INTRODUCTION TO LMI/SDP OPTIMIZATION

Denis ARZELIER
arzelier@laas.fr

26-27 June 2017
Outline: LMI optimization

1 Introduction: What is an LMI? What is SDP?
   Historical survey - applications - convexity - cones - polytopes

2 SDP duality
   Lagrangian duality - SDP duality - KKT conditions

3 Geometry of LMI sets
   Geometry - algebraic tricks

4 Solving LMIs
   Interior point methods - solvers - interfaces
Lecture material

References on convex optimization:

• A. Ben-Tal, A. Nemirovskii. Lectures on Modern Convex Optimization, SIAM, 2001

Modern state-space LMI methods in control:

• C. Scherer, S. Weiland. Course on LMIs in Control, Lecture Notes Delft & Eindhoven Univ Tech, NL, 2002

LMI and algebraic optimization:

• P. A. Parrilo, S. Lall. Semidefinite Programming Relaxations and Algebraic Optimization in Control, Workshop presented at the 42nd IEEE Conference on Decision and Control, Maui HI, USA, 2003
LMI OPTIMIZATION
PART 1

WHAT IS AN LMI ?
WHAT IS SDP ?

Denis ARZELIER
arzelier@laas.fr

Professeur Jan C Willems

26-27 June 2017
LMI - Linear Matrix Inequality

\[
F(x) = F_0 + \sum_{i=1}^{n} x_i F_i \succeq 0
\]

- \( F_i \in \mathbb{S}^m \) given symmetric matrices
- \( x_i \in \mathbb{R}^n \) decision variables

Fundamental property: feasible set is convex

\[ S = \{ x \in \mathbb{R}^n : F(x) \succeq 0 \} \]

\( S \) is the Spectrahedron

Nota : \( \succeq 0 \) (\( \succ 0 \)) means positive semidefinite (positive definite) e.g. real nonnegative eigenvalues (strictly positive eigenvalues) and defines generalized inequalities on PSD cone

Terminology coined out by Jan Willems in 1971

\[
F(P) = \begin{bmatrix}
A'P + PA + Q & PB + C' \\
B'P + C & R
\end{bmatrix} \succeq 0
\]

"The basic importance of the LMI seems to be largely unappreciated. It would be interesting to see whether or not it can be exploited in computational algorithms"
Historically, the first LMIs appeared around 1890 when Lyapunov showed that the autonomous system with LTI model:

\[ \frac{d}{dt} x(t) = \dot{x}(t) = Ax(t) \]

is stable (all trajectories converge to zero) iff there exists a solution to the matrix inequalities

\[ A'P + PA \prec 0 \quad P = P' \succ 0 \]

which are linear in unknown matrix \( P \)

Aleksandr Mikhailovich Lyapunov
(1857 Yaroslavl - 1918 Odessa)
Example of Lyapunov’s LMI

\[ \begin{bmatrix} -1 & 2 \\ 0 & -2 \end{bmatrix} \quad \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix} \]

\[ A'P + PA < 0 \quad P > 0 \]

\[ \begin{bmatrix} -2p_1 & 2p_1 - 3p_2 \\ 2p_1 - 3p_2 & 4p_2 - 4p_3 \end{bmatrix} < 0 \]

\[ \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix} > 0 \]

Matrices \( P \) satisfying Lyapunov LMI’s

\[ \begin{bmatrix} 2 & -2 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} p_1 + \begin{bmatrix} 0 & 3 & 0 & 0 \\ 3 & -4 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} p_2 + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} p_3 > 0 \]
Some history (1)

1940s - Absolute stability problem: **Lu’re, Postnikov** et al applied Lyapunov’s approach to control problems with **nonlinearity** in the actuator

\[
\dot{x} = Ax + b\sigma(x)
\]

- Stability criteria in the form of LMIs solved analytically by hand
- Reduction to **Polynomial** (frequency dependent) inequalities (small size)
Some history (2)

1960s: Yakubovich, Popov, Kalman, Anderson et al obtained the positive real lemma

The linear system $\dot{x} = Ax + Bu, \ y = Cx + Du$ is passive $H(s) + H(s)^* \geq 0 \ \forall \ s + s^* > 0$ iff

$$P \succ 0 \quad \left[ \begin{array}{cc}
A'P + PA & PB - C' \\
B'P - C & -D - D'
\end{array} \right] \preceq 0$$

- Solution via a simple graphical criterion (Popov, circle and Tsypkin criteria)

Mathieu equation: $\ddot{y} + 2\mu \dot{y} + (\mu^2 + a^2 - q \cos \omega_0 t) y = 0$
$$q < 2\mu a$$
Some history (3)

1971: Willems focused on solving algebraic Riccati equations (AREs)

\[ A'P + PA - (PB + C')R^{-1}(B'P + C) + Q = 0 \]

Numerical algebra

\[ H = \begin{bmatrix} A - BR^{-1}C & BR^{-1}B' \\ -C'R^{-1}C & -A' + C'R^{-1}B' \end{bmatrix} \quad V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \]

\[ P_{are} = V_2V_1^{-1} \]

By 1971, methods for solving LMIs:

- Direct for small systems
- Graphical methods
- Solving Lyapunov or Riccati equations
Some history (4)

1963: Bellman-Fan: infeasibility criteria for multiple Lyapunov inequalities (duality theory)

On Systems of Linear Inequalities in hermitian Matrix Variables

1975: Cullum-Donath-Wolfe: Optimality conditions, nondifferentiable criterion for multiple eigenvalues and algorithm for minimization of sum of maximum eigenvalues

The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices

1979: Khachiyan: polynomial bound on worst case iteration count for LP ellipsoid algorithm of Nemirovski and Shor

A polynomial algorithm in linear programming
Some history (5)

1981: Craven-Mond: **Duality theory**
Linear Programming with Matrix variables

1984: Karmarkar introduces **interior-point** (IP) methods for LP: improved complexity bound and efficiency

1985: Fletcher: **Optimality conditions** for non-differentiable optimization
Semidefinite matrix constraints in optimization

1988: Overton: **Nondifferentiable optimization**
On minimizing the maximum eigenvalue of a symmetric matrix

1988: Nesterov, Nemirovski, Alizadeh, Karmarkar and Thakur **extend** IP methods for convex programming
Interior-Point Polynomial Algorithms in Convex Programming

1990s: most papers on SDP are written (control theory, combinatorial optimization, approximation theory...
Mathematical preliminaries (1)

A set $\mathcal{C}$ is **convex** if the line segment between any two points in $\mathcal{C}$ lies in $\mathcal{C}$

$$\forall \ x_1, x_2 \in \mathcal{C} \quad \lambda x_1 + (1-\lambda)x_2 \in \mathcal{C} \quad \forall \ \lambda \quad 0 \leq \lambda \leq 1$$

The **convex hull** of a set $\mathcal{C}$ is the set of all convex combinations of points in $\mathcal{C}$

$$\text{co} \mathcal{C} = \left\{ \sum_i \lambda_i x_i : x_i \in \mathcal{C} \quad \lambda_i \geq 0 \quad \sum_i \lambda_i = 1 \right\}$$
A hyperplane is a set of the form:
\[ \mathcal{H} = \{ x \in \mathbb{R}^n \mid a'(x - x_0) = 0 \} \quad a \neq 0 \in \mathbb{R}^n \]

A hyperplane divides \( \mathbb{R}^n \) into two halfspaces:
\[ \mathcal{H}_- = \{ x \in \mathbb{R}^n \mid a'(x - x_0) \leq 0 \} \quad a \neq 0 \in \mathbb{R}^n \]
Mathematical preliminaries (3)

A polyhedron is defined by a finite number of linear equalities and inequalities

\[ \mathcal{P} = \{ x \in \mathbb{R}^n : a_j^t x \leq b_j, j = 1, \ldots, m, c_i^t x = d_i, i = 1, \ldots, p \} \]

= \{ x \in \mathbb{R}^n : Ax \preceq b, Cx = d \}

A bounded polyhedron is a polytope

• positive orthant is a polyhedral cone
• k-dimensional simplexes in \( \mathbb{R}^n \)

\[ \mathcal{X} = \text{co} \{ v_0, \ldots, v_k \} = \left\{ \sum_{i=0}^{k} \lambda_i v_i : \lambda_i \geq 0, \sum_{i=0}^{k} \lambda_i = 1 \right\} \]
A set $\mathcal{K}$ is a **cone** if for every $x \in \mathcal{K}$ and $\lambda \geq 0$ we have $\lambda x \in \mathcal{K}$. A set $\mathcal{K}$ is a **convex cone** if it is convex and a cone.

$\mathcal{K} \subseteq \mathbb{R}^n$ is called a **proper cone** if it is a closed solid **pointed** convex cone

\[
a \in \mathcal{K} \quad \text{and} \quad -a \in \mathcal{K} \quad \Rightarrow \quad a = 0
\]
Lorentz cone $\mathbb{L}^n$

3D Lorentz cone or ice-cream cone

$$x^2 + y^2 \leq z^2 \quad z \geq 0$$

arises in quadratic programming
PSD cone $S^n_+$

2D positive semidefinite cone

$$\begin{bmatrix} x & y \\ y & z \end{bmatrix} \succeq 0 \iff x \geq 0 \quad z \geq 0 \quad xz \geq y^2$$

arises in semidefinite programming
Mathematical preliminaries (5)

Every proper cone $\mathcal{K}$ in $\mathbb{R}^n$ induces a partial ordering $\succeq_{\mathcal{K}}$ defining **generalized inequalities** on $\mathbb{R}^n$

$$a \succeq_{\mathcal{K}} b \iff a - b \in \mathcal{K}$$

The positive orthant, the Lorentz cone and the PSD cone are all **proper cones**

- **positive orthant** $\mathbb{R}^n_+$: standard coordinatewise ordering (LP)
  $$x \succeq_{\mathbb{R}^n_+} y \iff x_i \geq y_i$$

- **Lorentz cone** $\mathbb{L}^n$
  $$x_n \geq \sqrt{\frac{n-1}{\sum_{i=1}^{n-1} x_i^2}}$$

- **PSD cone** $\mathbb{S}^n_+$: L"owner partial order
Mathematical preliminaries (6)

The set $\mathcal{K}^* = \{ y \in \mathbb{R}^n | x'y \geq 0 \quad \forall \ x \in \mathcal{K} \}$ is called the dual cone of the cone $\mathcal{K}$.

- $(\mathbb{R}_+^n)^* = \mathbb{R}_+$
- $(\mathbb{S}_+^n)^* = \mathbb{S}_+$

- $\mathbb{L}^n = \{(x, t) \in \mathbb{R}^{n+1} | \|x\| \leq t \}$, then $(\mathbb{L}^n)^* = \{(u, v) \in \mathbb{R}^{n+1} | \|u\|_* \leq v \}$ with $\|u\|_* = \sup \{u'x \mid \|x\| \leq 1\}$

$\mathcal{K}^*$ is closed and convex, $\mathcal{K}_1 \subseteq \mathcal{K}_2 \Rightarrow \mathcal{K}_2^* \subseteq \mathcal{K}_1^*$

$\preceq_{\mathcal{K}}$ is a dual generalized inequality

$$x \preceq_{\mathcal{K}} y \iff \lambda'x \leq \lambda'y \quad \forall \ \lambda \preceq_{\mathcal{K}}^* 0$$
Mathematical preliminaries (7)

\( f : \mathbb{R}^n \to \mathbb{R} \) is convex if \( \text{dom} f \) is a convex set and \( \forall \ x, \ y \in \text{dom} f \) and \( 0 \leq \lambda \leq 1 \)

\[
f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)
\]

If \( f \) is differentiable: \( \text{dom} f \) is a convex set and \( \forall \ x, \ y \in \text{dom} f \)

\[
f(y) \geq f(x) + \nabla f(x)'(y - x)
\]

If \( f \) is twice differentiable: \( \text{dom} f \) is a convex set and \( \forall \ x, \ y \in \text{dom} f \)

\[
\nabla^2 f(x) \succeq 0
\]

Quadratic functions:
\( f(x) = (1/2)x'Px + q'x + r \) is convex if and only if \( P \succeq 0 \)
Convex function $y = x^2$

Nonconvex function $y = -x^2$

Mind the sign!
LMI and SDP formalisms (1)

In mathematical programming terminology
LMI optimization = semidefinite programming (SDP)

<table>
<thead>
<tr>
<th>LMI (SDP dual)</th>
<th>SDP (primal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \min \quad c'x )</td>
<td>( \min \quad -\text{Tr}(F_0Z) )</td>
</tr>
<tr>
<td>under ( F_0 + \sum_{i=1}^{n} x_i F_i &lt; 0 )</td>
<td>under ( -\text{Tr}(F_i Z) = c_i )</td>
</tr>
<tr>
<td>( Z \succeq 0 )</td>
<td>( Z \succeq 0 )</td>
</tr>
</tbody>
</table>

\( x \in \mathbb{R}^n, \ Z \in \mathbb{S}^m, \ F_i \in \mathbb{S}^m, \ c \in \mathbb{R}^n, \ i = 1, \ldots, n \)

Nota:
In a typical control LMI

\[ A'P + PA = F_0 + \sum_{i=1}^{n} x_i F_i < 0 \]

individual matrix entries are decision variables
LMI and SDP formalisms (2)

\[ \exists \, x \in \mathbb{R}^n \mid F_0 + \sum_{i=1}^{n} x_i F_i < 0 \iff \min_{x \in \mathbb{R}^n} \lambda_{max}(F(x)) \]

The LMI feasibility problem is a convex and non differentiable optimization problem.

Example:

\[ F(x) = \begin{bmatrix} -x_1 - 1 & -x_2 \\ -x_2 & -1 + x_1 \end{bmatrix} \]

\[ \lambda_{max}(F(x)) = 1 + \sqrt{x_1^2 + x_2^2} \]
LMI and SDP formalisms (3)

\[
\begin{align*}
\min \quad & c'x \\
\text{s.t.} \quad & b - A'x \in K
\end{align*}
\]  
\[
\begin{align*}
\min \quad & b'y \\
\text{s.t.} \quad & Ay = c \\
& y \in K
\end{align*}
\]

Conic programming in cone $K$

- positive orthant (LP)
- Lorentz (second-order) cone (SOCP)
- positive semidefinite cone (SDP)

Hierarchy: LP cone $\subset$ SOCP cone $\subset$ SDP cone
LMI and SDP formalisms (4)

LMI optimization = generalization of linear programming (LP) to cone of positive semidefinite matrices = semidefinite programming (SDP)

Linear programming pioneered by
- Kantorovich (co-winner of the 1975 Nobel prize in economics)

Unfortunately, SDP has not reached maturity of LP or SOCP so far..
Applications of SDP

- control systems
- robust optimization
- signal processing
- sparse Principal Component Analysis
- structural design (trusses)
- geometry (ellipsoids)
- Euclidean distance matrices (sensor network localization, molecular conformation)
- graph theory and combinatorics (MAXCUT, Shannon capacity)
- facility layout problem (single-row facility layout problem, VLSI floorplanning)

and many others...

See Helmberg’s page on SDP
www-user.tu-chemnitz.de/~helmberg/semidef.html
Robust optimization (1)

In many real-life applications of optimization problems, exact values of input data (constraints) are seldom known
- Uncertainty about the future
- Approximations of complexity by uncertainty
- Errors in the data
- Variables may be implemented with errors

\[
\begin{align*}
\min_{x} \quad & f_0(x,u) \\
\text{under} \quad & f_i(x,u) \leq 0 \quad i = 1, \ldots, m
\end{align*}
\]

where \(x \in \mathbb{R}^n\) is the vector of decision variables and \(u \in \mathbb{R}^p\) is the parameters vector.

- Stochastic programming
- Sensitivity analysis
- Interval arithmetic
- Worst-case analysis

\[
\begin{align*}
\min_{x} \quad & \sup_{u \in U} f_0(x,u) \\
\text{under} \quad & \sup_{u \in U} f_i(x,u) \leq 0 \quad i = 1, \ldots, m
\end{align*}
\]
Robust optimization (2)

Case study by Ben Tal and Nemirovski: [Math. Programm. 2000]
90 LP problems from NETLIB + uncertainty
quite small (just 0.1%) perturbations of "obviously uncertain" data coefficients can make the "nominal" optimal solution $x^*$ heavily infeasible
Remedy: robust optimization, with robustly feasible solutions guaranteed to remain feasible at the expense of possible conservatism
Robust conic problem: [Ben Tal Nemirovski 96]

$$\min_{x \in \mathbb{R}^n} c'x$$
$$\text{s.t. } Ax - b \in \mathcal{K}, \quad \forall (A, b) \in \mathcal{U}$$

This last problem, the so-called robust counterpart is still convex, but depending on the structure of $\mathcal{U}$, can be much harder that original conic problem
### Robust optimization (3)

<table>
<thead>
<tr>
<th>Uncertainty</th>
<th>Problem</th>
<th>Optimization Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>polytopic</td>
<td>LP</td>
<td>LP</td>
</tr>
<tr>
<td>ellipsoid</td>
<td></td>
<td>SOCP</td>
</tr>
<tr>
<td>LMI</td>
<td></td>
<td>SDP</td>
</tr>
<tr>
<td>polytopic</td>
<td>SOCP</td>
<td>SOCP</td>
</tr>
<tr>
<td>ellipsoid</td>
<td></td>
<td>SDP</td>
</tr>
<tr>
<td>LMI</td>
<td></td>
<td>NP-hard</td>
</tr>
</tbody>
</table>

**Examples of applications:**

**Robust LP**: Robust portfolio design in finance [Lobo 98], discrete-time optimal control [Boyd 97], robust synthesis of antennae arrays [Lebret 94], **FIR filter** design [Wu 96]

**Robust SOCP**: robust least-squares in identification [El Ghaoui 97], robust synthesis of antennae arrays and **FIR filter** synthesis
Robust optimization (4)
Robust LP as a SOCP

Robust counterpart of robust LP

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad c'x \\
\text{s.t.} & \quad a'_i x \leq b_i, \quad i = 1, \ldots, m, \\
& \quad \forall a_i \in \mathcal{E}_i \\
& \quad \mathcal{E}_i = \{ \bar{a}_i + P_i u \mid ||u||_2 \leq 1 \text{ and } P_i \succeq 0 \} 
\end{align*}
\]

Note that

\[
\max_{a_i \in \mathcal{E}_i} a'_i x = \bar{a}_i' x + ||P_i x||_2 \leq b_i
\]

SOCP formulation

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad c'x \\
\text{s.t.} & \quad \bar{a}_i' x + ||P_i x|| \leq b_i, \quad i = 1, \ldots, m,
\end{align*}
\]
Robust optimization (5)

Example of Robust LP

\[ J_1^* = \max_{x,y} \quad 2x + y \]
\[ \text{s.t.} \quad x \geq 0, \ y \geq 0 \]
\[ x \leq 2 \]
\[ y \leq 2 \]
\[ x + y \leq 3 \]

\[ J_2^* = \max_{x,y} \quad 2x + y \]
\[ \text{s.t.} \quad x \geq 0, \ y \geq 0 \]
\[ \sqrt{x^2 + y^2} \leq 2 - x \]
\[ \sqrt{x^2 + y^2} \leq 2 - y \]
\[ \sqrt{x^2 + y^2} \leq 3 - x - y \]

\[ (x^*, y^*) = (2,1) \]
\[ J_1^* = 5 \]
\[ (x^*, y^*) = (0.8284, 0.8284) \]
\[ J_2^* = 2.4852 \]

\[ \mathcal{E}_1 = \mathcal{E}_2 = \left\{ \begin{bmatrix} 1 & 0 \end{bmatrix}^T + 1_2 u \mid ||u||_2 \leq 1 \right\} \]
\[ \mathcal{E}_3 = \left\{ \begin{bmatrix} 1 & 1 \end{bmatrix}^T + 1_2 u \mid ||u||_2 \leq 1 \right\} \]
Combinatorial optimization (1)

**Combinatorics:** Graph theory, polyhedral combinatorics, combinatorial optimization, enumerative combinatorics...

**Definition:** Optimization problems in which the solution space is discrete (finite collection of objects) or a decision-making problem in which each decision has a finite (possibly many) number of feasibilities.

Depending upon the formalism
- **0-1 Linear Programming problems:** 0-1 Knapsack problem,…
- **Propositional logic:** Maximum satisfiability problems…
- **Constraints satisfaction problems:** Airline crew assignment, maximum weighted stable set problem…
- **Graph problems:** Max-Cut, Shannon or Lovasz capacity of a graph, bandwidth problems, equipartition problems…
Combinatorial optimization (2)

SDP relaxation of QP in binary variables

\[
(BQP) \quad \max_{x \in \{−1, 1\}} x'Qx
\]

Noticing that \(x'Qx = \text{trace}(Qxx')\) we get the equivalent form

\[
(BQP) \quad \max_X \text{trace}(QX)
\]

\[
\text{diag}(X_{ii}) = e = [1 \cdots 1]
\]

s.t. \(X \succeq 0\)

\(\text{rank}(X) = 1\)

Dropping the non convex rank constraint leads to the SDP relaxation:

\[
(SDP) \quad \max_X \text{trace}(QX)
\]

\[
\text{diag}(X_{ii}) = e = [1 \cdots 1]
\]

s.t. \(X \succeq 0\)

Interpretation: lift from \(\mathbb{R}^n\) to \(\mathbb{S}^n\)
Combinatorial optimization (3)

Example

\[(BQP) \min_{x \in \{-1,1\}} x'Qx = x_1x_2 - 2x_1x_3 + 3x_2x_3\]

with \(Q = \begin{bmatrix} 0 & 0.5 & -1 \\ 0.5 & 0 & 1.5 \\ -1 & 1.5 & 0 \end{bmatrix}\)

SDP relaxation

\[(SDP) \min_X \text{trace}(QX) = X_1 - 2X_2 + 3X_3\]

s.t. \(X = \begin{bmatrix} 1 & X_1 & X_2 \\ X_1 & 1 & X_3 \\ X_2 & X_3 & 1 \end{bmatrix} \succeq 0\)

\(X^* = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}\) \(\text{rank}(X^*) = 1\)

From \(X^* = x^*x^*'\), we recover the optimal solution of (BQP)

\(x^* = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}'\)
Combinatorial optimization (4)

Example (continued)

Visualization of the feasible set of (SDP) in $(X_1, X_2, X_3)$ space:

$$
X = \begin{bmatrix}
1 & X_1 & X_2 \\
X_1 & 1 & X_3 \\
X_2 & X_3 & 1
\end{bmatrix} \succeq 0
$$

Optimal vertex is $[-1, 1, -1]$
LMI OPTIMIZATION
PART 2

Lagrangian and SDP duality

arzelier@laas.fr

26-27 June 2017
Duality

- Versatile notion
- Theoretical results and numerical methods
- **Certificates** of infeasibility

Lagrangian duality has many applications and interpretations (price or tax, game, geometry...)

Applications of SDP duality:
- numerical solvers design
- problems reduction
- new theoretical insights into control problems

In the sequel we will recall some basic facts about Lagrangian duality and SDP duality
Lagrangian duality

Let the **primal** problem

\[
\begin{align*}
    p^* &= \min_{x \in \mathbb{R}^n} f_0(x) \\
    \text{s.t.} & \quad f_i(x) \leq 0 \quad i = 1, \ldots, m \\
    & \quad h_i(x) = 0 \quad i = 1, \ldots, p
\end{align*}
\]

Define **Lagrangian** \( L(.,.,.) \) \( \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \)

\[
    L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)
\]

where \( \lambda, \mu \) are **Lagrange multipliers** vectors or **dual variables**

Let the **Lagrange dual** function

\[
    g(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu)
\]

- \( g \) is always **concave**
- \( g(\lambda, \mu) = -\infty \) if there is no finite infimum
Lagrangian duality (2)

A pair \((\lambda, \mu)\) s.t. \(\lambda \succeq 0\) and \(g(\lambda, \mu) > -\infty\) is dual feasible

For any primal feasible \(x\) and dual feasible pair \((\lambda, \mu)\)

\[ g(\lambda, \mu) \leq p^* \leq f_0(x) \]

\[ \min_x x^4 - 3x^2 - x \]

under \(x(x + 1) \leq 0\)
Lagrangian duality (3)

Lagrange dual problem

\[ d^* = \max_{\lambda, \mu} g(\lambda, \mu) \]
\[ \text{s.t. } \lambda \succeq 0 \]

The Lagrange dual problem is a convex optimization problem

<table>
<thead>
<tr>
<th>Primal</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \inf_{x \in \mathbb{R}^n} \sup_{\lambda, \mu} L(x, \lambda, \mu) )</td>
<td>( \sup_{\lambda, \mu} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) )</td>
</tr>
<tr>
<td>s.t. ( \lambda \succeq 0 )</td>
<td>s.t. ( \lambda \succeq 0 )</td>
</tr>
</tbody>
</table>

A Lagrangian relaxation consists in solving the dual problem instead of the primal problem
Weak and strong duality

Weak duality (max-min inequality):

\[ p^* \geq d^* \]

because

\[ g(\lambda, \mu) \leq f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \leq f_0(x) \]

for any primal feasible \( x \) and dual feasible \( \lambda, \mu \)

The difference \( p^* - d^* \geq 0 \) is called duality gap

Strong duality (saddle-point property):

\[ p^* = d^* \]

Sometimes, constraint qualifications ensure that strong duality holds

Example: Slater's condition = strictly feasible convex primal problem

\[ f_i(x) < 0, \ i = 1, \ldots, m \quad h_i(x) = 0, \ i = 1, \ldots, p \]
Geometric interpretation of duality (1)

Consider the primal optimization problem

$$p^* = \min_{x \in \mathbb{R}} f_0(x) \quad \text{s.t.} \quad f_1(x) \leq 0$$

with Lagrangian and dual function

$$L(x, \lambda) = f_0(x) + \lambda f_1(x) \quad g(\lambda) = \inf_x L(x, \lambda)$$

The dual problem is given by:

$$d^* = \max_{\lambda} g(\lambda) \quad \text{s.t.} \quad \lambda \succeq 0$$
Geometric interpretation of duality (2)

Set of values $\mathcal{G} = (f_1(x), f_0(x)), \forall x \in \mathcal{D}$

$$L(x, \lambda) = f_0(x) + \lambda f_1(x) = \begin{bmatrix} \lambda & 1 \end{bmatrix} \begin{bmatrix} f_1(x) \\ f_0(x) \end{bmatrix}$$

$$g(\lambda) = \inf_{x \in \mathcal{D}} L(\lambda, x) = \inf_{x \in \mathcal{D}} \left\{ \begin{bmatrix} \lambda & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \right\} \quad (u, v) \in \mathcal{G}$$

Supporting hyperplane with slope $-\lambda$

$$\begin{bmatrix} \lambda & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \geq g(\lambda) \quad (u, v) \in \mathcal{G}$$
Geometric interpretation of duality (3)

Three supporting hyperplanes, including the optimum $\lambda^*$ yielding $d^* < p^*$
No strong duality here

$$p^* - d^* > 0$$

Duality gap $\neq 0$
Geometric interpretation of duality (4)

\[ \mathcal{B} = \{(0, s) \in \mathbb{R} \times \mathbb{R} : s < p^*\} \]

- Separating hyperplane theorem for \( \mathcal{G} \) and \( \mathcal{B} \)
- The separating hyperplane is a supporting hyperplane to \( \mathcal{G} \) in \((0, p^*)\)
- Slater's condition ensures the hyperplane is non vertical
Suppose that strong duality holds, let \( x^* \) be primal optimal and \((\lambda^*, \mu^*)\) be dual optimal,

\[
\begin{align*}
f_0(x^*) &= g(\lambda^*, \mu^*) \\
&= \inf_x \left( f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \mu_i^* h_i(x) \right) \\
&\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \mu_i^* h_i(x^*) \\
&< f_0(x^*)
\end{align*}
\]

\[
\lambda_i^* f_i(x^*) = 0 \quad i = 1, \ldots, m
\]

This is \textbf{complementary slackness} condition

\[
\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0 \quad \text{or} \quad f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0
\]

In words, the \( i \)th optimal Lagrange multiplier is \textbf{zero} unless the \( i \)th constraint is \textbf{active} at the optimum
KKT optimality conditions

\[ f_i, h_i \] are differentiable and strong duality holds

\[
\begin{align*}
  h_i(x^*) &= 0, \quad i = 1, \ldots, p, \quad \text{(primal feasible)} \\
  f_i(x^*) &\leq 0, \quad i = 1, \ldots, m, \quad \text{(primal feasible)} \\
  \lambda_i^* &\geq 0, \quad i = 1, \ldots, m, \quad \text{(dual feasible)} \\
  \lambda_i^* f_i(x^*) &= 0, \quad i = 1, \ldots, m, \quad \text{(complementary)} \\
  \nabla f_0(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \mu_i^* \nabla h_i(x^*) &= 0
\end{align*}
\]

Necessary Karush-Kuhn-Tucker conditions satisfied by any primal and dual optimal pair \( x^* \) and \((\lambda^*, \mu^*)\)

For convex problems, KKT conditions are also sufficient
Feasibility of inequalities (1)

∃ \( x \in \mathbb{R}^n \) : \[
\begin{cases}
  f_i(x) \leq 0 & i = 1, \ldots, m \\
  h_i(x) = 0 & i = 1, \ldots, p 
\end{cases}
\]

Dual function: \( g(.,.) : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \)

\[
g(\lambda, \mu) = \inf_{x \in \mathcal{D}} \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)
\]

The dual feasibility problem is

∃ \( (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p \) : \[
\begin{cases}
  g(\lambda, \mu) > 0 \\
  \lambda \succeq 0
\end{cases}
\]

Theorem of weak alternatives
At most, one of the two (primal and dual) is feasible
If the dual problem is feasible then the primal problem is infeasible
Proof of the theorem of alternatives

Suppose \( \bar{x} \in \mathcal{D} \) is a feasible point for the primal problem

\[
g(\lambda, \mu) = \inf_{x \in \mathcal{D}} \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \\
\leq \sum_{i=1}^{m} \lambda_i f_i(\bar{x}) + \sum_{i=1}^{p} \mu_i h_i(\bar{x}) \leq 0 \\
\forall (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p
\]

and so \( g(\lambda, \mu) \leq 0 \) for all \( \lambda \succeq 0 \)

If \( f_i \) are convex functions, \( h_i \) are affine functions and some type of constraint qualification holds:

**Theorem of strong alternatives**

Exactly one of the two alternative holds

A dual feasible pair \((\lambda, \mu)\) gives a certificate (proof) of infeasibility of the primal
Feasibility of inequalities (3)
Geometric interpretation

\[ P = \left\{ (u, v) \in \mathbb{R}^2 : \begin{bmatrix} u \\ v \end{bmatrix} \preceq 0 \right\} \]

\[ H_\lambda = \left\{ (u, v) \in \mathbb{R}^2 : \lambda' \begin{bmatrix} u \\ v \end{bmatrix} = g(\lambda) \right\} \]

If \( g(\lambda) > 0 \) and \( \lambda \succeq 0 \) then \( H_\lambda \) is a separating hyperplane for \( P \) from

\[ G = \left\{ \begin{bmatrix} f_1(x) & f_2(x) \end{bmatrix} : x \in \mathbb{R}^n \right\} \]
Conic duality (1)

Let the primal:

\[
p^* = \min_{x \in \mathbb{R}^n} f_0(x) \\
\text{s.t. } f_i(x) \preceq K_i 0 \quad i = 1, \ldots, m
\]

Lagrange dual function: \( g(.): \mathbb{R}^m \to \mathbb{R} \)

\[
g(\lambda) = \inf_{x \in D} f_0(x) + \sum_{i=1}^{m} \lambda_i' f_i(x)
\]

Lagrange dual problem:

\[
d^* = \max_{\lambda \in \mathbb{R}^m} g(\lambda) \\
\text{s.t. } \lambda_i \succeq K_i^* 0, \quad i = 1, \ldots, m
\]
Conic duality (2)

- **Weak duality**
- **Strong duality:**
  - if primal is s.f. with finite $p^*$ then $d^*$ is reached by dual
  - if dual is s.f. with finite $d^*$ then $p^*$ is reached by primal
  - if primal and dual are s.f. then $p^* = d^*$
- **Complementary slackness:**
  \[
  \lambda_i^* f_i(x^*) = 0 \\
  \lambda_i^* \succ_{K_i} 0 \Rightarrow f_i(x^*) = 0 \\
  f_i(x^*) \prec_{K_i} 0 \Rightarrow \lambda_i^* = 0
  \]
- **KKT conditions:**
  \[
  f_i(x^*) \leq_{K_i} 0 \\
  \lambda_i^* \succeq_{K_i} 0 \\
  \nabla f_0(x^*) + \sum_{i=1}^{m} \nabla f_i(x^*)' \lambda_i^* = 0
  \]
Example of conic duality

Consider the primal conic program

\[
\begin{align*}
\min \quad & x_1 \\
\text{s.t.} \quad & \begin{bmatrix} x_1 - x_2 \\ 1 \\ x_1 + x_2 \end{bmatrix} \succeq_{\mathbb{L}^3} 0 \iff x_1 + x_2 > 0 \\
& 4x_1x_2 \geq 1
\end{align*}
\]

with dual

\[
\begin{align*}
\max \quad & -\lambda_2 \\
\text{s.t.} \quad & \begin{cases}
\lambda_1 + \lambda_3 = 1 \\
-\lambda_1 + \lambda_3 = 0 \\
\lambda \in \mathbb{L}^3
\end{cases} \iff \lambda_1 = \lambda_3 = 1/2 \\
& \frac{1}{2} \geq \sqrt{1/4 + \lambda_2^2}
\end{align*}
\]

The primal is strictly feasible and bounded below with \( p^* = 0 \) which is not reached since dual problem is infeasible \( d^* = -\infty \)
SDP duality (1)

Primal SDP:

\[ p^* = \min_{x \in \mathbb{R}^n} c'x \]
\[ \text{s.t. } F_0 + \sum_{i=1}^{n} x_i F_i \preceq 0 \]

Lagrange dual function:

\[ g(Z) = \inf_{x \in \mathcal{D}} \left( c'x + \text{tr} \ Z F(x) \right) \]
\[ = \begin{cases} 
\text{tr} \ F_0 Z & \text{if } \text{tr} \ F_i Z + c_i = 0 \quad i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases} \]

Dual SDP:

\[ d^* = \max_{Z \in \mathcal{S}^m} \text{tr} \ F_0 Z \]
\[ \text{s.t. } \text{tr} \ F_i Z + c_i = 0 \quad i = 1, \ldots, n \]
\[ Z \succeq 0 \]

Complementary slackness:

\[ \text{tr} \ F(x^*) Z^* = 0 \iff F(x^*) Z^* = Z^* F(x^*) = 0 \]
SDP duality (2)
KKT optimality conditions

\[ F_0 + \sum_{i=1}^{n} x_i F_i + Y = 0 \quad Y \succeq 0 \]

\[ \forall i \text{ trace } F_i Z + c_i = 0 \quad Z \succeq 0 \]

\[ Z^* F(x^*) = -Z^* Y^* = 0 \]

Nota:
Since \( Y^* \succeq 0 \) and \( Z^* \succeq 0 \) then

\[ \text{trace } F(x^*) Z^* = 0 \iff F(x^*) Z^* = Z^* F(x^*) = 0 \]

Theorem:
Under the assumption of strict feasibility for the primal and the dual, the above conditions form a system of necessary and sufficient optimality conditions for the primal and the dual
Consider the primal semidefinite program

\[
\begin{align*}
\text{min} & \quad x_1 \\
\text{s.t.} & \quad \begin{bmatrix} 0 & x_1 & 0 \\ x_1 & -x_2 & 0 \\ 0 & 0 & -1 - x_1 \end{bmatrix} \preceq 0
\end{align*}
\]

with dual

\[
\begin{align*}
\text{max} & \quad -z_6 \\
\text{s.t.} & \quad \begin{bmatrix} z_1 & (1 - z_6)/2 & z_4 \\ (1 - z_6)/2 & 0 & z_5 \\ z_4 & z_5 & z_6 \end{bmatrix} \succeq 0
\end{align*}
\]

In the primal \(x_1 = 0\) (\(x_1\) appears in a row with zero diagonal entry) so the primal optimum is \(x_1^* = 0\)

Similarly, in the dual necessarily \((1 - z_6)/2 = 0\) so the dual optimum is \(z_6^* = 1\)

There is a nonzero duality gap here \((p^* = 0) > (d^* = -1)\)
Conic theorem of alternatives

\[ f_i(x) \preceq_{\mathcal{K}_i} 0 \quad \mathcal{K}_i \subseteq \mathbb{R}^{k_i} \]

Lagrange dual function

\[ g(\lambda) = \inf_{x \in \mathcal{D}} \sum_{i=1}^{m} \lambda_i^t f_i(x) \quad \lambda_i \in \mathbb{R}^{k_i} \]

Weak alternatives:

1. \[ 1 - f_i(x) \preceq_{\mathcal{K}_i} 0 \quad i = 1, \ldots, m \]
2. \[ 2 - \lambda_i \succeq_{\mathcal{K}_i^*} 0 \quad g(\lambda) > 0 \]

Strong alternatives:

\( f_i \) \( \mathcal{K}_i \)-convex and \( \exists \ x \in \text{relintD} \)

1. \[ 1 - f_i(x) \prec_{\mathcal{K}_i} 0 \quad i = 1, \ldots, m \]
2. \[ 2 - \lambda_i \succeq_{\mathcal{K}_i^*} 0 \quad g(\lambda) \geq 0 \]
Theorem of alternatives for LMIs

For the LMI feasible set

\[ F(x) = F_0 + \sum_i x_i F_i < 0 \]

Exactly one statement is true
1- \( \exists \ x \ \text{s.t.} \ F(x) < 0 \)
2- \( \exists \ 0 \neq Z \succeq 0 \ \text{s.t.} \)
\[ \text{trace } F_0 Z \geq 0 \ \text{and trace } F_i Z = 0 \ \text{for } i = 1, \ldots, n \]

Useful for giving certificate of infeasibility of LMIs

Rich literature on theorems of alternatives for generalized inequalities, e.g. nonpolyhedral convex cones

Elegant proofs of standard results (Lyapunov, ARE) in linear systems control
S-procedure (1)

**S-procedure**: also frequently useful in robust and nonlinear control, also an outcome of the theorem of alternatives

1- if \( x' A_1 x \geq 0, \ldots, x' A_m x \geq 0 \)
then \( x' A_0 x \geq 0 \ \forall \ x \in \mathbb{R}^n \)

2- \( \exists \tau_j \geq 0 \) s.t. \( x' A_0 x - \sum_{j=1}^{m} \tau_j x' A_j x \geq 0 \)

The **S-procedure** consists in replacing 1 by 2

The converse also holds (no duality gap)

* when \( m = 1 \) for real quadratic forms and \( \exists \ x \ | \ x' A_1 x > 0 \) (from the theorem of alternatives)
* when \( m = 2 \) for complex quadratic forms
**S-procedure (2)**

Sketch of the proof for \( m = 1 \)

**Dines theorem:**

For \((A_0, A_1) \in \mathbb{S}_n\) then

\[
\mathcal{K} = \{(u, v) = (x'A_0x, x'A_1x) : x \in \mathbb{R}^n\}
\]

is a **closed convex cone** of \(\mathbb{R}^2\)

Suppose if \( v = x'A_1x \geq 0 \) then \( u = x'A_0x \geq 0 \)

Defining \( Q = \{v \geq 0, \ u < 0\} \) then \( \mathcal{K} \cap Q = \emptyset \)

**Separating Hyperplane Theorem:**

\[
\tau_1u - \tau_2v < 0 \quad (u, v) \in Q \quad \tau_2 \geq 0 \quad \tau_1 > 0
\]

\( \forall (u, v) \in \mathcal{K} \quad \exists \ \tau = \frac{\tau_2}{\tau_1} \geq 0 \quad u - \tau v \geq 0 \)
S-procedure (3)
Counter-example $m = 3$ and $n = 2$

Let the quadratic forms
\[ f_1(x, y) = -x^2 + 2y^2 \quad f_2(x, y) = 2x^2 - y^2 \]
\[ f_0(x, y) = xy \]
then
\[ Q = \{(x, y) \mid f_1(x, y) \geq 0 \text{ and } f_2(x, y) \geq 0\} \]
\[ = \left\{(x, y) \mid 1/\sqrt{2} \leq \left|\frac{x}{y}\right| \leq \sqrt{2}\right\} \]
and
\[ (x, y) = (1, 1) \mid f_1(x, y) > 0 \text{ and } f_2(x, y) > 0 \]
\[ f_0(x, y) \geq 0 \quad \forall (x, y) \in Q \]
But \(\not\exists (\tau_1, \tau_2) \succeq 0\) s.t.
\[ xy - \tau_1(-x^2 + 2y^2) - \tau_2(2x^2 - y^2) \geq 0 \]
Finsler’s (Debreu) lemma (1)

The following statements are equivalent

1. \( x'A_0x > 0 \) \( \forall x \neq 0 \in \mathbb{R}^n \), \( n \geq 3 \), s.t. \( x'A_1x = 0 \)

2. \( A_0 + \tau A_1 \succ 0 \) for some \( \tau \in \mathbb{R} \)

Theorem of alternatives

1. \( \exists \tau \in \mathbb{R} \mid \tau A_1 + A_0 \succ 0 \)

2. \( \exists Z \in \mathbb{S}_+^n : \text{tr}(ZA_1) = 0 \) and \( \text{tr}(A_0Z) \leq 0 \)

Paul Finsler
(1894 Heilbronn - 1970 Zurich)
Counter-example 1:

\[ f_0(x) = x_1^2 - 2x_2^2 - x_3^2 \quad f_1(x) = x_1 - x_2 \]

\[ f_0(x) \leq 0 \text{ if } f_1(x) = 0 \]

But, no \( \tau \) exists s.t. \( f_0(x) + \tau f_1(x) \leq 0 \)

\[ x' \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix} x + \tau \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} x \leq 0 \]

Pick out \( x = \begin{bmatrix} 4 & 0 & 0 \end{bmatrix} \) and \( x = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \)

Counter-example 2:

\[ f_0(x) = 2x_1x_2 \quad f_1(x) = x_1^2 - x_2^2 \]

\( f_0(x) > 0 \) for \( x \mid f_1(x) = 0 \) but no \( \tau \in \mathbb{R} \) exists s.t.

\[ f_0(x) + \tau f_1(x) = x' \begin{bmatrix} \tau & 1 \\ 1 & -\tau \end{bmatrix} x > 0 \]
Elimination lemma

The following statements are equivalent

1 – $H^\bot A H^{\bot*} \succ 0$ or $H H^* \succ 0$

2 – $\exists X \ | \ A + X H + H^* X^* \succ 0$

Theorem of alternatives

1 – $\exists X \in \mathbb{C}^{m \times n} \ | \ H X + (X H)^* + A \succ 0$

2 – $\exists Z \in \mathbb{S}_+^n : Z H = 0$ and $\text{tr}(A Z) \leq 0$

Nota: For $H \in \mathbb{C}^{n \times m}$ with rank $r$, $H^\bot \in \mathbb{C}^{(n-r) \times n}$ s.t.

$$H^\bot H = 0 \quad H^\bot H^{\bot*} \succ 0$$
Geometry of LMI sets

Given $F_i \in \mathbb{S}^m$ we want to characterize the shape in $\mathbb{R}^n$ of the LMI set

$$S = \{x \in \mathbb{R}^n : F(x) = F_0 + \sum_{i=1}^{n} x_i F_i \succeq 0\}$$

Matrix $F(x)$ is PSD iff its principal minors $f_i(x)$ are nonnegative

Principal minors are multivariate polynomials of indeterminates $x_i$

So the LMI set can be described as

$$S = \{x \in \mathbb{R}^n : f_i(x) \geq 0, \ i = 1, \ldots, n\}$$

which is a semialgebraic set

Moreover, it is a convex set
Example of 2D LMI feasible set

\[ F(x) = \begin{bmatrix} 1 - x_1 & x_1 + x_2 & x_1 \\ x_1 + x_2 & 2 - x_2 & 0 \\ x_1 & 0 & 1 + x_2 \end{bmatrix} \succeq 0 \]

Feasible iff all principal minors nonnegative

System of polynomial inequalities \( f_i(x) \geq 0 \)

1st order minors

\[ f_1(x) = 1 - x_1 \geq 0 \]
\[ f_2(x) = 2 - x_2 \geq 0 \]
\[ f_3(x) = 1 + x_2 \geq 0 \]
2nd order minors

\[ f_4(x) = (1 - x_1)(2 - x_2) - (x_1 + x_2)^2 \geq 0 \]
\[ f_5(x) = (1 - x_1)(1 + x_2) - x_1^2 \geq 0 \]
\[ f_6(x) = (2 - x_2)(1 + x_2) \geq 0 \]
3rd order minor

\[ f_7(x) = (1 + x_2)((1 - x_1)(2 - x_2) - (x_1 + x_2)^2) \]
\[ -x_1^2(2 - x_2) \geq 0 \]
LMI feasible set = intersection of semialgebraic sets $f_i(x) \geq 0$ for $i = 1, \ldots, 7$
Example of 3D LMI feasible set

LMI set

$$S = \{x \in \mathbb{R}^3 : \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{bmatrix} \succeq 0\}$$

arising in SDP relaxation of MAXCUT

Semialgebraic set

$$S = \{x \in \mathbb{R}^3 : 1 + 2x_1x_2x_3 - (x_1^2 + x_2^2 + x_3^2) \geq 0, \frac{x_1^2}{1} \leq 1, \frac{x_2^2}{1} \leq 1, \frac{x_3^2}{1} \leq 1\}$$
Intersection of LMI sets

Intersection of LMI feasible sets

$F(x) \succeq 0 \quad x_1 \geq -2 \quad 2x_1 + x_2 \leq 0$

is also an LMI

$$
\begin{bmatrix}
F(x) & 0 & 0 \\
0 & x_1 + 2 & 0 \\
0 & 0 & -2x_1 - x_2
\end{bmatrix} \succeq 0
$$
Reformulations
Linear LMI constraint = projection in subspace
Using explicit subspace basis, more efficient formulations (less decision variables) can be obtained
Example: original problem

\[
\begin{align*}
\max & \quad 2x_1 + 2x_2 \\
\text{s.t.} & \quad \begin{bmatrix} 1 + x_1 & x_2 \\ x_2 & 1 - x_1 \end{bmatrix} \succeq 0
\end{align*}
\]

with dual

\[
\begin{align*}
\min & \quad \text{trace} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} Z \\
\text{s.t.} & \quad \text{trace} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} Z = 2 \\
& \quad \text{trace} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} Z = 2 \\
& \quad Z \succeq 0
\end{align*}
\]
Reformulations (2)

Denoting

\[ Z = \begin{bmatrix} z_{11} & z_{21} \\ z_{21} & z_{22} \end{bmatrix} \]

the linear trace constraints on \( Z \) can be written

\[
\begin{bmatrix} -1 & 0 & 1 \\ 0 & -2 & 0 \end{bmatrix} \begin{bmatrix} z_{11} \\ z_{21} \\ z_{22} \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}
\]

Particular solution and explicit null-space basis

\[
\begin{bmatrix} z_{11} \\ z_{21} \\ z_{22} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \bar{z}
\]

so we obtain the equivalent dual problem with less variables

\[
\begin{array}{ll}
\text{min} & 2 \bar{z} \\
\text{s.t.} & \begin{bmatrix} \bar{z} - 1 & -1 \\ -1 & \bar{z} + 1 \end{bmatrix} \preceq 0
\end{array}
\]

and primal

\[
\begin{array}{ll}
\text{max} & \text{trace} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \bar{X} \\
\text{s.t.} & \text{trace} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \bar{X} = 2 \\
& \bar{X} \succeq 0
\end{array}
\]
We can use the **Schur complement** to convert a non-linear matrix inequality into an LMI

\[
A(x) - B(x)C^{-1}(x)B'(x) \succeq 0 \\
C(x) \succ 0
\]

\[
\iff \\
\begin{bmatrix}
A(x) & B(x) \\
B(x) & C(x)
\end{bmatrix} \succeq 0 \\
C(x) \succ 0
\]

Issai Schur  
(1875 Mogilyov - 1941 Tel Aviv)
SOLVING LMIs

Denis Arzelier
www.laas.fr/~arzelier
arzelier@laas.fr

26-27 June 2017
History

Convex programming

- Logarithmic barrier function [K. Frisch 1955]
- Method of centers ([P. Huard 1967]

Interior-point (IP) methods for LP

- Ellipsoid algorithm [Khachiyan 1979]
  polynomial bound on worst-case iteration count
- IP methods for LP [Karmarkar 1984]
  improved complexity bound and efficiency - About 50% of commercial LP solvers

IP methods for SDP

- Self-concordant barrier functions [Nesterov, Nemirovski 1988], [Alizadeh 1991]
- IP methods for general convex programs (SDP and LMI)
  Academic and commercial solvers (MATLAB)
Interior point methods (1)

For the optimization problem

\[
\min_{x \in \mathbb{R}^n} f_0(x) \\
\text{s.t. } f_i(x) \geq 0 \quad i = 1, \cdots, m
\]

where the \( f_i(x) \) are twice continuously differentiable convex functions

Sequential minimization techniques: Reduction of the initial problem into a sequence of unconstraint optimization problems

[Fiacco - Mc Cormick 68]

\[
\min_{x \in \mathbb{R}^n} f_0(x) + \mu \phi(x)
\]

where \( \mu > 0 \) is a parameter sequentially decreased to 0 and the term \( \phi(x) \) is a barrier function

Barrier functions go to infinity as the boundary of the feasible set is approached
Interior point methods (2)

Descent methods

To solve an unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

we produce a minimizing sequence

$$x_{k+1} = x_k + t_k \Delta x_k$$

where $\Delta x_k \in \mathbb{R}^n$ is the step or search direction and $t_k \geq 0$ is the step size or step length.

A descent method consists in finding a sequence $\{x_k\}$ such that

$$f(x^*) \leq \cdots f(x_{k+1}) < f(x_k)$$

where $x^*$ is the optimum.

General descent method

0. $k = 0$; given starting point $x_k$
1. determine descent direction $\Delta x_k$
2. line search: choose step size $t_k > 0$
3. update: $k = k + 1; x_k = x_{k-1} + t_{k-1} \Delta x_{k-1}$
4. go to step 1 until a stopping criterion is satisfied
Interior point methods (3)

Newton’s method

A particular choice of search direction is the Newton step

\[ \Delta x = -\nabla^2 f(x)^{-1}\nabla f(x) \]

where
- \( \nabla f(x) \) is the gradient
- \( \nabla^2 f(x) \) is the Hessian

This step \( y = \Delta x \) minimizes the second-order Taylor approximation

\[ \hat{f}(x + y) = f(x) + \nabla f(x)'y + y'\nabla^2 f(x)y/2 \]

and it is the steepest descent direction for the quadratic norm defined by the Hessian

**Quadratic convergence** near the optimum
For the conic optimization problem
\[
\min_{x \in \mathbb{R}^n} \quad f_0(x) \\
\text{s.t.} \quad f_i(x) \preceq_K 0 \quad i = 1, \ldots, m
\]
suitable barrier functions are called self-concordant

Smooth convex 3-differentiable functions \( f \) with second derivative Lipschitz continuous w.r. to the local metric induced by the Hessian

\[
|f'''(x)| \leq 2f''(x)^{3/2}
\]

- goes to infinity as the boundary of the cone is approached
- can be efficiently minimized by Newton’s method
- Each convex cone \( K \) possesses a self-concordant barrier
- Such barriers are only computable for some special cones
Barrier function for LP (1)

For LP and positive orthant $\mathbb{R}^n_+$, the logarithmic barrier function

$$
\phi(y) = -\sum_{i=1}^{n} \log(y_i) = \log \prod_{i=1}^{n} y_i^{-1}
$$

is convex in the interior $y \succ 0$ of the feasible set and is instrumental to design IP algorithms

$$
\begin{align*}
\max_{\mu \in \mathbb{R}^p} & \quad b'y \\
\text{s.t.} & \quad c_i - a_i y \succeq 0, \quad i = 1, \cdots, m, \quad (y \in \mathcal{P})
\end{align*}
$$

$$
\phi(y) = -\log \prod_{i=1}^{m} (c_i - a_i y) = -\sum_{i=1}^{m} \log(c_i - a_i y)
$$

The optimum

$$
y_c = \arg \left[ \min_y \phi(y) \right]
$$

is called the analytic center of the polytope
Barrier function for LP (2)

Example

\[ J^*_1 = \max_{x, y} \quad 2x + y \]

\[ \text{s.t.} \quad x \geq 0 \quad y \geq 0 \quad x \leq 2 \]

\[ y \leq 2 \quad x + y \leq 3 \]

\[ \phi(x, y) = -\log(xy) - \log(2 - x) - \log(2 - y) - \log(3 - x - y) \]

\( (x_c, y_c) = \left( \frac{6 - \sqrt{6}}{5}, \frac{6 - \sqrt{6}}{5} \right) \)
Barrier function for an LMI (1)

Given an LMI constraint $F(x) \succeq 0$

Self-concordant barriers are smooth convex 3-differentiable functions $\phi : \mathbb{S}_+^n \to \mathbb{R}$ s.t. for $\overline{\phi}(\alpha) = \phi(X + \alpha H)$ for $X \succ 0$ and $H \in \mathbb{S}^n$

$$|\overline{\phi}'''(0)| \leq 2\overline{\phi}''(0)^{3/2}$$

Logarithmic barrier function

$$\phi(x) = -\log \det F(x) = \log \det F(x)^{-1}$$

This function is analytic, convex and self-concordant on $\{x : F(x) \succ 0\}$

The optimum

$$x_c = \arg \left[ \min_x \phi(x) \right]$$

is called the analytic center of the LMI
Barrier function for an LMI (2)
Example (1)

\[ F(x_1, x_2) = \begin{bmatrix} 1 - x_1 & x_1 + x_2 & x_1 \\ x_1 + x_2 & 2 - x_2 & 0 \\ x_1 & 0 & 1 + x_2 \end{bmatrix} \succeq 0 \]

Computation of analytic center:

\[ \nabla_{x_1} \log \det F(x) = 2 + 3x_2 + 6x_1 + x_2^2 = 0 \]

\[ \nabla_{x_2} \log \det F(x) = 1 - 3x_1 - 4x_2 - 3x_2^2 - 2x_1x_2 = 0 \]

\[ x_{1c} = -0.7989 \quad x_{2c} = 0.7458 \]
The barrier function $\phi(x)$ is flat in the interior of the feasible set and sharply increases toward the boundary.
IP methods for SDP (1)

Primal / dual SDP

\[
\begin{align*}
\min_Z & \quad -\text{trace}(F_0 Z) \\
\text{s.t.} & \quad -\text{trace}(F_i Z) = c_i \\
& \quad Z \succeq 0
\end{align*}
\]

\[
\begin{align*}
\min_{x, Y} & \quad c'x \\
\text{s.t.} & \quad Y + F_0 + \sum_{i=1}^{m} x_i F_i = 0 \\
& \quad Y \succeq 0
\end{align*}
\]

Remember KKT optimality conditions

\[
\begin{align*}
F_0 + \sum_{i=1}^{m} x_i F_i + Y &= 0 \quad Y \succeq 0 \\
\forall \ i \ & \text{trace} \ F_i Z + c_i = 0 \quad Z \succeq 0 \\
Z^* F(x^*) = -Z^* Y^* &= 0
\end{align*}
\]
IP methods for SDP (2)
The central path

Perturbed KKT optimality conditions = Centrality conditions

\[ F_0 + \sum_{i=1}^{m} x_i F_i + Y = 0 \quad Y \succeq 0 \]

\[ \forall i \text{ trace } F_i Z + c_i = 0 \quad Z \succeq 0 \]

\[ ZY = \mu 1 \]

where \( \mu > 0 \) is the centering parameter or barrier parameter

For any \( \mu > 0 \), centrality conditions have a unique solution \( Z(\mu), x(\mu), Y(\mu) \) which can be seen as the parametric representation of an analytic curve: The central path

The central path exists if the primal and dual are strictly feasible and converges to the analytic center when \( \mu \to 0 \)
IP methods for SDP (3)
Primal methods

\[
\min_{Z} -\text{trace}(F_0 Z) - \mu \log \det Z \\
\text{s.t. } \text{trace}(F_i Z) = -c_i
\]

where parameter \( \mu \) is sequentially decreased to zero.

Follow the primal central path approximately:
Primal path-following methods

The function \( f^\mu_p(Z) \)

\[
f^\mu_p(Z) = -\frac{1}{\mu} \text{trace}(F_0 Z) - \log \det Z
\]

is the primal barrier function and the primal central path corresponds to the minimizers \( Z(\mu) \) of \( f^\mu_p(Z) \)

- The projected Newton direction \( \Delta Z \)
- Updating of the centering parameter \( \mu \)
IP methods for SDP (4)
Dual methods (1)

\[
\begin{align*}
\min_{x,Y} & \quad c'x - \mu \log \det Y \\
\text{s.t.} & \quad Y + F_0 + \sum_{i=1}^{m} x_i F_i = 0
\end{align*}
\]

where parameter \( \mu \) is sequentially decreased to zero.
The function \( f^\mu_d(x,Y) \)

\[
f^\mu_d(x,Y) = \frac{1}{\mu} c'x - \log \det Y
\]
is the dual barrier function and the dual central path corresponds to the minimizers \( (x(\mu), Y(\mu)) \) of \( f^\mu_d(x,Y) \).

\( Y_k \succeq 0 \) ensured via Newton process:
- Large decreases of \( \mu \) require damped Newton steps
- Small updates allow full (deep) Newton steps
Dual methods (2)
Newton step for LMI

The centering problem is

$$\min \phi(x) = \frac{1}{\mu} c' x - \log \det(-F(x))$$

and at each iteration Newton step $\Delta x$ satisfies the linear system of equations (LSE)

$$H \Delta x = -g$$

where gradient $g$ and Hessian $H$ are given by

$$H_{ij} = \text{trace } F(x)^{-1} F_i F(x)^{-1} F_j$$

$$g_i = c_i/\mu - \text{trace } F(x)^{-1} F_i$$

LSE typically solved via Cholesky factorization or QR decomposition (near the optimum)

Nota: Expressions for derivatives of $\phi(x) = -\log \det F(x)$

Gradient:

$$(\nabla \phi(x))_i = -\text{trace } F(x)^{-1} F_i$$

$$= -\text{trace } F(x)^{-1/2} F_i F(x)^{-1/2}$$

Hessian:

$$(\nabla^2 \phi(x))_{ij} = \text{trace } F(x)^{-1} F_i F(x)^{-1} F_j$$

$$= \mu \text{trace } (F(x)^{-1/2} F_i F(x)^{-1/2})(F(x)^{-1/2} F_j F(x)^{-1/2})$$
IP methods for SDP (4)
Primal-dual methods (1)

\[
\begin{align*}
\min_{x, Y, Z} & \quad \text{trace } YZ - \mu \log \det YZ \\
\text{s.t.} & \quad -\text{trace } F_i Z = c_i \\
& \quad Y + F_0 + \sum_{i=1}^m x_i F_i = 0
\end{align*}
\]

Minimizers \((x(\mu), Y(\mu), Z(\mu))\) satisfy optimality conditions

\[
\begin{align*}
\text{trace } F_i Z &= -c_i \\
\sum_{i=1}^m x_i F_i + Y &= -F_0 \\
YZ &= \mu I \\
Y, Z &\succeq 0
\end{align*}
\]

The duality gap:

\[
-\text{trace}(F_0 Z) - c' x = \text{trace}(YZ) \geq 0
\]

is minimized along the central path
For primal-dual IP methods, primal and dual directions $\Delta Z$, $\Delta x$ and $\Delta Y$ must satisfy non-linear and over determined system of conditions

\[
\begin{align*}
\text{trace}(F_i \Delta Z) &= 0 \\
\sum_{i=1}^{m} \Delta x_i F_i + \Delta Y &= 0 \\
(Z + \Delta Z)(Y + \Delta Y) &= \mu I \\
Z + \Delta Z &\succeq 0 \\
\Delta Z &= \Delta Z' \\
Y + \Delta Y &\succeq 0
\end{align*}
\]

These centrality conditions are solved approximately for a given $\mu > 0$, after which $\mu$ is reduced and the process is repeated.

Key point is in linearizing and symmetrizing the latter equation.
The non-linear equation in the centrality conditions is replaced by

\[ H_P(\Delta ZY + Z\Delta Y) = \mu 1 - H_P(ZY) \]

where \( H_P \) is the linear transformation

\[ H_P(M) = \frac{1}{2} \left[ PMP^{-1} + P^{-1}M'P' \right] \]

for any matrix \( M \) and the scaling matrix \( P \) gives the symmetrization strategy.

Following the choice of \( P \), long list of primal-dual search directions, (AHO, HRVW, KSH, M, NT...), the most known of which is Nesterov-Todd's

Algorithms differ in how the symmetrized equations are solved and how \( \mu \) is updated (long step methods, dynamic updates of for predictor-corrector methods)
IP methods in general

Generally for LP, QP or SDP primal-dual methods outperform primal or dual methods

General characteristics of IP methods:

- **Efficiency**: About 5 to 50 iterations, almost independent of input data (problem), each iteration is a least-squares problem (well established linear algebra)
- **Theory**: Worst-case analysis of IP methods yields polynomial computational time
- **Structure**: Tailored SDP solvers can exploit problem structure

For more information see the Linear, Cone and SDP section at

[www.optimization-online.org](http://www.optimization-online.org)

and the Optimization and Control section at

SDP solvers

Primal-dual algorithms:
• **SeDuMi** (J. Sturm, I. Polik)
• **SDPT3** (K.C. Toh, R. Tütüncü, M. Todd)
• **CSDP** (B. Borchers)
• **SDPA** (M. Kojima and al.)
• **SMCP** (E. Andersen and L. Vandenberghe)
• **MOSEK** (E. Andersen)

Bundle methods:
• **ConicBundle** (C. Helmberg)

Dual-scaling potential reduction algorithms:
• **DSDP** (S. Benson, Y. Ye)

Barrier method and augmented Lagrangian:
• **PENSDP** (M. Kočvara, M. Stingl)
• **SDPLR** (S. Burer, R. Monteiro)
Matrices as variables

Generally, in control problems we do not encounter the LMI in canonical or semidefinite form but rather with matrix variables

Lyapunov’s inequality

\[ A'P + PA < 0 \quad P = P' > 0 \]

can be written in canonical form

\[ F(x) = F_0 + \sum_{i=1}^{m} F_i x_i < 0 \]

with the notations

\[ F_0 = 0 \quad F_i = A'B_i + B_iA \]

where \( B_i, i = 1, \ldots, n(n+1)/2 \) are matrix bases for symmetric matrices of size \( n \)

Most software packages for solving LMIs however work with canonical or semidefinite forms, so that a (sometimes time-consuming) pre-processing step is required
LMI solvers

Available under the Matlab environment

Projective method: project iterate on ellipsoid within PSD cone = least squares problem

- **LMI Control Toolbox** (P. Gahinet, A. Nemirovski)
  exploits structure with rank-one linear algebra
  warm-start + generalized eigenvalues
  originally developed for INRIA’s Scilab

LMI parser to SDP solvers

- **YALMIP** (Y. Löfberg)

See Helmberg’s page on SDP

and Mittelmann’s page on optimization software with benchmarks