weights in H.

Use (*square*) perturbed optimality conditions.

$$
\begin{array}{ll}
2H^{(2)} \circ (P - A) - \Lambda = 0 & \text{dual feasibility} \\
-P + \mu\Lambda^{-1} = 0 & \text{perturbed C.S.}
\end{array}
$$

Linearization of second equation
and solve for h and l

$$
h = \mu\Lambda^{-1} - \mu\Lambda^{-1} l \Lambda^{-1} - P
$$

$$
l = \frac{1}{\mu}\{-\Lambda(P + h)\Lambda\} + \Lambda
$$

Dual Step First:
(if many elements of P are free)

We can eliminate the primal step h and solve for the dual step l.

$$\begin{aligned} l &= 2H^{(2)} \circ h + \left(2H^{(2)} \circ (P - A) - \Lambda\right) \\ &= 2H^{(2)} \circ \left(\mu\Lambda^{-1} - \mu\Lambda^{-1} l \Lambda^{-1} - P\right) \\ &\qquad + \left(2H^{(2)} \circ (P - A) - \Lambda\right). \end{aligned}$$

Equivalently, we get the Newton equation

$$2H^{(2)} \circ \left(\mu\Lambda^{-1} l \Lambda^{-1}\right) + l = 2H^{(2)} \circ \left(\mu\Lambda^{-1} - A\right) - \Lambda.$$

l, Λ have same sparsity pattern as H,
order is number of nonzeros/2 in H.

dim	toler	Hdens./infty	Apsd	cond(A)	Hpd	min/max	iters
60	10^{-6}	.01/.001	yes	79.7	no	15/23	16.8
65	10^{-6}	.015/.001	yes	49.9	yes	18/24	21.3
83	10^{-6}	.007/.001	no	235.1	no	24/29	25.5
85	10^{-5}	.008/.001	yes	94.7	no	11/17	13.1
85	10^{-6}	.0075/.001	no	299.9	no	23/27	25.2
87	10^{-6}	.006/.001	yes	74.2	yes	14/19	16.9
89	10^{-6}	.006/.001	no	179.3	no	23/28	15.2
110	10^{-6}	.007/.001	yes	172.3	yes	15/20	17.8
155	10^{-6}	.01/0	yes	643.9	yes	14/18	15.3
655	10^{-6}	.017/0	yes	1.4	no	14/14	14.
755	10^{-6}	.002/0	yes	1.5	no	15/15	15.

data for dual-step-first (20 problems per test): dimension; tolerance for duality gap;
density of nonzeros in H/ density of infinite values in H;
positive semidefiniteness of A; condition number of A; positive definiteness of H;

(only one test for: 655,755)

Euclidean Distance Matrix Completion Problem

(with Abdo Alfakih)

What are EDMs?

—–

A **pre-distance matrix** (or dissimilarity matrix):

• an $n \times n$ symmetric matrix $D = (d_{ij})$ with nonnegative elements and zero diagonal

—–

A (squared) **Euclidean distance matrix** (EDM):

• a pre-distance matrix such that there exists points $x^1, x^2, \ldots, x^n$ in $\Re^r$ such that

$$d_{ij} = \left\| x^i - x^j \right\|^2, \quad i, j = 1, 2, \ldots, n.$$

—–

The smallest value of r is called **the embedding dimension** of D. (r is always $\leq n - 1$)

EDM problem:

Given a partial symmetric matrix A with certain elements specified, the Euclidean distance matrix completion problem (EDMCP) consists in finding the unspecified elements of A that make A a EDM.

WHY?

e.g.:

• The shape of an enzyme determines it chemical function. Once the shape is known, then the proper drug can be designed.

• distance geometry on molecules: Atoms are points in space with pairwise distances; find a set of points which yield those distances.

For approximate EDMCP:

A is a pre-distance matrix, H is an $n \times n$ symmetric weight matrix,

$$f(D) := \|H \circ (A - D)\|_F^2,$$

$$(CDM_0) \quad \begin{array}{rcc} \mu^* := & \min & f(D) \\ & \text{subject to} & D \in \mathcal{E}, \end{array}$$

where $\mathcal{E}$ denotes the cone of EDMs.

DISTANCE GEOMETRY A pre-distance matrix D is a EDM if and only if D is negative semidefinite on

$$M := \left\{x \in \Re^n : x^t e = 0\right\},$$

where e is the vector of all ones.

Define **centered** and **hollow** subspaces

$$\begin{aligned} \mathcal{S}_C &:= \{B \in \mathcal{S}^n : Be = 0\}, \\ \mathcal{S}_H &:= \{D \in \mathcal{S}^n : \operatorname{diag}(D) = 0\}. \end{aligned}$$

Define two linear operators

$$\mathcal{K}(B) := \operatorname{diag}(B)\, e^t + e \operatorname{diag}(B)^t - 2B,$$

$$\mathcal{T}(D) := -\tfrac{1}{2} JDJ.$$

The operator $-2\mathcal{T}$ is an orthogonal projection onto $\mathcal{S}_C$.

THEOREM The linear operators satisfy

$$\mathcal{K}(\mathcal{S}_C) = \mathcal{S}_H,$$
$$\mathcal{T}(\mathcal{S}_H) = \mathcal{S}_C,$$

and $\mathcal{K}_{|\mathcal{S}_C}$ and $\mathcal{T}_{|\mathcal{S}_H}$ are inverses of each other. $\blacksquare$

A hollow matrix D is EDM
if and only if
$B = \mathcal{T}(D) \succeq 0$ (positive semidefinite)

D is EDM
if and only if
$D = \mathcal{K}(B)$, for some B with $Be = 0$ and $B \succeq 0$.

In this case the embedding dimension r is given by the rank of B. Moreover if $B = XX^t$, then the coordinates of the points $x^1, x^2, \ldots, x^n$ that generate D are given by the rows of X and, since $Be = 0$, it follows that the origin coincides with the centroid of these points.

For Projection: V $n \times (n-1)$, full column rank with $V^t e = 0$.

$$J := VV^{\dagger} = I - \frac{ee^t}{n}$$

is orthogonal projection onto M, where $V^{\dagger}$ denotes Moore-Penrose generalized inverse.

The cone of EDMs, $\mathcal{E}$, has empty interior. This can cause problems for interior-point methods.

$$V \cdot V : \mathcal{S}_{n-1} \to \mathcal{S}_n$$

$$V \cdot V : \mathcal{P}_{n-1} \to \mathcal{P}_n$$

Define the composite operators

$$\mathcal{K}_V(X) \quad := \quad \mathcal{K}(VXV^t),$$

and

$$\mathcal{T}_V(D) \quad := \quad V^\dagger \mathcal{T}(D)(V^\dagger)^t = -\tfrac{1}{2} V^\dagger D (V^\dagger)^t.$$

LEMMA

$$\mathcal{K}_V(\mathcal{S}_{n-1}) = \mathcal{S}_H,$$
$$\mathcal{T}_V(\mathcal{S}_H) = \mathcal{S}_{n-1},$$

and $\mathcal{K}_V$ and $\mathcal{T}_V$ are inverses of each other on these two spaces. ■

COROLLARY

$$\begin{aligned} \mathcal{K}_V(\mathcal{P}) &= \mathcal{E}, \\ \mathcal{T}_V(\mathcal{E}) &= \mathcal{P}. \end{aligned}$$

■

Summary

(Re)Define the closest EDM problem:

$$\begin{aligned} f(X) &:= \|H \circ (A - \mathcal{K}_V(X))\|_F^2 \\ &= \|H \circ \mathcal{K}_V(B - X)\|_F^2, \end{aligned}$$

where $B = \mathcal{T}_V(A)$.

($\mathcal{K}_V$ and $\mathcal{T}_V$ are both linear operators)

$$(CDM) \quad \begin{array}{lll} \mu^* := & \min & f(X) \\ & \text{subject to} & X \succeq 0. \end{array}$$

Primal-Dual Interior-Point Framework:

STEPS:

1. derive a dual program
2. state optimality conditions for log-barrier problem (perturbed primal-dual optimality conditions)
3. find a search direction for solving the perturbed optimality conditions
4. take a step and backtrack to stay strictly feasible (positive definite)
5. Update and go to Step 3 (adaptive update of log-barrier parameter)

Step 1. derive a dual program:

$\Lambda \in \mathcal{S}_{n-1}, \Lambda \succeq 0$ and $y \in R^m$,
Lagrangian is

$$L(X, y, \Lambda) = f(X) + \langle y, b - \mathcal{A}(X)\rangle - \langle \Lambda, X\rangle$$

primal program (CDM) is

$$= \min_{X} \max_{\substack{y \\ \Lambda \succeq 0}} L(X, y, \Lambda).$$

dual program is:

$$= \max_{\substack{y \\ \Lambda \succeq 0}} \min_{X} L(X, y, \Lambda),$$

The inner minimization of the convex, in X, Lagrangian is unconstrained so we add the hidden constraint which makes the minimization redundant.

dual program (DCDM)

$$\max_{\substack{\nabla f(X) - \mathcal{A}^* y = \Lambda \\ \Lambda \succeq 0}} f(X) + \langle y, b - \mathcal{A}(X) \rangle - \operatorname{trace} \Lambda X.$$

or

$$\begin{array}{cc} \max & f(X) + \langle y, b - \mathcal{A}(X) \rangle - \operatorname{trace} \Lambda X \\ \text{subject to} & \nabla f(X) - \mathcal{A}^* y - \Lambda = 0 \\ & \Lambda \succeq 0, (X \succeq 0). \end{array}$$

the duality gap,
$f(X) - (f(X) + \langle y, b - \mathcal{A}(X)\rangle - \operatorname{trace} \Lambda X)$,
in the case of primal and dual feasibility, is given by the complementary slackness condition:

$$\operatorname{trace} X(\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) - \mathcal{A}^* y) = 0,$$

or equivalently

$$X(\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) - \mathcal{A}^* y) = 0,$$

where $H^{(2)} = H \circ H$.

THEOREM Suppose that Slater's condition holds. Then $\bar{X} \succeq 0$, and $\bar{y}$, $\bar{\Lambda} \succeq 0$ solve (CDM) and (DCDM), respectively, if and only if the following three equations hold.

$$\begin{array}{rl} \mathcal{A}(\bar{X}) = b & \text{prim. feas.} \\ 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B)) - \mathcal{A}^*\bar{y} - \bar{\Lambda} = 0 & \text{dual feas.} \\ \operatorname{trace} \bar{\Lambda}\bar{X} = 0 & \text{C.S.} \end{array}$$

■

LEMMA Let H be an $n \times n$ symmetric matrix with nonnegative elements and 0 diagonal such that the graph of H is connected. Then

$$\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(I)) \succ 0,$$

where $I \in \mathcal{S}_{n-1}$ is the identity matrix. ■

Step 2. state optimality conditions for log-barrier problem (perturbed primal-dual optimality conditions):

The log-barrier problem for (CDM) is

$$\min_{X \succ 0} B_\mu(X) := f(X) - \mu \log\det(X),$$

where $\mu \downarrow 0$.

For each $\mu > 0$ we take one Newton step for solving the stationarity condition

$$\nabla B_\mu(X) = 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) - \mu X^{-1} = 0.$$

Let

$$C := 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(B)) = 2\mathcal{K}_V^*(H^{(2)} \circ A).$$

Then the stationarity condition is equivalent to

$$\nabla B_\mu(X) = 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(X)\right) - C - \mu X^{-1} = 0.$$

equating $\Lambda = \mu X^{-1}$ and multiplying through by X

optimality conditions, $F := \begin{pmatrix} F_d \\ F_c \end{pmatrix} = 0,$

$$\begin{array}{rcll} 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(X)\right) - C - \Lambda & = & 0 & \text{dual feas.} \\ \Lambda X - \mu I & = & 0 & \text{pert. C.S.,} \end{array}$$

(an OVERDETERMINED nonlinear system since ΛX not symmetric)

estimate of the barrier parameter

$$\mu = \frac{1}{n-1} \text{ trace } \Lambda X$$

σ_k centering parameter

$\mathcal{F}^0$ set of strictly feasible primal-dual points

F' derivative of F

Algorithm 1 *(p-d i-p framework:)*

Given $(X^0, \Lambda^0) \in \mathcal{F}^0$

for $k = 0, 1, 2 \ldots$

solve *for the search direction*

$$F'(X^k, \Lambda^k) \begin{pmatrix} \delta X^k \\ \delta \Lambda^k \end{pmatrix} = \begin{pmatrix} -F_d \\ -\Lambda^k X^k + \sigma_k \mu_k I \end{pmatrix}$$

where σ_k *centering,* $\mu_k = \frac{\operatorname{trace} X^k \Lambda^k}{(n-1)}$

$$(X^{k+1}, \Lambda^{k+1}) = (X^k, \Lambda^k) + \alpha_k (\delta X^k, \delta \Lambda^k)$$

so that $(X^{k+1}, \Lambda^{k+1}) \succ 0$

end (for).

For the EDM: search direction (Gauss-Newton direction) is the Frobenius norm lss of $F's = -F$, i.e.

$$\begin{aligned} 2\mathcal{K}_V^*\left(H^{(2)} \circ \mathcal{K}_V(h)\right) - l &= -F_d \\ \Lambda h + lX &= -F_c. \end{aligned}$$

$t(n) = \frac{(n+1)n}{2}$ dimension of $\mathcal{S}^n$.

$$F's = \begin{bmatrix} F'_{u1} & F'_{u2} \\ F'_{l1} & F'_{l2} \end{bmatrix} \begin{pmatrix} h \\ l \end{pmatrix} = rhs = \begin{pmatrix} rhs_1 \\ rhs_2 \end{pmatrix}.$$

$F' : \Re^{2(t(n-1))} \to \Re^{t(n-1)+(n-1)^2}$

Larger Models

Instead of projecting and reducing the dimension to get Slater's condition, add a variable and increase the dimension.

Lemma 1 *Let*

$$
\begin{array}{rcl}
\mathcal{F} & := & \{X \in \mathcal{S}^n : v^T e = 0 \quad \Rightarrow \quad v^T X v \leq 0\}, \\
\mathcal{F}_0 & := & \{X \in \mathcal{S}^n : X - \alpha e e^t \preceq 0, \quad \text{for some } \alpha \geq 0\}, \\
\mathcal{F}_1 & := & \{X \in \mathcal{S}^n : X - \alpha e e^t \preceq 0, \quad \forall\, \alpha \geq \bar{\alpha}, \\
& & \qquad \text{for some } \bar{\alpha} \geq 0\}.
\end{array}
$$

Then

$$
\mathrm{ri}\,(\mathcal{F}) \subset \mathcal{F}_0 = \mathcal{F}_1 \subset \mathcal{F} \subset \overline{\mathcal{F}_0}. \tag{1}
$$

Proof. Suppose that $\bar{X} \in \mathrm{ri}\,(\mathcal{F})$ (i.e. $v^T e = 0, v \neq 0 \Rightarrow v^T \bar{X} v < 0$) but $\bar{X} \notin \mathcal{F}_0$. Then, for each $\alpha \geq 0$, there exists w_α with $||w_\alpha|| = 1$, such that $w_\alpha \to \bar{w}$, as $\alpha \to \infty$ and

$$w_\alpha^T (\bar{X} - \alpha e e^t) w_\alpha > 0, \quad \forall\, \alpha \geq 0,$$

i.e.

$$w_\alpha^T \bar{X} w_\alpha > \alpha w_\alpha^T e e^t w_\alpha, \quad \forall\, \alpha \geq 0.$$

Since w_α converges and the left-hand-side of the above inequality must be finite, this implies that $e^t \bar{w} = \bar{w}^T \bar{X} \bar{w} = 0$, a contradiction. Therefore, $\mathrm{ri}\,(\mathcal{F}) \subset \mathcal{F}_0$. That $\mathcal{F}_0 = \mathcal{F}_1$ is clear.

Now suppose that $\bar{X} - \alpha e e^t \preceq 0,\ \alpha \geq 0$. Let $v^T e = 0$. Then $0 \geq v^T (\bar{X} - \alpha e e^t) v = v^T \bar{X} v$, i.e. $\mathcal{F}_0 \subset \mathcal{F}$. The final inclusion comes from the first and the fact that $\mathcal{F}$ is closed. ■

Corollary 1 *Let*

$$\begin{array}{rcl}
\mathcal{E} & := & \{X \in \mathcal{S}_H : v^T e = 0 \quad \Rightarrow \quad v^T X v \leq 0\}, \\
\mathcal{E}_0 & := & \{X \in \mathcal{S}_H : X - \alpha e e^t \preceq 0, \quad \textit{for some } \alpha\}, \\
\mathcal{E}_1 & := & \{X \in \mathcal{S}_H : X - \alpha e e^t \preceq 0, \quad \forall\ \alpha \geq \bar{\alpha}, \\
 & & \qquad \textit{for some } \bar{\alpha}\}.
\end{array}$$

Then

$$\mathcal{E} = \mathcal{E}_0 = \mathcal{E}_1. \tag{2}$$

Proof. (Similar to Lemma 1.) For closure, suppose $0 \neq X_k \in \mathcal{E}_0$, i.e. $\operatorname{diag}(X_k) = 0, X_k \preceq \alpha_k E$, for some α_k; and, suppose $X_k \to \bar{X}$. Since X_k is hollow it has exactly one positive eigenvalue which must be smaller than α_k. However, since X_k converges to $\bar{X}$, $\bar{X} \leq \lambda_{\max}(\bar{X})E$, where $\lambda_{\max}(\bar{X})$ is the largest eigenvalue of $\bar{X}$. ■

let: $E = ee^t$; $f(P) := \|H \circ (A - P)\|_F^2$;
$\mathcal{K}$ lin. operator with constraint $\operatorname{diag}(P) = 0$.
primal problem is:

$$
\text{(CDM)} \quad \begin{array}{rcc} \mu^* := & \min & f(P) \\ & \text{subject to} & \alpha E - P \succeq 0 \end{array}
$$

and **dual problem (DCDM)** is

$$
\begin{array}{rcr} \nu^* := & \max & f(P) + \langle y, b - KP \rangle - \operatorname{trace} \Lambda(\alpha E - P) \\ & \text{subject to} & \nabla_P f(P) - \mathcal{K}^* y + \Lambda = 0 \\ & & -\operatorname{trace} \Lambda E = 0 \\ & & \Lambda \succeq 0. \end{array}
$$

(Slater's holds for primal but fails for dual.)

(perturbed) Optimality Conditions are:

$$
\begin{array}{rl}
\operatorname{diag}(P) = 0 & \text{primal feas.} \\
\left.\begin{array}{r} 2H^{(2)} \circ (P - A) - \operatorname{Diag}(y) + \Lambda = 0 \\ -\operatorname{trace} \Lambda E = 0 \end{array}\right\} & \text{dual feas.} \\
-(\alpha E - P) + \mu \Lambda^{-1} = 0, & \text{pert. C.S.}
\end{array}
$$

$$
\begin{array}{rl}
h & \text{denotes the step for} \quad P \\
w & \text{denotes the step for} \quad \alpha \\
l & \text{denotes the step for} \quad \Lambda \\
s & \text{denotes the step for} \quad y.
\end{array}
$$

maintain

$$\operatorname{diag}(h) = \operatorname{diag}(P) = 0$$

linearization of complementary slackness

$$-(\alpha + w)E + (P + h) + \mu\Lambda^{-1} - \mu\Lambda^{-1} l \Lambda^{-1} \quad = \quad 0,$$

solve for h

$$h \quad = \quad -\mu\Lambda^{-1} + \mu\Lambda^{-1} l \Lambda^{-1} - P + (\alpha + w)E.$$

(or solve for l)

linearization dual feasibility

$$\begin{aligned} 2H^{(2)} \circ h - \operatorname{Diag}(s) + l \quad &= \quad -(2H^{(2)} \circ (P - A) \\ &\qquad - \operatorname{Diag}(y) + \Lambda) \\ -\operatorname{trace} lE \quad &= \quad \operatorname{trace} \Lambda E \end{aligned}$$

substitute for h, s

Newton equation is

$$
\begin{aligned}
2H^{(2)} \circ \left(wE + \mu\Lambda^{-1} l \Lambda^{-1}\right) - \operatorname{Diag}\operatorname{diag}(l) + l \\
= 2H^{(2)} \circ \left\{\mu\Lambda^{-1} + A - \alpha E\right\} + \operatorname{Diag}(y) - \Lambda \\
\operatorname{diag}\ \left(\mu\Lambda^{-1} l \Lambda^{-1}\right) + we \\
= \operatorname{diag}\ \left(\mu\Lambda^{-1}\right) - \alpha e \\
\operatorname{trace}(lE) = -\operatorname{trace}(\Lambda E).
\end{aligned}
$$

square system, order $1 + nnz$ where nnz are the number of nonzeros in the upper triangular part of H, ($\operatorname{diag}(H) = 0$).

F denotes $(nnz+n)\times 2$ matrix
row p contains indices of the p-th nonzero, upper triangular, element of $H+I$ ordered by columns,

$$\{(F_{p1}, F_{p2})_{p=1,\dots nnz+n}\} \\ = \{ij : H_{ij} \neq 0, i \leq j, \text{ ordered by columns}\}.$$

δ_{ij} is *Kronecker delta function*
$\delta_{(ij)(kl)}$ is 1 if $(ij)=(kl)$, 0 otherwise.

$E_{ij} = \left(e_i e_j^t + e_j^t e_i\right)/\sqrt{2}$, ij unit matrix in $\mathcal{S}^n$, where
$E_{ij} = \left(e_i e_j^t + e_j^t e_i\right)/2$ if $i=j$.
(orthonormal basis of $\mathcal{S}^n$)

operator equation:

$$
\begin{aligned}
& k \neq l, i \neq j \text{ LHS} = \\
& = \operatorname{trace} E_{kl} \left\{ 2H^{(2)} \circ \left(\mu \Lambda^{-1} E_{ij} \Lambda^{-1} \right) - \operatorname{Diag} \operatorname{diag} \left(E_{ij} \right) \right. \\
& \qquad\qquad\qquad \left. + E_{ij} \right\} \\
& = \mu \operatorname{trace} \left(e_k e_l^t + e_l e_k^t \right) \left(H^{(2)} \circ \Lambda^{-1} (e_i e_j^t + e_j e_i^t) \Lambda^{-1} \right) \\
& \qquad\qquad\qquad + \delta_{(ij)(kl)} \\
& = \mu \left[2e_l^t \left(H^{(2)} \circ \Lambda^{-1}_{:,i} \Lambda^{-1}_{j:} \right) e_k + 2e_k^t \left(H^{(2)} \circ \Lambda^{-1}_{:,i} \Lambda^{-1}_{j:} \right) e_l \right] \\
& \qquad\qquad + \delta_{(ij)(kl)}; \\
& k \neq l, i \neq j \text{ LHS} = \\
& = 2\mu H^{(2)}_{kl} \left(\Lambda^{-1}_{li} \Lambda^{-1}_{jk} + \Lambda^{-1}_{ki} \Lambda^{-1}_{jl} \right) + \delta_{(ij)(kl)}
\end{aligned}
$$

$k \neq l, i = j$ LHS $=$

$$= \operatorname{trace} E_{kl} \left\{ 2\mu H^{(2)} \circ \left[\Lambda^{-1} E_{jj} \Lambda^{-1} \right] - \operatorname{Diag} \operatorname{diag}\left(E_{jj}\right) + E_{jj} \right\}$$

$$= 2\sqrt{2}\mu \operatorname{trace} e_k e_l^t \left(H^{(2)} \circ \Lambda^{-1} e_j e_j^t \Lambda^{-1} \right)$$

$$= 2\sqrt{2}\mu H_{kl}^{(2)} \left(\Lambda_{lj}^{-1} \Lambda_{jk}^{-1} \right);$$

$k = l, i \neq j$ LHS $=$

$$= \sqrt{2}\mu \Lambda_{ki}^{-1} \Lambda_{jk}^{-1}, \quad k = 1, \ldots n;$$

$k = l, i = j$ LHS $=$

$$= \mu \Lambda_{ki}^{-1} \Lambda_{ik}^{-1}, \quad k = 1, \ldots n.$$

last column of LHS, matrix $l = 0$ and $w = 1$:

$$\begin{aligned} w = 1, k \neq l \text{ LHS} = \\ \text{trace } \left(E_{kl}(2H^{(2)} \circ E) \right); \\ w = 1, k = l \text{ LHS} = 1. \end{aligned}$$

last row of LHS:

$$\begin{aligned} i \neq j \text{ LHS} = &= \text{trace } (E_{ij}E)) = \sqrt{2}; \\ i = j \text{ LHS} = &= 1. \end{aligned}$$

Newton system is:

$$\mathrm{sMat}\,[L(\mathrm{svec}\,(l))] = \mathrm{sMat}\,[\mathrm{svec}\,(RHS)]\,,$$

svec (S) vector formed from nonzero elements of columns of upper triangular part, where strict upper triangular part is multiplied by $\sqrt{2}$. (trace $XY = \mathrm{svec}\,(X)^t \mathrm{svec}\,(Y)$, i.e. isometry) sMat is inverse

Solve for svec (l):

$$L(\mathrm{svec}\,(l)) = \mathrm{svec}\,(RHS).$$

$$L_{pq} =$$

$$\begin{cases} 2\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_1}} + \Lambda^{-1}_{F_{p_1},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_2}}\right) \\ \qquad \text{if } p \neq q,\ k \neq l,\ i \neq j; \\ 2\sqrt{2}\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_2}} \Lambda^{-1}_{F_{q_2},F_{p_1}}\right) \\ \qquad \text{if } p \neq q,\ k \neq l,\ i = j; \\ 2\sqrt{2}\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_2}} \Lambda^{-1}_{F_{q_2},F_{p_1}}\right) \\ \qquad \text{if } p = q,\ k \neq l,\ i = j; \\ 2\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_1}} + \Lambda^{-1}_{F_{p_1},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_2}}\right) + 1 \\ \qquad \text{if } p = q, k \neq l,\ i \neq j; \\ \sqrt{2}\mu \Lambda^{-1}_{F_{p_1},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_1}} \\ \qquad \text{if } k = l,\ i \neq j; \end{cases}$$

$=$

$$\begin{cases} \mu \Lambda^{-1}_{F_{p_1},F_{q_1}} \Lambda^{-1}_{F_{q_1},F_{p_1}} & \text{if } k = l, \ i = j \\ 2\sqrt{(}2) H^{(2)}_{F_{p_2},F_{p_1}} & \text{if } w = 1, \ k \neq l \\ 1 & \text{if } w = 1, \ k = l. \end{cases}$$

The p-th row calculated using Hadamard product of pairs of columns of Λ^{-1},

$$\Lambda^{-1}_{F_{p_2},F_{:,1}} \circ \Lambda^{-1}_{F_{p_1},F_{:,2}}.$$

complete vectorization (preliminary numerics are very promising)

$p = kl, k \leq l$, and last row, component of the right-hand-side of the system is

$RHS_p =$

$$\begin{cases} \sqrt{2}\left(2H_p^{(2)} \circ \{\mu\Lambda_p^{-1} + A_p - \alpha\} - \Lambda_p\right), & \text{if } k \neq l \\ \mu\Lambda_{kk}^{-1} - \alpha & \text{if } k = l \\ -\text{trace}\,(\Lambda E) & \text{last row} \end{cases}$$