Summary

objective function

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

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

(Re)Define the closest EDM problem

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

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

Lagrangian

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

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

primal program (CDM)

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

dual program

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

Dual DCDM

Wolfe Dual

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

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

or

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

Duality Gap

Using complementary slackness

$f(X) - (f(X) + \langle y, b - \mathcal{A}(X)\rangle - \text{trace}\, \Lambda X)$,
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 $H^{(2)} = H \circ H$.

Optimality

Primal-Dual Conditions - Theorem

Suppose that Slater's condition holds. Then the p-d pair $\bar{X} \succeq 0$, and $\bar{y}$, $\bar{\Lambda} \succeq 0$ solve (CDM) and (DCDM), respectively, if and only if

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

Slater point for connected graph

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: (perturbed) optimality conditions for log-barrier problem

take $\Lambda = \mu X^{-1}$ and multiplying through by X to get the (overdetermined) optimality conditions, $F := \begin{pmatrix} F_d \\ F_c \end{pmatrix} = 0,$ or

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

estimate of the barrier parameter

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

Algorithm

p-d i-p paradigm

σ_k centering param.; $\mathcal{F}^0$ strictly feasible p-d set
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 search direction

Gauss-Newton Direction

What is a good (local) search direction/linearization?
(Still an active area of research.)
We use the **Gauss-Newton direction** (Frobenius norm),

$$F's = -F,$$

i.e. lss of linearization

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

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

Motivation

solve a nonlinear system of equations In numerical analysis

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.

GN Method

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

should we symmetrize ZX, e.g. $ZX + XZ$, many others ...
use a line search?

SYMMETRIZATION

use log-barrier $X - \mu Z^{-1} = 0$, with linearization

$X + \Delta X - \mu \left(Z^{-1} - Z^{-1} \Delta Z Z^{-1} \right) = 0$

less nonlinear version: $ZX - \mu I = 0$, with linearization

$ZX - \mu I + Z \Delta X + \Delta Z X = 0$

symmetrizations

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

Search Directions

AHO direction

$XZ + ZX$:
advantage: fast convergence and stability;
disadvantage: solve Lyapunov equation

HRVW-KSH-M direction

Heuristic: symmetrize ΔX after it is calculated:
advantages: fast convergence, inexpensive
disadvantage: instability

NT direction; adaptive scaling

advantage: self-scaling, symmetry
disadvantage: expensive, instability

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

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

Block Elimination

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.

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

substitute back into dual feasible system; obtain $(n^2 \times t(n))$

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

$$= F_d X - F_c - Z_B \mathcal{A}_B^{-1} F_p$$

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

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

More Adjoints

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

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; however expensive