Algorithms for linear SDP

Michal Kočvara

Institute of Information Theory and Automation

Academy of Sciences of the Czech Republic

and

Czech Technical University

kocvara@utia.cas.cz

http://www.utia.cas.cz/kocvara

SDP notations

$\mathbb{S}^n$ … symmetric matrices of order $n \times n$
$A \succeq 0$ … A positive semidefinite
$A \succeq B$ … $A - B \geq 0$
$\langle A, B \rangle :=$ Trace(AB) … inner product on $\mathbb{S}^n$
$\mathcal{A}[\mathbb{R}^m \to \mathbb{S}^n]$ … linear matrix operator defined by

$$\mathcal{A}(y) := \sum_{i=1}^{m} y_i A_i \quad \text{with } A_i \in \mathbb{S}_n$$

$\mathcal{A}^*[S^n \to \mathbb{R}^m]$ … adjoint operator to $\mathcal{A}$ defined by

$$\mathcal{A}^*(X) := [\langle A_1, X \rangle, \ldots, \langle A_m, X \rangle]^T$$

and satisfying

$$\langle \mathcal{A}^*(X), y \rangle = \langle \mathcal{A}(y), X \rangle \qquad \text{for all } y \in \mathbb{R}^m$$

Primal-Dual SDP pair

$$\inf_{X} \langle C, X\rangle := \text{Trace}(CX) \qquad \text{(P)}$$
$$\text{s.t.} \quad \mathcal{A}^*(X) = b \qquad [\langle A_i, X\rangle = b_i,\ i = 1, \ldots, m]$$
$$X \succeq 0$$

$$\inf_{y,S} \langle b, y\rangle := \sum b_i y_i \qquad \text{(D)}$$
$$\text{s.t.} \quad \mathcal{A}(y) + S = C \qquad [\sum y_i A_i + S = C]$$
$$S \succeq 0$$

Weak duality: Feasible X, y, S satisfy

$$\langle C, X\rangle - \langle b, y\rangle = \langle \mathcal{A}(y) + S, X\rangle - \sum y_i \langle A_i, X\rangle = \langle S, X\rangle \geq 0$$

duality gap **nonnegative** for feasible points.

SDP duality

Points with zero duality gap

$$dg := \langle C, X \rangle - \langle b, y \rangle = \langle S, X \rangle = 0$$

are optimal.

LP: (P)/(D) has optimal solution $\Rightarrow$(D)/(P) has opt. sol. and $dg = 0$
. . . strong duality

SDP: Strong duality under Slater Constraint Qualification: (SCQ)

$$\exists X \succ 0 \quad \text{and} \quad S \succ 0 .$$

Without SCQ (examples):
(P) solvable, (D) infeasible
$dg > 0$ at optimality
etc

SDP Optimality Conditions (1st order)

Theorem: Under SCQ, necessary and sufficient optimality conditions for (P) and (D) are

$$\mathcal{A}^*(X) = b\,, \quad X \succeq 0$$
$$\mathcal{A}(y) + S = C\,, \quad S \succeq 0$$
$$XS = 0\,.$$

Note: $\langle X, S \rangle = 0 \Leftrightarrow XS = 0$ since $X \succeq 0,\ S \succeq 0$.

Logarithmic barrier methods

Primal Log-Barrier method:

For $\mu \searrow 0$ solve

$$\min_{X} \langle C, X\rangle - \mu \log\det(X)$$
$$\text{s.t.} \quad \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m$$

Dual Log-Barrier method:

For $\mu \searrow 0$ solve

$$\min_{y,S} \langle b, y\rangle - \mu \log\det(S)$$
$$\text{s.t.} \quad \mathcal{A}(y) + S = C$$

Logarithmic barrier methods

Primal-Dual Log-Barrier method:

Minimize the duality gap

$$\langle C, X\rangle - \langle b, y\rangle = \langle S, X\rangle$$

using primal-dual barrier function

$$-(\log\det(X) + \log\det(S)) = -\log\det(XS)$$

For $\mu \searrow 0$ solve

$$\min_{X,y,S} \langle X, S\rangle - \mu \log\det(XS)$$

s.t. $\langle A_i, X\rangle = b_i, \quad i = 1, \dots, m$

$$\mathcal{A}(y) + S = C$$

Central path

Perturb OC by $\mu > 0$:

$$\boxed{\begin{array}{rl} \mathcal{A}^*(X) = b\,, & X \succeq 0 \\ \mathcal{A}(y) + S = C\,, & S \succeq 0 \\ XS = \mu I\,. & \end{array}} \qquad (\mathrm{OC}_\mu)$$

Theorem: System (OC$_\mu$) has a unique solution.

Proof: Consider the primal log-barrier problem

$\min_{X \succ 0} \left\{ f_p^\mu := \frac{1}{\mu}\langle C, X\rangle - \log\det(X) \mid \mathcal{A}^*(X) = b \right\}$.

Function f_p^μ is strictly convex. The KKT conditions for this problem are

$$\nabla f_P^\mu := \frac{1}{\mu} C - X^{-1} = \sum \hat{y}_i A_i$$

$$\mathcal{A}^*(X) = b, X \succ 0$$

Def. $S = C - \sum y_i A_i$ where $y_i = \mu \hat{y}_i$ to get (OC)$_\mu$. If (D) strictly feasible, level-sets of f_p^μ are compact (without proof here). Hence there is a unique minimizer X^* of f_p^μ over ri (P) and it is a unique solution of (OC)$_\mu$ with $S := \mu(X^*)^{-1}$.

Central path

Definition: The curve defined by solutions $(X(\mu), S(\mu), y(\mu))$ of $(\text{OC})_\mu$ is called *central path*.

Theorem: The central path exists if (P) and (D) are strictly feasible.

Theorem: The pair

$$X^* = \lim_{\mu \searrow 0} X(\mu), \qquad S^* = \lim_{\mu \searrow 0} S(\mu)$$

is a maximally complementary solution pair (matrices with highest rank).

Primal-dual path-following methods

Given $\mu > 0$, the pair $(X(\mu), S(\mu))$ is the target point on the central path, associated with target duality gap $\langle X, S \rangle = n\mu$.

Idea: iteratively compute approximations of $X(\mu), S(\mu)$ and thus follow the central path while decreasing μ.
Assume $X \succ 0$, $S \succ 0$, solve the OC for the P-D problem

$$\begin{aligned} \mathcal{A}^*(X) &= b \\ \mathcal{A}(y) + S &= C \\ XS &= \mu I \end{aligned}$$

by the Newton method:

Newton method for

$$\begin{aligned} \mathcal{A}^*(X) &= b \\ \mathcal{A}(y) + S &= C \\ XS &= \mu I \end{aligned}$$

Find $\Delta X, \Delta S, \Delta y$:

$$\begin{aligned} (i) \quad & \langle A_i, \Delta X \rangle = R_p := b - \langle A_i, X \rangle, \quad i = 1, \ldots, m \\ (ii) \quad & \mathcal{A}(\Delta y) + \Delta S = R_d := C - S - \mathcal{A}(y) \\ (iii) \quad & X\Delta S + \Delta X S = R_c := \mu I - XS \qquad (-\Delta X \Delta S) \end{aligned}$$

Remark: Solutions $\Delta S, \Delta X$ of (iii) generally nonsymmetric.

ΔS symmetric from (ii) but ΔX may be nonsymmetric.

Symmetrization of (iii) needed.

Symmetrization techniques

Replace $XS = \mu I$ by symmetrization

$$H_p(XS) = \mu I$$

where $H_p(M) = \frac{1}{2}(PMP^{-1} + P^{-T}M^TP^T)$.
Thus (iii) becomes

$$(iii)' \qquad H_P(\Delta XS + \Delta SX) = \mu I - H_p(XS)\,.$$

The scaling matrix P determines the symmetrization strategy.

P	reference
$[X^{\frac{1}{2}}(X^{\frac{1}{2}}SX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}}]^{\frac{1}{2}}$	Nesterov-Todd (NT)
$X^{-\frac{1}{2}}$	Monteiro and others
$S^{\frac{1}{2}}$	Monteiro, Helmberg at al., . . .
I	Alizadeh-Haeberly-Overton

NT direction

Recall $F(X) = \log\det(X)$ (the barrier function)

We require $F''(D)X = S$ for a scaling matrix D.

Direct computation: $D^{-1}XD^{-1} = S$ and thus

$$D^{-\frac{1}{2}}XD^{-\frac{1}{2}} = D^{\frac{1}{2}}SD^{\frac{1}{2}} := V$$

Note that $V^2 = D^{-\frac{1}{2}}XSD^{\frac{1}{2}} \sim XS$ (have the same eigvs.)

Thus $D = S^{-\frac{1}{2}}(S^{\frac{1}{2}}XS^{\frac{1}{2}})^{\frac{1}{2}}S^{-\frac{1}{2}}$

NT equation:

$$\boxed{\Delta X + D\Delta S D = \mu S^{-1} - X}$$

NT direction

NT equation

$$\Delta X + D\Delta S D = \mu S^{-1} - X$$

Can be written in the form (Todd-Toh-Tütüncü)

$$\mathcal{E}\Delta X + \mathcal{F}\Delta S = \mu S^{-1} - \Sigma$$

where

$$\mathcal{E} = P^{-T} \circledast PS, \quad \mathcal{F} = P^{-T}X \circledast P, \quad P^T P = D, \quad \Sigma = P^{-T} X P^{-1}$$

and

$$G \circledast H(M) := \frac{1}{2}(HMG^T + GMH^T)$$

is the symmetric Kronecker product.

Primal-Dual path-following algorithms

Define the centrality function

$$\delta(X,S,\mu) := \frac{1}{2}\|\sqrt{\mu}V^{-1} - \frac{1}{\sqrt{\mu}}V\| \qquad \left(V = D^{-\frac{1}{2}}XD^{-\frac{1}{2}} = D^{\frac{1}{2}}SD^{\frac{1}{2}}\right)$$

Note:

$$\delta(X,S,\mu) = 0 \Longleftrightarrow V^2 = \mu I \Longleftrightarrow XS = \mu I$$

Primal-Dual path-following algorithms

Input: $(X^0, S^0) \in \mathcal{P} \times \mathcal{D}$

Parameters: $\tau < 1$ and $\mu_0 > 0$ such that $\delta(X^0, S^0, \mu_0) \leq \tau$

Algorithm (generic): Set $X := X^0, S := S^0, \mu := \mu_0$

while $\langle XS \rangle > \varepsilon$ **do**

compute $\Delta X, \Delta S$ by solving $(i), (ii), (iii)'$
choose a steplength $\alpha \in (0, 1]$
$X := X + \alpha \Delta X, \quad S := S + \alpha \Delta S$
choose $0 < \theta < 1$ and set $\mu := (1 - \theta)\mu$

end

Denote
$(X^+, S^+) := (X + \Delta X, S + \Delta S) \ldots$ full NT step
$(X_\alpha, S_\alpha) := (X + \alpha \Delta X, S + \alpha \Delta S) \ldots$ damped NT step

Primal-Dual path-following algorithms

Lemma: If $\delta(X, S, \mu) < \frac{1}{\sqrt{2}}$ then $\delta(X^+, S^+, \mu) < \delta^2(X, S, \mu)$

…quadratic convergence to the $\mu-$center (near the path).

Theorem: If $\tau = \frac{1}{\sqrt{2}}$ and $\theta = \frac{1}{2\sqrt{n}}$, then the primal-dual path-following algorithm with full NT steps terminates after at most

$$O\left(2\sqrt{n}\log\frac{n\mu^o}{\varepsilon}\right)$$

iterations.

Long-step p-d path-following algorithm

Input: $\tau > 0 \ldots$ centering parameter
$(X^0, S^0) \in \text{ri}\mathcal{P} \times \mathcal{D}$
$\mu_0 > 0$ such that $\delta(X^0, S^0, \mu_0) \le \tau$
$\theta < 1 \ldots$ updating parameter

Algorithm (long-step): Set $X := X^0, S := S^0, \mu := \mu_0$

while $\langle XS \rangle > \varepsilon$ **do**

  if $\delta(X, S, \mu) \le \tau$ **do**

    $\mu := (1 - \theta)\mu$

  else if $\delta(X, S, \mu) > \tau$ **do**

    compute $\Delta X, \Delta S$ by solving $(i), (ii), (iii)'$
    find $\alpha := \arg\min f_{pd}^{\mu}(X + \alpha \Delta X, S + \alpha \Delta S, \mu)$
    $X := X + \alpha \Delta X, \quad S := S + \alpha \Delta S$

  end

end

Long-step p-d path-following algorithm

Theorem: The long-step primal-dual path-following algorithm terminates after at most

$$\boldsymbol{O}\left(\boldsymbol{n}\log\frac{\boldsymbol{n}\boldsymbol{\mu}^{\boldsymbol{o}}}{\boldsymbol{\varepsilon}}\right)$$

iterations.

Solving the linear systems

Primal-dual system:

$$\begin{pmatrix} 0 & A^T & 0 \\ A & 0 & I \\ 0 & \mathcal{E} & \mathcal{F} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta X \\ \Delta S \end{pmatrix} \begin{pmatrix} R_p \\ R_d \\ R_c \end{pmatrix}$$

There exists a unique solution (Todd-Toh-Tütüncü).

Define $\mathcal{U} := \mathcal{F}^{-1}\mathcal{E}$ $(= D^{-1} \circledast D^{-1})$ and substitute

$$\Delta S = R_d - A^T \Delta y$$

to get

$$\begin{pmatrix} -\mathcal{U} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta X \\ \Delta y \end{pmatrix} \begin{pmatrix} \mathcal{R} := R_d - \mathcal{F}^{-1} R_c \\ R_p \end{pmatrix}$$

. . . **augmented system**

Solving the linear systems

Further eliminate ΔX

$$\Delta X = \mathcal{U}^{-1}(A^T \Delta y - R_d + \mathcal{F}^{-1} R_c)$$

to get

$$\boxed{\underbrace{A\mathcal{U}^{-1}A^T}_{M} \Delta y = h} := R_p + A\mathcal{U}^{-1} R_d - A\mathcal{E}^{-1} R_c$$

... **Schur complement equation** (normal eq.)

<u>**Most popular strategy:**</u> solve SCE by direct Cholesky factorization

M typically <u>fully dense</u> even if A_i sparse

Use sparse linear algebra to compute M

A dual scaling algorithm

S. Benson and Y. Ye $\longrightarrow$ code DSDP

$$\min_{y,S} \langle b, y\rangle - \mu \log\det(S)$$
$$\text{s.t.} \quad \mathcal{A}(y) + S = C$$

KKT conditions:

$$\mathcal{A}^*(X) = b, \ \mathcal{A}(y) + S = C \ \mu S^{-1} = X$$

The corresponding Newton system:

$$\mathcal{A}^*(X + \Delta X) = b$$
$$\mathcal{A}(\Delta y) + \Delta S = 0$$
$$\mu S^{-1} \Delta S S^{-1} + \Delta X = \mu S^{-1} - X$$

A dual scaling algorithm

$$\mathcal{A}^*(X + \Delta X) = b$$
$$\mathcal{A}(\Delta y) + \Delta S = 0$$
$$\mu S^{-1} \Delta S S^{-1} + \Delta X = \mu S^{-1} - X$$

From (i) and (iii):

$$-\mathcal{A}^*(S^{-1} \Delta S S^{-1}) = \frac{1}{\mu} b - \mu \mathcal{A}^*(S^{-1})$$

Substitute $\Delta S = \mathcal{A}(\Delta y)$ (from (ii)):

$$\mathcal{A}^*(S^{-1} \mathcal{A}(\Delta y) S^{-1}) = \frac{1}{\mu} b - \mu \mathcal{A}^*(S^{-1})$$

Set $\mu = \frac{z - b^T y}{\rho}$, $z = \langle C, X \rangle$

A dual scaling algorithm

$$\begin{pmatrix} \langle A_1, S^{-1}A_1S^{-1}\rangle & \cdots & \langle A_1, S^{-1}A_mS^{-1}\rangle \\ \vdots & \ddots & \vdots \\ \langle A_m, S^{-1}A_1S^{-1}\rangle & \cdots & \langle A_m, S^{-1}A_mS^{-1}\rangle \end{pmatrix} \Delta y = \frac{\rho}{z - b^T y} b - \mathcal{A}^*(S^{-1})$$

Remarks:
$z^{k+1} = \langle C, X_k \rangle$ computed as

$$z^{k+1} = \langle b, y^k \rangle + \langle X_k, S^k \rangle = \langle b, y^k \rangle + \frac{z^k - \langle b, y^k \rangle}{\rho} (\Delta y^T \mathcal{A}^*((S^k)^{-1}) + n)$$

$X_k = \mu S^{-1} - \mu S^{-1} \Delta S S^{-1}$ satisfies the primal constraint $\mathcal{A}^*(X_k) = b$.

A first-order algorithm

S. Burer and R. Monteiro $\longrightarrow$ code SDPLR

$$\begin{aligned} & \inf_X \langle C, X\rangle := \text{Trace}(CX) && \text{(P)} \\ \text{s.t.} \quad & \mathcal{A}^*(X) = b \qquad [\langle A_i, X\rangle = b_i,\ i = 1,\ldots,m] \\ & X \succeq 0 \end{aligned}$$

Theorem: (Pataki) Let $\overline{X}$ be an extreme point of SDP-P. Then $\overline{r} = \text{rank}(\overline{X})$ satisfies $\overline{r}(\overline{r}+1)/2 \le m$.

Consequence: We may restrict our search to $X : \text{rank}(X) \le r$, where $r := \min\{\overline{r} : \overline{r}(\overline{r}+1)/2 \ge m\}$ (note: $r \approx \sqrt{2m}$).

Thus SDP-P is equivalent to the *nonlinear program*

$$\min_R \{\langle C, RR^T\rangle : \mathcal{A}^*(RR^T) = b,\ R \in \mathbb{R}^{n\times r}\}$$

A first-order algorithm

$$\min\{\langle C, RR^T\rangle : \mathcal{A}^*(RR^T) = b, R \in \mathbb{R}^{n\times r}\} \qquad (\star)$$

Often $m < n$ and thus $r \ll n \rightarrow$ avoid storing full X

When C and A_i sparse, use first-order NLP code, maintaining the sparsity

SDPLR: solve ($\star$) by an Augmented Lagrangian code.
The subproblems (unconstrainend NLPs) solved by an L-BFGS code.