Semidefinite Programming, Combinatorial Optimization and Real Algebraic Geometry

assoc. prof. Janez Povh, Ph.D.[1]

[1]UNM - Faculty of information studies

Edinburgh, 16. September 2014

Outline

Some notation

- I - unit matrix,
- $X \in \mathbb{R}^{m \times n} \Rightarrow x = \mathrm{vec}(X) \in \mathbb{R}^{mn}$
- $\langle x, y \rangle = x^T y = \sum_i x_i y_i$
- $\langle A, B \rangle = \mathrm{trace}(A^T B) = \sum_{i,j} a_{ij} b_{ij}$

Some notation

- **Symmetric matrices** $\mathcal{S}_n := \{X \in \mathbb{R}^{n\times n}\colon X^T = X\}$
- Cone of **positive semidefinite** matrices $\mathcal{S}_n^+ := \{X \in \mathcal{S}_n\colon u^T X u \geq 0, \forall u \in \mathbb{R}^n\}$ - *closed convex pointed cone*
- **Positive definite** matrices $\mathcal{S}_n^{++} := \{X \in \mathcal{S}_n\colon u^T X u > 0, \forall u \in \mathbb{R}^n\}$
- The **dual** cone:
  $(\mathcal{S}_n^+)^* = \{Y \in \mathcal{S}_n\colon \langle Y, X\rangle \geq 0,\ \forall X \in \mathcal{S}_n^+\} = \mathcal{S}_n^+$
  (Note: $\inf_{X\succeq 0}\langle X,\ Y\rangle = 0 \iff Y \succeq 0$)

Some notation

- **Symmetric matrices** $\mathcal{S}_n := \{X \in \mathbb{R}^{n\times n} \colon X^T = X\}$
- Cone of **positive semidefinite** matrices $\mathcal{S}_n^+ := \{X \in \mathcal{S}_n \colon u^T X u \geq 0, \forall u \in \mathbb{R}^n\}$ - *closed convex pointed cone*
- **Positive definite** matrices $\mathcal{S}_n^{++} := \{X \in \mathcal{S}_n \colon u^T X u > 0, \forall u \in \mathbb{R}^n\}$
- The **dual** cone:
  $(\mathcal{S}_n^+)^* = \{Y \in \mathcal{S}_n \colon \langle Y, X\rangle \geq 0,\ \forall X \in \mathcal{S}_n^+\} = \mathcal{S}_n^+$
  (Note: $\inf_{X\succeq 0}\langle X,\ Y\rangle = 0 \iff Y \succeq 0$)
- For $X \in \mathcal{S}_n^+$ we use $X \succeq 0$.

Few properties of PSD matrices

- For $A \in \mathcal{S}_n$ the following are equivalent
  - $A \in \mathcal{S}_n^+$,
  - all eigenvalues of A are nonnegative real numbers,
  - there exist P and $D = \mathrm{Diag}(d)$, $d \geq 0$, such that $A = PDP^T$,
  - $\det A_{II} \geq 0$ for every main submatrix A_{II} (A_{II} is main submatrix, if $A = [a_{ij}]$, for $i, j \in I \subseteq \{1, \ldots, n\}$).
  - $BAB^T \in \mathcal{S}_n^+$, for some non-singular B.

Few properties of PSD matrices

- For $A \in \mathcal{S}_n$ the following are equivalent
  - $A \in \mathcal{S}_n^+$,
  - all eigenvalues of A are nonnegative real numbers,
  - there exist P and $D = \mathrm{Diag}(d)$, $d \geq 0$, such that $A = PDP^T$,
  - $\det A_{II} \geq 0$ for every main submatrix A_{II} (A_{II} is main submatrix, if $A = [a_{ij}]$, for $i, j \in I \subseteq \{1, \ldots, n\}$).
  - $BAB^T \in \mathcal{S}_n^+$, for some non-singular B.

Few properties of PSD matrices

- For $X, Y \in \mathcal{S}_n^+$: $\langle X,\ Y\rangle \geq 0$ and $\langle X,\ Y\rangle = 0 \iff XY = 0$.

Few properties of PSD matrices

- For $X, Y \in \mathcal{S}_n^+$: $\langle X,\ Y\rangle \geq 0$ and $\langle X,\ Y\rangle = 0 \iff XY = 0$.
- **Schur's complement:** if $A \in \mathcal{S}_n^{++}$ then

$$\begin{bmatrix} A & B \\ B^t & C \end{bmatrix} \succeq 0 \iff C - B^t A^{-1} B \succeq 0.$$

Primal and the dual SDP

- Primal **semidefinite programming problem** (PSDP)

$$\boxed{\begin{array}{rrcl} & \inf & \langle C, X\rangle & \\ \text{s. t.} & \langle A_i, X\rangle & = & b_i, \quad \forall i, \\ & X & \in & \mathcal{S}_n^+ \end{array}}$$

Primal and the dual SDP

- Primal **semidefinite programming problem** (PSDP)

$$\boxed{\begin{array}{rrcl} & \inf & \langle C, X\rangle & \\ \text{s. t.} & \langle A_i, X\rangle & = & b_i, \quad \forall i, \\ & X & \in & \mathcal{S}_n^+ \end{array}}$$

- Dual **semidefinite programming problem** (DSDP)

$$\boxed{\begin{array}{rrcl} & \sup & b^T y & \\ \text{s. t.} & \sum_i y_i A_i + Z & = & C \\ & Z & \in & \mathcal{S}_n^+ \end{array}}$$

Primal and the dual SDP

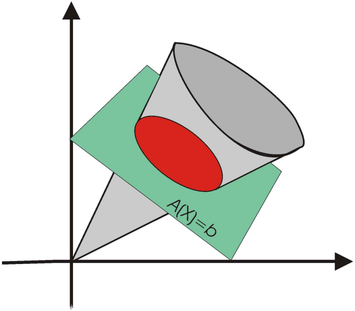
- Primal **semidefinite programming problem** (PSDP)

$$\boxed{\begin{array}{rrcl} & \inf & \langle C, X\rangle & \\ \text{s. t.} & \langle A_i, X\rangle & = & b_i, \quad \forall i, \\ & X & \in & \mathcal{S}_n^+ \end{array}}$$

- Dual **semidefinite programming problem** (DSDP)

$$\boxed{\begin{array}{rrcl} & \sup & b^T y & \\ \text{s. t.} & \sum_i y_i A_i + Z & = & C \\ & Z & \in & \mathcal{S}_n^+ \end{array}}$$

- PSDP and DSDP are dual of each other.



Example 1: linear programming (LP)

Linear programming problem (LP) in standard primal form

$$\boxed{\begin{array}{llll} & \min & c^T x & \\ \text{p. p.} & Ax & = & b \\ & x & \geq & 0 \end{array}}$$

- LP as PSDP: $C = \mathrm{Diag}(c),\ X = \mathrm{Diag}(x),\ A_i = \mathrm{Diag}(A(i,:))$

Example 1: linear programming (LP)

Linear programming problem (LP) in standard primal form

$$\boxed{\begin{array}{rrcl} & \min & c^T x & \\ \text{p. p.} & Ax & = & b \\ & x & \geq & 0 \end{array}}$$

- LP as PSDP: $C = \mathrm{Diag}(c),\ X = \mathrm{Diag}(x),\ A_i = \mathrm{Diag}(A(i,:))$

$$\boxed{\begin{array}{rrcl} & \min & \langle C, X\rangle & \\ \text{p. p.} & \langle A_i, X\rangle & = & b_i,\ \ 1 \leq i \leq m, \\ & X & \in & \mathcal{S}_n^+ \end{array}}$$

Example 2: convex quadratic programming

Def. Convex quadratic programming problem (CCP):

$$\boxed{\begin{array}{llll} \inf & f_0(x) & & \\ \text{p. p.} & f_i(x) & \leq & 0, \quad i = 1, \ldots, m. \quad \text{(CCP)} \end{array}}$$

where: $f_i(x) = x^t U_i x - v_i^t x - z_i$, $U_i \succeq 0$.

- Note
$$f_i(x) \leq 0 \iff A_i = \begin{bmatrix} I & U_i^{1/2} x \\ (U_i^{1/2} x)^t & v_i^t x + z_i \end{bmatrix} \succeq 0.$$

Rem.

$$A_i = \begin{bmatrix} I & 0 \\ 0 & z_i \end{bmatrix} + x_1 \begin{bmatrix} 0 & U_i^{1/2}(:,1) \\ U_i^{1/2}(1,:) & v_{i1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 & U_i^{1/2}(:,n) \\ U_i^{1/2}(n,:) & v_{in} \end{bmatrix}.$$

Example 2: convex quadratic programming cnt.

- Finding min of $f_0(x)$ is equiv. to finding min. of t with additional constraint
$$f_0(x) \leq t.$$
- (CCP) is equivalent to:
$$\begin{array}{lll} \inf & t & \\ \text{p. p.} & \mathrm{Diag}(A_0, A_1, \ldots, A_m) & \succeq & 0, \end{array}$$
- where
$$A_0 = \begin{bmatrix} I & U_0^{1/2}x \\ (U_0^{1/2}x)^t & v_0^t x + z_0 + t \end{bmatrix}.$$

Weak duality

- Let us define

$$\begin{aligned}
\langle A_i, X\rangle &= b_i \qquad =: \mathcal{A}(X) = b \\
\sum_i y_i A_i &=: \mathcal{A}^T(y) \\
OPT_P &= \inf\{\langle C, X\rangle ; \mathcal{A}(X) = b,\ X \succeq 0\} \quad \text{and} \\
OPT_D &= \sup\{b^t y ; \mathcal{A}^t(y) + Z = C,\ Z \succeq 0,\ y \in \mathbb{R}^m\}.
\end{aligned}$$

- More definitions $\sup \emptyset = -\infty$, $\inf \emptyset = \infty$.

Thm.: $OPT_P \geq OPT_D$.

Weak duality

- Let us define

$$
\begin{aligned}
\langle A_i, X\rangle &= b_i \qquad =: \mathcal{A}(X) = b \\
\sum_i y_i A_i &=: \mathcal{A}^T(y) \\
OPT_P &= \inf\,\{\langle C, X\rangle\,;\,\mathcal{A}(X) = b,\ X \succeq 0\} \quad \text{and} \\
OPT_D &= \sup\,\{b^t y\,;\,\mathcal{A}^t(y) + Z = C,\ Z \succeq 0,\ y \in \mathbb{R}^m\}.
\end{aligned}
$$

- More definitions $\sup\emptyset = -\infty$, $\inf\emptyset = \infty$.

Thm.: $OPT_P \geq OPT_D$.

Proof:

- **Duality gap:**
$OPT_P - OPT_D = \langle C, X_{opt}\rangle - b^T y_{opt} = \langle X_{opt}, Z_{opt}\rangle$.

Strong duality

Def.: **PSDP** is **strictly feasible**, if there exists $X \in \mathcal{S}_n^{++}$ such that $\mathcal{A}(X) = b$.
DSDP is **strictly feasible**, if there exists pair $(y, Z) \in \mathbb{R}^m \times \mathcal{S}_n^{++}$ such that $\mathcal{A}^T(y) + Z = C$.

Thm.: Let *(DSDP)* be strictly feasible. Then

- $OPT_P = OPT_D$.
- If $OPT_P < \infty$ then there exists $X \succeq 0$ s.t. $\mathcal{A}(X) = b$ and $\langle C, X \rangle = OPT_P$.

Example 3: optimum is not attained

$$\min \quad \left\langle \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, X \right\rangle$$

$$\text{p. p.} \quad \left\langle \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, X \right\rangle = 2, \quad X \succeq 0.$$

- Feasible solutions: $X = \begin{bmatrix} x_{11} & 1 \\ 1 & x_{22} \end{bmatrix}$ with $x_{22} > 0$ in $x_{11} x_{22} \geq 1$.
- $OPT_P = 0$, but OPT_P is not attained.
- Interpretation : DSDP has no strictly feasible solution.

Example 4: positive duality gap

$$\min \quad \left\langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \right\rangle$$

$$\text{p. p.} \quad \left\langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \right\rangle = 0, \quad \left\langle \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}, X \right\rangle = 2, \quad X \succeq 0.$$

- $OPT_P = 1$, $OPT_D = 0$.

Optimal conditions for SDP

Let PSDP and DSDP be strictly feasible.

Thm.: X^* in (y^*, Z^*) are optimal for PSDP and DSDP if and only if:

$$\begin{array}{lrclrcl} \text{(prim. dop.)} & \mathcal{A}(X) & = & b, & X & \succeq & 0 \\ \text{(dual. dop.)} & \mathcal{A}^T(y) + Z & = & C, & Z & \succeq & 0 \\ \text{(zero duality gap)} & XZ & = & 0 & (b^T y & = & \langle C, X \rangle) \end{array}$$

Central path

- **Assumptions**
  - eqs. $\langle A_i, X\rangle = b_i$ linearly independant
  - PSDP and DSDP strictly feasible
- Then the system

$$\begin{array}{lrcccc} \text{(prim. dop.)} & \mathcal{A}(X) & = & b, & X & \succeq & 0 \\ \text{(dual. dop.)} & \mathcal{A}^T(y) + Z & = & C, & Z & \succeq & 0 \\ & XZ & = & \mu I & \mu & > & 0 \end{array}$$

has a unique solution (X_μ, y_μ, Z_μ).

Def.: The **central path:** $\{(X_\mu, y_\mu, Z_\mu) \colon \mu > 0\}$.

Example

- PSDP

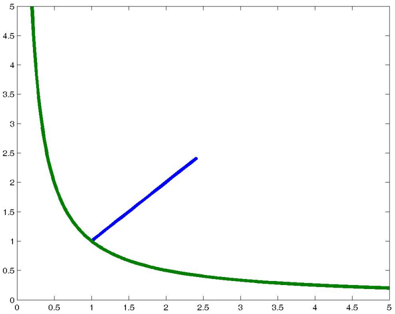
$$\min \quad \left\langle \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, X \right\rangle$$

$$\text{p. p.} \quad \left\langle \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, X \right\rangle = 2, \quad X \succeq 0.$$

- DSDP

$$\max \ 2y \ \ p.\ p.\ Z = \begin{bmatrix} 1 & -y \\ -y & 1 \end{bmatrix} \succeq 0.$$

The central path (primal part)



The properties of the central path

Thm: The central path is a smooth curve parameterized by μ.

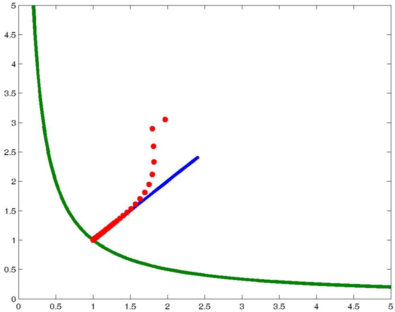
Thm: The central path always converge $\lim_{\mu\downarrow 0}(X_\mu, y_\mu, Z_\mu) = (X^*, y^*, Z^*)$. Limit point is *maximally complementary* primal dual optimal solution.

Def: Let (X^*, y^*, Z^*) be primal dual optimal solution. It is **maximally complementary** if $\operatorname{rank}(X^*)$ is maximal among all primal optimal solutions and $\operatorname{rank}(Y^*)$ is maximal among all dual optimal solutions.

Proof: See e.g. H. Wolkowicz, R. Saigal, L. Vanderberghe (ed.): *Handbook of Semidefinite Programming*. Kluwer Academic Publishers, Boston-Dordrecht-London, 2000.

Path following IPMs

Idea: We solve approximately the equations defining the central path (we follow the central path approximately).



Path following IPMs - more precisely

Input: $\mathcal{A}$, $b \in \mathbb{R}^m$, $C \in \mathcal{S}_n$, $\varepsilon > 0$, $\sigma \in (0,1)$ *and* (X_0, y_0, Z_0) *strictly feasible for PSDP in DSDP*

1. *Set* $k := 0$.
2. **Repeat**
   2.1 $\mu_k = \langle X_k, Z_k \rangle / n$
   2.1 *Solve*

$$\begin{aligned} \mathcal{A}(X_k + \Delta X) &= b, \\ \mathcal{A}^T(y_k + \Delta y) + Z + \Delta Z &= C, \\ (X_k + \Delta X)(Z_k + \Delta Z) &= \sigma \mu_k I. \end{aligned}$$

   2.2 $\Delta X = (\Delta X + (\Delta X)^T)/2$.
   2.3 $(X_{k+1}, y_{k+1}, Z_{k+1}) := (X_k + \Delta X, y_k + \Delta y, Z_k + \Delta Z)$.
   2.4 $k := k + 1$.
3. **until** $\langle X_{k+1}, Z_{k+1} \rangle \leq \varepsilon$.

Output: (X_k, y_k, Z_k).

Solving the system

- Note that the starting point is strictly feasible.

$$
\begin{aligned}
\mathcal{A}(\Delta X) &= 0, \\
\mathcal{A}^T(\Delta y) + \Delta Z &= 0, \\
\Delta X Z_k + X_k \Delta Z &= \mu_k I - X_k Z_k, \\
\Delta X &= (\Delta X + (\Delta X)^T)/2,
\end{aligned}
$$

- **FIRSTLY:** $\Delta Z \;=\; -\mathcal{A}^T(\Delta y)$
- **THEN:** $\Delta X \;=\; (\mu I - X_k Z_k - X_k \Delta Z) Z_k^{-1}$
- **FINALLY:** $\mathcal{A}(X_k \mathcal{A}^T(\Delta y) Z_k^{-1}) \;=\; -\mathcal{A}(\mu Z_k^{-1}) + \mathcal{A}(X_k) \;=\; b - \mathcal{A}(\mu Z_k^{-1}).$

Solving the system: bottlenecks

- On each step we have to
  - compute one inverse (Z_k^{-1})
  - compose the system matrix $M\Delta y = \tilde{b}$. Each $m_{ij} = \langle A_i, XA_jZ_k^{-1}\rangle$ takes $\mathcal{O}(n^3)$ flops (there are $m(m+1)/2$ of them).
  - Solve the system $M\Delta y = \tilde{b}$ (takes $\mathcal{O}(m^3)$ flops)
  - Compute ΔZ in ΔX (takes $\mathcal{O}(mn^2 + n^3)$ flops)

Theoretical guaranty

Thm.: Let (X_0, y_0, Z_0) be strictly feasible starting point with

$$\|Z^{1/2}XZ^{1/2} - \mu I\| \leq \theta\langle X, Z\rangle.$$

The described path following IPM gives ε-optimal solution in at most $\lceil \sqrt{n}/\delta \log(\varepsilon^{-1} \langle X_0,\, S_0\rangle)\rceil$ iterations, where

- $\delta = \sqrt{n}(1 - \sigma)$.
- $\frac{(1+\theta)^{\frac{1}{2}}}{2(1-\theta)^{\frac{3}{2}}}(\theta^2 + n(1 - \sigma)^2) \leq \sigma\theta.$

Boundary point method (Povh, Rendl, Wiegele, 2005)

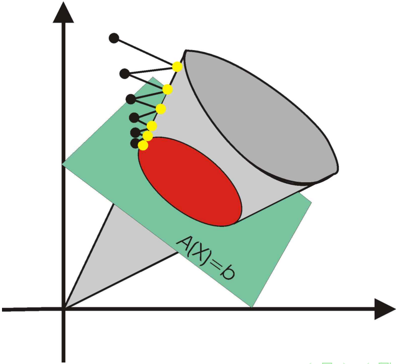
- Relaxed optimality conditions

$$
\begin{array}{lrcll}
\text{(primal feas.)} & \mathcal{A}(X) & \approx & b, & X \succeq 0 \\
\text{(dual feas.)} & \mathcal{A}^T(y) + Z & \approx & C, & Z \succeq 0 \\
\text{(zero opt. dual. gap)} & XZ & = & 0 &
\end{array}
$$

Boundary point method (Povh, Rendl, Wiegele, 2005)

- Relaxed optimality conditions

$$\begin{array}{lrcll} \text{(primal feas.)} & \mathcal{A}(X) & \approx & b, & X \succeq 0 \\ \text{(dual feas.)} & \mathcal{A}^T(y) + Z & \approx & C, & Z \succeq 0 \\ \text{(zero opt. dual. gap)} & XZ & = & 0 & \end{array}$$



-

Idea behind the Boundary point method

- Augmented Lagrangian approach to DSDP

$$\inf\{b^T y \colon \mathcal{A}^T y - Z = C, Z \in \mathcal{S}_n^+\}$$

Idea behind the Boundary point method

- Augmented Lagrangian approach to DSDP

$$\inf\{b^T y \colon \mathcal{A}^T y - Z = C, Z \in \mathcal{S}_n^+\}$$

- Replace DSDP by **DSDP-L**:

$$\inf\ \{b^T y + \langle X,\ C - \mathcal{A}^T(y) + Z\rangle + \frac{\sigma}{2}\|C - \mathcal{A}^T(y) + Z\|^2 \colon Z \succeq 0\}$$

BPM: details

INPUT: $\mathcal{A}, b, C \in \mathcal{S}_n$.

1. *Select* $\sigma > 0$, $X^0 \succeq 0$.
2. $k = 0$.
3. **Repeat**
   - 3.1 *Solve DSDP-L approximately to get* y^k *and* $Z^k \succeq 0$.
   - 3.2 *Update* $X^{k+1} := X^k + \sigma(Z^k - C + \mathcal{A}^T(y^k))$
   - 3.3 $k := k + 1$.
4. **Until:** *stopping criteria is satisfied.*

OUTPUT: X^k, y^k, Z^k.

BPM: solving the inner problem DSDP-L

- **Optimality conditions** for DSDP-L

$$
\begin{array}{rcl}
\mathcal{A}(\mathcal{A}^T(y)) & = & \mathcal{A}(C - Z) + \frac{1}{\sigma}(\mathcal{A}(X) - b) \\
V & = & \sigma(C - \mathcal{A}^T(y) - Z) + X \\
V & \succeq & 0, \quad Z \succeq 0, \quad VZ = 0.
\end{array}
$$

- **Efficient idea**: alternating y and (V, Z).

Solving SDP in praxis - we use software packages

- Nekaj najbolj razširjenih in robustnih paketa
  - SEDUMI (http://sedumi.ie.lehigh.edu/).
  - SDPT3 (http://www.math.nus.edu.sg/ mattohkc/sdpt3.html).
  - SDPA (http://sdpa.sourceforge.net/).
  - MOSEK (`http://www.mosek.com/`).
  - YALMIP (`http://users.isy.liu.se/johanl/yalmip/`).
- To solve large SDP (m is large) we can use:
  - SDPLR (http://dollar.biz.uiowa.edu/ burer/software/SDPLR/)
  - spectral bundle method (http://www-user.tu-chemnitz.de/ helmberg/SBmethod/).
  - boundary point method (http://www.math.uni-klu.ac.at/or/Software/).

Literature

1. Miguel F. Anjos and Jean B. Lasserre. Handbook of Semidefinite, Conic and Polynomial Optimization: Theory, Algorithms, Software and Applications, volume 166 of International Series in Operational Research and Management Science. Springer, 2012
2. E. de Klerk: *Aspects of Semidefinite Programming - Interior Point Algorithms and Selected Applications*. Kluwer Academic Publishers, Dordrecht, 2002.
3. M. Laurent. Sums of squares, moment matrices and optimization over polynomials, volume 149 of The IMA Volumes in Mathematics and its Applications, str. 157–270. Springer, 2009.
4. J. Povh: Semidefinitno programiranje. Obz. mat. fiz., 2002, letn. 49, št. 6, str. 161-173
5. J. Povh: Towards the optimum by semidefinite and copositive programming. VDM Verlag, 2009.
6. H. Wolkowicz, R. Saigal, L. Vanderberghe (ed.): *Handbook of Semidefinite Programming*. Kluwer Academic Publishers, Boston-Dordrecht-London, 2000.