EUCLIDEAN DISTANCE MATRIX COMPLETION PROBLEMS VIA SEMIDEFINITE PROGRAMMING

Henry Wolkowicz (University of Waterloo)
in collaboration with
Abdo Alfakih, Amir Khandani, Serge Kruk

Outline:

- SDP model for EDM
- Primal-dual interior-point (p-d i-p) method for SDP
- Gauss-Newton (G-N) search direction
- Elimination technique for G-N direction

Semidefinite Programming

looks just like

Linear Programming

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

$\preceq$ denotes the Löwner partial order
$A \preceq B$ if $B - A \succeq 0$

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

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

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

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

replaces

$\Re^n_+$ - nonnegative orthant

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 **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

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

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 the approximate EDMCP, let A be a pre-distance matrix, H be an $n \times n$ symmetric matrix with nonnegative elements, and let $\|A\|_F = \sqrt{\text{trace}\, A^t A}$ denote the *Frobenius norm* of A. Consider the objective function

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

where $\circ$ denotes *Hadamard product*.

weighted, closest Euclidean distance matrix problem

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

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 V $n \times (n-1)$, full column rank such that $V^t e = 0$. Then

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

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

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

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

Define the two linear operators

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

and

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

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

Theorem 1 *The linear operators satisfy*

$$\begin{aligned} \mathcal{K}(\mathcal{S}_C) &= \mathcal{S}_H, \\ \mathcal{T}(\mathcal{S}_H) &= \mathcal{S}_C, \end{aligned}$$

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

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.

Difficulty

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

$$V \cdot V^t : \mathcal{S}_{n-1} \to \mathcal{S}_n \quad \text{(into)}$$

$$V \cdot V^t : \mathcal{P}_{n-1} \to \mathcal{P}_n \quad \text{(onto face)}$$

Define the composite operators

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

and

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

Lemma 1

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

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

Corollary 1

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

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)

$$(CDM) \qquad \mu^* := \begin{array}{ll} \min & f(X) \\ \textbf{subject to} & \mathcal{A}X = b \\ & X \succeq 0. \end{array}$$

The additional constraint using $\mathcal{A} : \mathcal{S}_{n-1} \longrightarrow \Re^m$, could represent some of the fixed elements in the given matrix A.

Primal-Dual Interior-Point Framework:

Step 1 derive a dual program

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

Step 3 find a search direction for **solv**ing the **perturbed optimality conditions**

Step 4 take a step and backtrack to stay strictly feasible (positive definite)

Step 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)

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

or

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

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

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

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 2 *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.*

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

Lemma 2 *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.*

---

i.e. when $\mathcal{A} = 0$, we have a **Slater point** (now assumed for what follows) for the dual (and primal)

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

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

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

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

(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,* $\mu_k = \text{trace}\, X^k \Lambda^k / (n - 1)$

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

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

Step 3. find a search direction for solving the perturbed optimality conditions:

What is a good (local) search direction?
What linear system should we solve?
(Still an active area of research.)

We use the **Gauss-Newton direction**; the (Frobenius norm) lss of

$$F's = -F,$$

i.e.

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

Let $t(n) = \frac{(n+1)n}{2}$ be dimension of $\mathcal{S}_n$.
F' maps $\Re^{2(t(n-1))}$ to $\Re^{t(n-1)+(n-1)^2}$.

Motivation

In numerical analysis, how do we solve a nonlinear system of equations?

The standard approach is to solve the equivalent nonlinear least squares problem using "Gauss-Newton" method.

$$\min \frac{1}{2}||F_\mu(X, y, Z)||_2^2$$

subject to $X, Z > 0$.

Advantages: obvious merit function; always descent direction; direction always exists; known convergence analysis.

In the LP case (or symmetrized SDP case), when we differentiate we get a square system and Gauss-Newton reduces to Newton's method. (So G-N is always being used.)

i.e. solving

$$\left(F'_\mu\right)^t F'_\mu(\Delta s) = -\left(F'_\mu\right)^t F_\mu$$

(where F'_μ is the Jacobian)
is equivalent to solving

$$F'_\mu(\Delta s) = -F_\mu$$

for the search direction Δs.

Complications in the SDP case:

Overdetermined system, i.e. should we symmetrize ZX, e.g. $ZX + XZ$, many others ...

line search

SYMMETRIZATION

from log-barrier we get:
$X - \mu Z^{-1} = 0$
with linearization:
$X + \Delta X - \mu \left(Z^{-1} - Z^{-1} \Delta Z Z^{-1} \right) = 0$

more linear version - multiply by Z:
$ZX - \mu I = 0$
with linearization:
$ZX - \mu I + Z\Delta X + \Delta Z X = 0$

symmetrize:
$ZX + ZX - 2\mu I = 0$
with linearization:
$ZX + ZX - 2\mu I + Z\Delta X + \Delta X Z + \Delta Z X + X \Delta Z = 0$

Others:
$Z^{1/2} X Z^{1/2} - \mu I = 0$, etc...
$H_P(M) := \frac{1}{2} \left(PMP^{-1} + (PMP^{-1})^t \right)$
$M = XZ$ and P any nonsingular matrix

AHO direction $XZ + ZX$:
advantage: fast convergence and stability;
disadvantage: solve Lyapunov equation when finding Newton direction

HRVW-KSH-M direction: Heuristic; symmetrize ΔX after it is calculated:
advantages: fast convergence and inexpensive Newton direction, especially for max-cut problem!
disadvantage: stability

NT direction: adaptive scaling
advantage: self-scaling, symmetry
disadvantage: expensive

For a standard (linear) SDP, **finding the GN-direction** is equivalent to solving the normal equations

$$\left((F_\mu)'\right)^* (F_\mu)'(\Delta s) = -\left((F_\mu)'\right)^* (F_\mu)\,.$$

which is equivalent to finding the least squares solution of the overdetermined system

$$(F_\mu)'(\Delta s) = -(F_\mu)\,.$$

Or:

$$\begin{bmatrix} 0 & -\mathcal{A}^* & I \\ -\mathcal{A} & 0 & 0 \\ Z\cdot & 0 & \cdot X \end{bmatrix} \begin{pmatrix} \Delta X \\ \Delta y \\ \Delta Z \end{pmatrix} = \begin{pmatrix} -F_d \\ -F_p \\ -F_c \end{pmatrix}.$$

Use QR? Preconditioned CG?

Basic Elimination

Use block elimination and

$$\Delta Z = \mathcal{A}^*(\Delta y) - F_d$$

to obtain the $(m + n^2) \times (t(n) + m)$ reduced *dual feasible* system

$$\begin{bmatrix} -\mathcal{A} & 0 \\ Z\cdot & \mathcal{A}^* \cdot X \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta y \end{bmatrix} = \begin{bmatrix} -F_p \\ F_d X - F_c \end{bmatrix}.$$

write

$$\mathcal{A} = [\mathcal{A}_B \mathcal{A}_N],$$

$\mathcal{A}_B$ is a subset of size $m \times m$ of $\mathcal{A}$ that is easily invertible.

$$\Delta X_B = \mathcal{A}_B^{-1} F_p - \mathcal{A}_B^{-1} \mathcal{A}_N \Delta X_N,$$

which we substitute back into the dual feasible system to obtain the $(n^2 \times t(n))$ system

$$\left[\begin{array}{cc} -Z_B \mathcal{A}_B^{-1} \mathcal{A}_N + Z_N & \mathcal{A}^* \cdot X \end{array} \right] \left[\begin{array}{c} \Delta X_N \\ \Delta y \end{array} \right] = F_d X - F_c - Z_B \mathcal{A}_B^{-1} F_p.$$

Exploiting Special Structure

Consider the system of linear equations

$$Ax - By = c$$

The vectors x^*, y^* are (best) least squares solutions if $x^* = A^\dagger(By^* + c)$, where $A^\dagger$ is the Moore-Penrose inverse of A.

Therefore, we can eliminate x and solve the equivalent least squares problem with objective

$$||(I - AA^\dagger)(By + c)||_2^2$$

Adjoints? Generalized Inverses?

Let

$$\mathcal{Z}(\cdot) := Z\cdot$$

and

$$\mathcal{X}(\cdot) := \cdot X$$

Then the adjoint operators are

$$\mathcal{Z}^*(M) = \frac{1}{2}\left(M^t Z + ZM\right)$$

and

$$\mathcal{X}^*(M) = \frac{1}{2}\left(XM^t + MX\right).$$

$$\mathcal{Z}^*\mathcal{Z}(S) = \frac{1}{2}\left(SZ^2 + Z^2S\right)$$

and

$$\mathcal{X}^*\mathcal{X}(S) = \frac{1}{2}\left(X^2S + SX^2\right).$$

Both $\mathcal{Z}^*\mathcal{Z}(S)$ and $\mathcal{X}^*\mathcal{X}(S)$ are nonsingular operators.

$$\mathcal{Z}^\dagger(M) = \frac{1}{2}\left(M^tZ^{-1} + Z^{-1}M\right)$$

and

$$\mathcal{X}^\dagger(M) = \frac{1}{2}\left(X^{-1}M^t + MX^{-1}\right)$$

The orthogonal projections are:

$$\mathcal{X}\mathcal{X}^{\dagger}(M) = \frac{1}{2}\left(X^{-1}M^{t}X + M\right)$$

and

$$\mathcal{Z}\mathcal{Z}^{\dagger}(M) = \frac{1}{2}\left(ZM^{t}Z^{-1} + M\right).$$

Concluding Remarks

• **EDM:** Typically, low rank solutions were obtained for problems where no completion exists; (explanation using complementary slackness;)
purification to reduce rank.

• **GN-direction:** always exists; full rank Jacobian at each iteration and at solution; high accuracy in optimal solution; elimination approach work per iteration is competitive
($n = 42$, sec per iteration: 1000 cpu G-N $\rightarrow$ 200 cpu elimination).

• In progress: more tests; better rank reduction techniques.