EUCLIDEAN DISTANCE MATRIX COMPLETION PROBLEMS VIA SEMIDEFINITE PROGRAMMING: NUMERICAL CONSIDERATIONS

3 hot areas:

- Euclidean Distance Matrices
- Semidefinite Programming
- Interior Point Methods

Henry Wolkowicz (University of Waterloo)
in collaboration with
Abdo Alfakih, Miguel Anjos, 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 techinique for G-N direction
- Conjugate gradient approach using structure of linear operators

Semidefinite Programming
looks just like
Linear Programming

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

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

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

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

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

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

replaces

$\Re^n_+$ - nonnegative orthant

Why use SEMIDEFINITE PROGRAMMING?

Quadratic approximations are better than linear approximations.

Quadratic approximations are too hard to solve!

But, we can solve relaxations of quadratic approximations efficiently using semidefinite programming. This leads to MANY applications.

How are SDPs solved?

Primal-Dual Interior-Point Algorithms.

What are EDMs?

A *pre-distance matrix* (or dissimilarity matrix): an $n \times n$ symmetric matrix $D = (d_{ij})$ with non-negative 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 $\|A\|_F = \sqrt{\operatorname{trace} A^t A}$ denote the *Frobenius norm* of A. Consider the objective function

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

where $\circ$ denotes *Hadamard product*. The weighted, *closest Euclidean distance matrix problem* is

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

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

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

Define the two linear operators

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

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.

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

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

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

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

where $B = \mathcal{T}_V(A)$. ($\mathcal{K}_V$ and $\mathcal{T}_V$ are both linear operators)

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

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 solving 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)

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

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,
$f(X) - (f(X) + \langle y, b - \mathcal{A}(X)\rangle - \text{trace}\,\Lambda X)$,
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.*

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

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 for the dual (and primal)

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

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

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

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

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.

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

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

$$F's = \left[\begin{array}{cc} F'_{u1} & F'_{u2} \\ F'_{l1} & F'_{l2} \end{array} \right] \left(\begin{array}{c} h \\ l \end{array} \right) = rhs = \left(\begin{array}{c} rhs_1 \\ rhs_2 \end{array} \right).$$

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

Step 3. Gauss-Newton Search Direction:

For general SDP:

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

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.

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

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

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

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

The dual is obtained from the optimal strategy of the competing player, Y.

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

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

for the primal

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

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

$\mathcal{A}^* y - Z = C$ dual feasibility

$\mathcal{A} X = b$ primal feasibility

$Z X = 0$ complementary slackness

$Z X = \mu I$ perturbed, $\mu > 0$

Forms the basis for:

primal simplex method
dual simplex method
interior point methods

INTERIOR-POINT METHODS
for
LP and SDP

started with Karmarkar 1984

primal-dual interior-point methods:

1. potential reduction

2. path following

3. infeasible-interior-point

4. predictor-corrector

(see e.g. recent book by Steve Wright, or interior-point archives)

Connection between Karmarkar's algorithm and log-barrier problems (Stanford gang of 4: Gill, Murray, M. Wright, Saunders)

$$
(\mathbf{D}) \qquad \begin{array}{rcc} d^* = & \text{min} & b^t y \\ & \text{s.t.} & \mathcal{A}^* y - Z = C \\ & & Z \succeq 0 \end{array}
$$

log-barrier problem for dual SDP, with fixed $\mu > 0$

$$
\begin{array}{ll} \text{min} & b^t y - \mu \log\det Z \\ \text{s.t.} & \mathcal{A}^* y - Z = C \end{array}
$$

and $\mu \downarrow 0$

Lagrangian for log-barrier

$$
\min b^t y + \operatorname{trace} X\,(C - \mathcal{A}^* y + Z) - \mu \log\det Z
$$

Optimality conditions for the
dual log-barrier
solve $F_\mu(X, y, Z) = 0$, with $X, Z \succ 0$

$\frac{\partial}{\partial X}$: $C - \mathcal{A}^* y + Z = 0$ dual feasibility

$\frac{\partial}{\partial y}$: $b - \mathcal{A}X = 0$ primal feasibility

$\frac{\partial}{\partial Z}$: $X - \mu Z^{-1} = 0$ pert. compl. slack.

or, to avoid ill-conditioning, use:

$ZX - \mu I = 0$ perturbed complementary slackness

Note:
the nonlinear system $F_\mu(X, y, Z) = 0$ is overdetermined if (more linear) $ZX - \mu I = 0$ is used.
(this is **not** the case for LP)

CENTRAL PATH

For each $\mu > 0$, there exists a unique (X_μ, y_μ, Z_μ), with $X_\mu, Z_\mu \succ 0$, that solves the nonlinear system $F_\mu(X, y, Z) = 0$.

$$\mathcal{C} = \{(X_\mu, y_\mu, Z_\mu) : \mu > 0\}$$

well defined,
converges to optimum as $\mu \downarrow 0$,
long steps possible in neighbourhoods near $\mathcal{C}$

p-d i-p methods followed:
Megiddo 1987, Kojima, Mizuono, and Yoshise 1987, many others ....

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 $X, Z > 0$.

In the LP case, when we differentiate we get a square system and Gauss-Newton reduces to Newton's method.

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.

framework p-d i-p:
Given $(X^0, y^0, Z^0) \in \mathcal{F}^0$
for $k = 0, 1, 2 \ldots$
solve for the search direction

$$F'_\mu(X^k, y^k, Z^k) \begin{pmatrix} \Delta X^k \\ \Delta y^k \\ \Delta Z^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k Z^k + \sigma_k \mu_k I \end{pmatrix}$$

where σ_k centering, $\mu_k = \operatorname{trace} X^k Z^k / n$

$$\begin{aligned}(X^{k+1}, y^{k+1}, Z^{k+1}) = \\ (X^k, y^k, Z^k) + \alpha_k(\Delta X^k, \Delta y^k, \Delta Z^k)\end{aligned}$$

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

Complications in the SDP case:

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

line search

strict complementarity may fail

duality gaps

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

Still many open questions, e.g. why symmetrize?
Can apply Gauss-Newton method to solving overdetermined system of nonlinear equations.

"best direction": Newton direction for least squares problem?

best compromise: Gauss-Newton direction

other directions: are approximations of the Gauss-Newton direction

Summary:
two ways to avoid ill-conditioning of log-barrier problem.

1. Increase the dimension of the problem by one - still finds the Newton direction .

2. Change the optimality conditions (perturb). (In which case, direction based solely on the optimality conditions and not on a log-barrier problem
(not a Newton direction)

Alternate motivation for changing true log-barrier direction:
$X - \mu Z^{-1} = 0$
with linearization:
$X + \Delta X - \mu\left(Z^{-1} - Z^{-1}\Delta Z Z^{-1}\right) = 0$

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

MORE LINEARITY
SPEEDS UP NEWTON'S METHOD
but
results in overdetermined system

(now solve using G-N or symmetrize)

Shrinking the Overdetermined System

We can find redundant equations and solve a smaller system. We can remove the lower triangular part from the nonlinear equation and get the equivalent equation

$$\mathcal{T}\left(ZX - \mu I\right) = 0,$$

where the linear operator $\mathcal{T} : \mathcal{M}_n \to \mathcal{S}_n$ by ignoring the strictly lower triangular part of the matrix, i.e. the i,j components are

$$(\mathcal{T}(W))_{ij} = \begin{cases} W_{ij} & \text{if } i \leq j \\ W_{ji} & \text{if } i > j \end{cases}$$

The resulting system is now square.

$\mathcal{T}$ is an orthogonal projection

The optimality conditions become

$$\begin{pmatrix} Z + C - \mathcal{A}^* y \\ b - \mathcal{A}(X) \\ \mathcal{T}(ZX - \mu I) \end{pmatrix} = 0.$$

This does not introduce new nonlinearities but does guarantee that we map between the same spaces. In fact, we can prove that we do not lose information in the optimality conditions when we only consider the upper triangular parts.

LEMMA
Suppose that $Z \succ 0$. Then the linear operator $\mathcal{T}_Z(\cdot) := \mathcal{T}(Z\cdot)$ is a one-one mapping on $\mathcal{S}_n$. In addition, if $\mathcal{T}_Z(X) = \mu I$, for some $X = X^t$, then $ZX = \mu I$.

The Newton direction is obtained from solving the system

$$\begin{aligned} \Delta Z - \mathcal{A}^*(\Delta y) &= -F_d \\ -\mathcal{A}(\Delta X) &= -F_p \\ \mathcal{T}(Z\Delta X + \Delta Z X) &= -\mathcal{T}(ZF_{Z,X}). \end{aligned}$$

Equivalently,

$$\begin{bmatrix} 0 & -\mathcal{A}^* & I \\ -\mathcal{A} & 0 & 0 \\ \mathcal{T}_Z(\cdot) & 0 & \mathcal{T}(\cdot X) \end{bmatrix} \begin{pmatrix} \Delta X \\ \Delta y \\ \Delta Z \end{pmatrix} = \begin{pmatrix} -F_d \\ -F_p \\ -\mathcal{T}_Z(X) + \mu I \end{pmatrix}.$$

Back to Gauss-Newton direction

Equivalently, we are solving the system of non-linear equations

$$F_\mu(X,y,Z) = \begin{pmatrix} Z + C - \mathcal{A}^* y \\ b - \mathcal{A}(X) \\ ZX - \mu I \end{pmatrix} =: \begin{pmatrix} F_d \\ F_p \\ F_c \end{pmatrix} = 0,$$

Solve the nonlinear least squares problem:

$$R^* = \min_s f_\mu(s) := \frac{1}{2} ||F_\mu(s)||_2^2.$$

The optimal value R^* is called the residual.

G-N direction – drop second order terms from the Newton direction.

Let $F_c = F_c(X, Z) = ZX - \mu I$.

The second order terms (curvature) are:

$$\left\langle F_c(Z, X), F_c''(Z, X)(\cdot, \cdot)\right\rangle.$$

If changes in Z, X are l, h respectively, then

$$(Z + l)(X + h) - \mu I = ZX - \mu I + Zl - lX + lh.$$

The curvature in this direction is therefore:

$$\operatorname{trace}(ZX - \mu I)lh.$$

dropped terms from the Newton direction

$$H := \begin{bmatrix} 0 & (ZX - \mu I) \\ (ZX - \mu I) & 0 \end{bmatrix}$$

Finding the GN-direction is therefore 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 get

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

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

$$\begin{bmatrix} -Z_B \mathcal{A}_B^{-1} \mathcal{A}_N + Z_N & \mathcal{A}^* \cdot X \end{bmatrix} \begin{bmatrix} \Delta X_N \\ \Delta y \end{bmatrix} = F_d X - F_c \\ - Z_B \mathcal{A}_B^{-1} F_p.$$

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

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

Back to Euclidean Distance Matrices

Recall:

$$
\begin{aligned}
\mathcal{K}(B) &:= \operatorname{diag}(B)\, e^t + e \operatorname{diag}(B)^t - 2B, \\
\mathcal{K}^*(D) &= 2\left(\operatorname{Diag}(De) - D\right) \\
\mathcal{K}_V(X) &:= \mathcal{K}(VXV^t), \\
\mathcal{K}_V^*(D) &= V^t \mathcal{K}^*(D) V
\end{aligned}
$$

The linearized system is

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

solve system using elimination

$$l = 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(h) \right) + F_d$$

find least squares solution of *overdetermined* system in h

$$\Lambda h + \left\{ 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(h) \right) \right\} X = -F_c - F_d X.$$

Exploit Structure? Solve

$$\Lambda(h) + \mathcal{X}((\mathcal{L}(h))V) = r.$$

where $\mathcal{L}(h) := 2V^t\mathcal{K}^*(H^{(2)} \circ \mathcal{K}_V(h))$

i.e. solve system for vector x

$$(C + DEF)x = b$$

where C, D, E, F are sparse but EF is dense.

(V has a completely dense row)

Sparse V

$$V := \frac{1}{\sqrt{2}} \begin{bmatrix} I \\ -e^t \end{bmatrix}.$$

i.e. $V^t e = 0$, $V^\dagger V = I$, and $VV^\dagger = J$

Columns of Operators with $i \neq j$

Now, for $i \neq j$ we have

$$\begin{aligned}
\mathcal{K}_V(E_{ij}) &= \mathcal{K}(V_i V_j^t + V_j V_i^t) \\
&= \mathrm{diag}\,(V_i V_j^t + V_j V_i^t) e^t \\
&\qquad + e\,\mathrm{diag}\,(V_i V_j^t + V_j V_i^t)^t \\
&\qquad - 2(V_i V_j^t + V_j V_i^t).
\end{aligned}$$

$$\begin{aligned}
V_i V_j^t &= \frac{1}{2} \begin{pmatrix} e_i \\ -1 \end{pmatrix} \begin{pmatrix} e_j^t & -1 \end{pmatrix} \\
&= \frac{1}{2} \begin{bmatrix} e_i e_j^t & -e_i \\ -e_j^t & 1 \end{bmatrix}.
\end{aligned}$$

Thus

$$\mathrm{diag}\,(V_i V_j^t) = \frac{1}{2} e_n$$

and

$$\mathrm{diag}\,(V_i V_j^t + V_j V_i^t) = e_n.$$

So

$$\mathrm{diag}\,(V_i V_j^t + V_j V_i^t) e^t = E_n,$$

$$\begin{aligned}\operatorname{diag}(V_iV_j^t + V_jV_i^t)e^t + e\operatorname{diag}(V_iV_j^t + V_jV_i^t)^t \\ = E_n + E_n^t \\ = \quad \begin{bmatrix} 0 & e \\ e^t & 2 \end{bmatrix}.\end{aligned}$$

Now calculate

$$\begin{aligned}\mathcal{K}_V(E_{ij}) &= \operatorname{diag}(V_iV_j^t + V_jV_i^t)e^t \\ &+ \quad e\operatorname{diag}(V_iV_j^t + V_jV_i^t)^t \\ &- \quad 2(V_iV_j^t + V_jV_i^t) \\ &= \begin{bmatrix} 0 & e \\ e^t & 2 \end{bmatrix} - \begin{bmatrix} E_{ij} & -e_i - e_j \\ (-e_i - e_j)^t & 2 \end{bmatrix}.\end{aligned}$$

Therefore, with

$h = H^{(2)}_{:,1:n-1},\ h_i = (H_{in})^2 e_i,\ H^{(2)}_{ij} =$ ...submatrix...,

we get the simple, sparse, expression

$$H^{(2)} \circ \mathcal{K}_V(E_{ij}) = \begin{bmatrix} -H^{(2)}_{ij} & h + h_i + h_j \\ (h + h_i + h_j)^t & 0 \end{bmatrix}.$$

Continuing, we let

$$d = \text{Diag}\left(\ h + h_i + h_j - (H_{ij})^2 e_i - (H_{ij})^2 e_j\ \right),\ \ a = h^t e + (H_{ni})^2 + (H_{nj})^2.$$

$$\begin{array}{rcl} \mathcal{K}^*(H^{(2)} \circ \mathcal{K}_V(E_{ij})) & = & 2\text{Diag}\left[\left(H^{(2)} \circ \mathcal{K}_V(E_{ij})\right) e\right] - 2H^{(2)} \circ \mathcal{K}_V(E_{ij}) \\ & = & -2\left[\begin{array}{cc} -H^{(2)}_{ij} - d & h + h_i + h_j \\ (h + h_i + h_j)^t & -a \end{array}\right]. \end{array} \tag{1}$$

Define the linear operator $\mathcal{L} : \mathcal{S}_{n-1} \to \mathcal{M}_{n-1,n}$ with

$$\mathcal{L}(h) := 2V^t \mathcal{K}^*(H^{(2)} \circ \mathcal{K}_V(h)) \tag{2}$$

and

$$\mathcal{Q}(h) := 2V^t \mathcal{K}^*(H^{(2)} \circ \mathcal{K}_V(h))V = 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(h)). \tag{3}$$

Note that $\mathcal{Q}$ is self-adjoint. Finally we evaluate

$$\begin{array}{rcl} \mathcal{L}(E_{ij}) & = & -4V^t \left[\begin{array}{cc} -H^{(2)}_{ij} - d & h + h_i + h_j \\ (h + h_i + h_j)^t & -a \end{array}\right] \\ & = & -2\sqrt{2}\left[\ -H^{(2)}_{ij} - d - e\left(h + h_i + h_j\right)^t \quad h + h_i + h_j + ae\ \right] \end{array} \tag{4}$$

$$
\begin{aligned}
\mathcal{Q}(E_{ij}) &= \mathcal{L}(E_{ij})V \\
&= -2\left[\ -H^{(2)}_{ij} - d - e\left(h + h_i + h_j\right)^t - \left(h + h_i + h_j + ae\right)e^t\ \right] \\
&= -2\left[\ -H^{(2)}_{ij} - d - aee^t - e\left(h + h_i + h_j\right)^t - \left(h + h_i + h_j\right)e^t\ \right]
\end{aligned}
\tag{5}
$$

The left hand side of the overdetermined linear system is

$$
\begin{aligned}
\wedge E_{ij} + \mathcal{Q}(E_{ij})X &= \wedge E_{ij} - 2\left[\ -H^{(2)}_{ij} - d - aee^t - e\left(h + h_i + h_j\right)^t - \left(h + h_i + h_j\right)e^t\ \right]X \\
&= \wedge E_{ij} + 2H^{(2)}_{ij} \circ E_{ij}X???split - into - fixed - n - variable
\end{aligned}
\tag{6}
$$

Now

$$VX = \begin{bmatrix} X \\ e^t X \end{bmatrix}, \; VX^2 = \begin{bmatrix} X^2 \\ e^t X^2 \end{bmatrix}.$$

Also,

$$IT(VX) = \left((VX)^t \otimes I\right) \text{vec}(T)$$

We define the linear operators

$$\Omega(\cdot) := \Lambda\cdot, \; \mathcal{X}(\cdot) := \cdot X.$$

Then the adjoint operators are

$$\Omega^*(M) = \frac{1}{2}\left(M^t\Lambda + \Lambda M\right)$$

and

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

Now

$$\begin{aligned}
\mathcal{X}^*\mathcal{X}(\mathcal{Q}(h)) &= \mathcal{X}^*(\mathcal{Q}(h)X) \\
&= \mathcal{X}^*(\mathcal{L}(h)VX) \\
&= \frac{1}{2}\left(X(\mathcal{L}(h)VX)^t + \mathcal{L}(h)VX)X\right) \\
&= \frac{1}{2}\left(X^2V^t(\mathcal{L}(h))^t + \mathcal{L}(h)VX^2\right) \\
&= \frac{1}{2}\left(X^2\mathcal{Q}(h) + \mathcal{Q}(h)X^2\right),
\end{aligned}$$

since $\mathcal{Q}(h)$ is symmetric.

We need to *efficiently* calculate the following part of the normal equation, i.e. exploit sparsity or the structure of, $V^t\mathcal{L}^*\left(\mathcal{X}^*\mathcal{X}(\mathcal{L}(E_{ij})V)\right)$.

Let's look at

$$V^t \mathcal{L}^* \left(\mathcal{L}(E_{ij}) V X^2 \right) \;=\; \mathcal{Q}^* \left(\mathcal{Q}(E_{ij}) X^2 \right)$$

Let x^k be the multiple of the k-th column of X^2 so that

$$H_{ij}^{(2)} X^2 = x^i e_j^t + x^j e_i^t.$$

We first look at

$$\begin{aligned}
\mathcal{K}_V \left(x^i e_j^t \right) &= \mathcal{K} \left(V x^i e_j^t V^t \right) \\
&= \frac{1}{2} \mathcal{K} \left(\begin{pmatrix} x^i \\ e^t x^i \end{pmatrix} \begin{pmatrix} e_j \\ 1 \end{pmatrix}^t \right) \\
&= \frac{1}{2} \left\{ \mathrm{diag}\,(V x^i V_j^t) e^t + e\,\mathrm{diag}\,(V x^i V_j^t)^t - 2(V x^i V_j^t) \right\} \\
&= \frac{1}{2} \left\{ \begin{pmatrix} (x^i)_j e_j \\ e^t x^i \end{pmatrix} + \begin{pmatrix} (x^i)_j e_j \\ e^t x^i \end{pmatrix}^t \right\} - 2 \left(\begin{pmatrix} x^i \\ e^t x^i \end{pmatrix} \begin{pmatrix} e_j \\ 1 \end{pmatrix}^t \right)
\end{aligned}$$