Convex Optimization and LMIs

Kostas Margellos / Paul Goulart

University of Oxford
Convex Optimization

- Optimization programs
- Convex sets
- Convex functions
- Operations that preserve convexity
- Convex optimization programs

Linear Matrix Inequalities (LMIs)

- What do they look like?
- Are they convex?
- Who cares?
References

Convex Optimization & Duality Theory:

Convex Optimization, Cambridge University Press.

Bertsekas (2009)
Convex Optimization Theory, Athena Scientific.

Rockafellar (1970)

Linear Matrix Inequalities (LMIs):

Boyd, El Ghaoui, Feron & Balakrishnan (1994)
Linear Matrix Inequalities in System and Control Theory, SIAM.

VanAntwerp & Braatz (2000)
A tutorial on linear and bilinear matrix inequalities, J. Process Control.
A more common problem format:

\[
\min_{x \in \mathcal{X}} \quad f_0(x) \\
\text{subject to: } \quad f_i(x) \leq 0 \quad i = 1, \ldots, m \\
\quad h_i(x) = 0 \quad i = 1, \ldots, p
\]

- **Objective function** \( f_0 : \mathcal{X} \rightarrow \mathbb{R} \)
- **Domain** \( \mathcal{X} \subseteq \mathbb{R}^n \) of the objective function, from which the decision variable \( x := (x_1; x_2; \ldots; x_n) \) must be chosen.
- **Inequality constraint functions** \( f_i : \mathbb{R}^n \rightarrow \mathbb{R} \), for \( i = 1, \ldots, m \)
- **Equality constraint functions** \( h_i : \mathbb{R}^n \rightarrow \mathbb{R} \), for \( i = 1, \ldots, p \)

\( \Rightarrow \) Maximization fit the framework with a change of sign.
Consider the problem

\[ p^* = \min_{x \in \mathcal{X}} f(x) \]

- If \( p^* = -\infty \), then the problem is **unbounded below**.
- If the set \( \mathcal{X} \) is empty, then the problem is **infeasible** (and we set \( p^* = +\infty \)).
- If \( \mathcal{X} = \mathbb{R}^n \), the problem is **unconstrained**.
- There might be more than one solution. The set of solutions is:

\[ \arg \min_{x \in \mathcal{X}} f(x) := \{ x \in \mathcal{X} \mid f(x) = p^* \} \]
Geometric view

- Objective function contour
- Feasible set
- Local optimum
- Inequality constraint
- Objective function minimum (infeasible)
- \( x^* \) - Global optimum
Under convexity it is easier ...

**Linear Program (LP):**

$$
\min_x \quad c^T x \\
\text{subject to:} \quad Gx \leq h \quad Ax = b
$$

**Convex Quadratic Program (QP):**

$$
\min_x \quad \frac{1}{2} x^T Px + q^T x \\
\text{subject to:} \quad Gx \leq h \quad Ax = b
$$

$\Rightarrow$ **Convex programs:** Local optimum = Global optimum
Convex sets

**Definition (Convex Set)**

A set $\mathcal{X}$ is **convex** if and only if for any pair of points $x$ and $y$ in $\mathcal{X}$, any **convex combination** of $x$ and $y$ lies in $\mathcal{X}$:

$$\mathcal{X} \text{ is convex } \iff \lambda x + (1 - \lambda)y \in \mathcal{X}, \forall \lambda \in [0, 1], \forall x, y \in \mathcal{X}$$

**Interpretation:** All line segments starting and ending in $\mathcal{X}$ stay within $\mathcal{X}$.
Convex sets

Definitions (Hyperplanes and halfspaces)

A hyperplane is defined by \( \{ x \in \mathbb{R}^n \mid a^\top x = b \} \) for \( a \neq 0 \), where \( a \in \mathbb{R}^n \) is the normal vector to the hyperplane.

A halfspace is defined by \( \{ x \in \mathbb{R}^n \mid a^\top x \leq b \} \) for \( a \neq 0 \). It can either be open (strict inequality) or closed (non-strict inequality).

For \( n = 2 \), hyperplanes define lines. For \( n = 3 \), hyperplanes define planes.

A hyperplane

A closed halfspace
**Convex sets**

**Definitions (Polyhedra and polytopes)**

A **polyhedron** is the intersection of a *finite* number of closed halfspaces:

\[ \mathcal{X} = \{ x \mid a_1^\top x \leq b_1, \ a_2^\top x \leq b_2, \ldots, \ a_m^\top x \leq b_m \} = \{ x \mid Ax \leq b \} \]

where \( A := [a_1, a_2, \ldots, a_m]^\top \) and \( b := [b_1, b_2, \ldots, b_m]^\top \).

A **polytope** is a *bounded* polyhedron.

Polyhedra and polytopes are always convex.
Norms

Definition (Vector norm)

A norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying the following conditions:

- $f(x) \geq 0$ and $f(x) = 0 \Rightarrow x = 0$.
- $f(tx) = |t|f(x)$ for scalar $t$.
- $f(x + y) \leq f(x) + f(y)$, for all $x, y \in \mathbb{R}^n$.

Definition ($\ell_p$ norm)

The $\ell_p$ norm on $\mathbb{R}^n$ is denoted $\|x\|_p$, and is defined for any $p \geq 1$ by

$$\|x\|_p := \left[ \sum_{i=1}^{n} |x_i|^p \right]^{1/p}$$
Norms

By far the most common $\ell_p$ norms are:

- $p = 2$ (Euclidean norm):
  \[ \|x\|_2 = \sqrt{\sum_i x_i^2} \]

- $p = 1$ (Sum of absolute values):
  \[ \|x\|_1 = \sum_i |x_i| \]

- $p = \infty$ (Largest absolute value):
  \[ \|x\|_\infty = \max_i |x_i| \]

The **norm ball**, defined by \( \{ x \mid \|x - x_c\| \leq r \} \) where \( x_c \) is the centre of the ball and \( r \geq 0 \) is the radius, is always convex for any norm.
Definition (Ellipsoid)

An ellipsoid is a set defined as

\[ \mathcal{E} = \{ x \mid (x - x_c)^\top A^{-1} (x - x_c) \leq 1 \} , \]

where \( x_c \) is the centre of the ellipsoid, and \( A \succ 0 \).

Alternatively,

\[ E = \{ x \mid T(x) \leq 0 \} \]

where

\[ T(x) = x^\top Ax + 2x^\top b + c, \quad \text{with} \quad A = A^\top \succ 0. \]
**Intersection of convex sets**

**Theorem**

*The intersection of two or more convex sets is itself convex.*

**Proof (for two sets):** Consider any two points \( a \) and \( b \) which *both* lie in *both* of two convex sets \( \mathcal{X} \) and \( \mathcal{Y} \). For any \( \lambda \in [0, 1] \), \( \lambda a + (1 - \lambda) b \) is in both \( \mathcal{X} \) and \( \mathcal{Y} \). Therefore \( \lambda a + (1 - \lambda) b \in \mathcal{X} \cap \mathcal{Y}, \forall \lambda \in [0, 1] \). This satisfies the definition of convexity for set \( \mathcal{X} \cap \mathcal{Y} \).

Think of simultaneous constraint satisfaction.
Convex functions

Definitions (Convex function)

A function $f : \text{dom}(f) \rightarrow \mathbb{R}$ is **convex** if and only if its domain $\text{dom}(f)$ is convex and

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \quad \forall \lambda \in (0, 1), \quad \forall x, y \in \text{dom}(f)$$

The function $f$ is **strictly convex** if this inequality is strict.
Convex functions – 1st-order condition

A differentiable function \( f : \text{dom}(f) \rightarrow \mathbb{R} \) with a convex domain is **convex if and only if**

\[
f(y) \geq f(x) + \nabla f(x)^\top (y - x), \quad \forall x, y \in \text{dom}(f)
\]

i.e. a first order approximator of \( f \) around any point \( x \) is a global underestimator of \( f \).

The gradient is given by

\[
\nabla f(x) = \left[ \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right]^\top
\]
Convex functions – 2nd-order condition

A twice-differentiable function $f : \text{dom}(f) \to \mathbb{R}$ is convex if and only if its domain $\text{dom}(f)$ is convex and

$$
\nabla^2 f(x) \succeq 0, \quad \forall x \in \text{dom}(f),
$$

where the Hessian $\nabla^2 f(x)$ is defined by

$$
\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}
$$

If $\text{dom}(f)$ is convex and $\nabla^2 f(x) \succ 0$ for all $x \in \text{dom}(f)$, then $f$ is strictly convex.
Operations that preserve convexity

**Theorem (Non-negative weighted sum)**

If $f$ is a function convex, then $\alpha f$ is convex for $\alpha \geq 0$. For several convex functions $f_i$, $\sum_i \alpha_i f_i$ is convex if all $\alpha_i \geq 0$.

**Theorem (Composition with affine function)**

If $f$ is a convex function, then $f(Ax + b)$ is convex.

**Example:** $||Ax - b||$ is convex for any norm; Exponential functions.

**Theorem (Pointwise maximum)**

If $f_1, \ldots, f_m$ are convex functions, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.

**Example:** Piecewise linear functions $\max_{i=1, \ldots, m}\{a_i^T x + b\}$ are convex.
Convex optimization program – standard form

A standard form convex optimization problem:

\[
\min_{x \in \mathcal{X}} f_0(x)
\]

subject to:

\[
f_i(x) \leq 0 \quad i = 1, \ldots, m
\]

\[
Ax = b \quad A \in \mathbb{R}^{p \times m}
\]

This problem is convex if:

- The domain \( \mathcal{X} \) is a convex set.
- The objective function \( f_0 \) is a convex function.
- The inequality constraint functions \( f_i \) are all convex.
- The equality constraint functions \( h_i(x) = a_i^T x \) are all affine.
Theorem

For a convex optimization problem, every locally optimal solution is globally optimal.

Proof:

- Assume that $x$ is locally optimal, but not globally optimal.
- Therefore there is some other point $y$ such that $f(y) < f(x)$.
- $x$ locally optimal implies that there is some $R > 0$ such that
  \[ \|z - x\|_2 < R \Rightarrow f(x) \leq f(z) \]
- The problem can’t be convex.
What are LMIs?

A **Linear Matrix Inequality** (LMI) is a constraint of the form:

\[ x_1 A_1 + x_2 A_2 + \cdots + x_n A_n \preceq B \]

where the matrices \( A_1, \ldots, A_n, B \in \mathbb{R}^{m \times m} \) are all symmetric.

- This is a constraint that imposes matrix

\[
B - \sum_{i}^{n} x_i A_i
\]

to be positive semidefinite (positive definite if \( \preceq \) replaced by \( < \)).

- It is equivalent to imposing \( m \) polynomial inequalities
  - *Not* element-wise constraints.
  - All leading principle minors of this matrix are positive.
What are LMIs?

A **Linear Matrix Inequality** (LMI) is a constraint of the form:

\[ x_1 A_1 + x_2 A_2 + \cdots + x_n A_n \preceq B \]

where the matrices \((A_1, \ldots, A_n, B)\) are all symmetric.

1. Consider the constraint

\[ Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \preceq 0 \]

2. This is equivalent to 2 inequalities:

\[ Q_{11} \geq 0 \]

\[ \det(Q) \geq 0 \iff Q_{11} Q_{22} - Q_{12} Q_{21} \geq 0 \]
General form LMIs

Example 1: \( y - x^2 > 0, \ y > 0 \iff \begin{bmatrix} y & x \\ x & 1 \end{bmatrix} \succ 0 \)

- Check leading principle minors
- That is an LMI; rewrite as

\[
\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + y \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \succ 0
\]

Example 2: \( x_1^2 + x_2^2 < 1 \iff \begin{bmatrix} 1 & 0 & x_1 \\ 0 & 1 & x_2 \\ x_1 & x_2 & 1 \end{bmatrix} \succ 0 \)

- Leading principle minors are: \( 1 > 0 \), \( \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} > 0 \), and

\[
1 \cdot \det \begin{bmatrix} 1 & x_2 \\ x_2 & 1 \end{bmatrix} - 0 \cdot \det \begin{bmatrix} 1 & x_1 \\ x_1 & 1 \end{bmatrix} + x_1 \cdot \det \begin{bmatrix} 0 & 1 \\ x_1 & x_2 \end{bmatrix} > 0
\]
Consider a congruence transformation $x = Mz$, with $M$ nonsingular

$$A > 0 \iff x^T Ax > 0 \text{ for all } x \neq 0$$

$$\iff z^T M^T AMz > 0 \text{ for all } z \neq 0, M \text{ nonsingular}$$

$$\iff M^T AM > 0$$

LMIs can be then represented in multiple ways; their feasible sets however remain the same

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \succeq 0 \iff \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \succeq 0$$

$$\iff \begin{bmatrix} D & C \\ B & A \end{bmatrix} \succeq 0$$
The following LMI constraint is convex.

\[ F(x) = B - \sum_{i=1}^{n} x_i A_i \succeq 0 \]

Proof: Let \( x, y \) such that \( F(x), F(y) \succeq 0 \), and \( \lambda \in (0, 1) \).

\[
F(\lambda x + (1 - \lambda)y) = B - \sum_{i} (\lambda x_i + (1 - \lambda)y_i) A_i \\
= \lambda B + (1 - \lambda)B - \lambda \sum_{i} x_i A_i - (1 - \lambda) \sum_{i} y_i A_i \\
= \lambda F(x) + (1 - \lambda) F(y) \\
\geq 0
\]
LMI constraints are convex

**Theorem**

The following LMI constraint is convex.

\[
F(x) = B - \sum_{i=1}^{n} x_i A_i \succeq 0
\]

**Alternative proof:** We want to show that the set \( \{ x : F(x) \succeq 0 \} \) is convex. We have that...

\[
\{ x : F(x) \succeq 0 \} = \{ x : z^\top F(x) z \succeq 0, \text{ for all } z \}
\]

\[
= \bigcap_z \{ x : z^\top F(x) z \succeq 0 \}
\]

... but this is an infinite intersection of sets affine in \( x \) ... so it is convex!

- LMI much harder than linear constraints – an infinite number of them!
- Result can be piecewise affine – LMI nonlinear!
Why are LMIs interesting?

Linear Matrix Inequalities:

- Appear in many common control design problems (more later on)
- Most of the problems presented so far can be written using LMI constraints

Linear constraints

\[ Ax \leq b \iff \text{diag}(Ax) \leq \text{diag}(b) \]

Quadratic constraints (explained later...)

\[ x^\top Q x + b^\top x + c \geq 0, \quad Q \succ 0 \iff \begin{bmatrix} c + b^\top x & x^\top \\ x & -Q^{-1} \end{bmatrix} \succeq 0 \]
Summary

1. Introduction to convex optimization
   - Under convexity: local = global optima
   - Recognizing convexity makes life easier
   - Interplay between convex functions and sets (epigraphic reformulation)

2. Linear Matrix Inequalities (LMIs)
   - Nonlinear constraints
   - LMI constraints are convex!
   - Generalize many of the well known constraints (e.g. linear, quadratic)
Duality Theory

- The Lagrangian function
- The dual problem
- Weak and strong duality
- Optimality conditions
- Game theoretic view

LMIs in optimization

- Semidefinite programming (SDP)
- The dual of an SDP
Consider the following optimization program

\[
\text{(SDP)}: \quad \begin{align*}
\min \quad & c^\top x \\
\text{subject to:} \quad & x_1 A_1 + x_2 A_2 + \cdots x_n A_n \leq B
\end{align*}
\]

where the matrices \((A_1, \ldots, A_n, B)\) are all symmetric.

- We could also have equality constraints
- Optimization over LMI constraints

Why is this class of optimization programs interesting?

- Semidefinite programming (SDP)
- Many control analysis and synthesis problems can be written as SDPs
- Most of the problems presented so far can be written as SDPs
Consider the following optimization program

\[
\begin{align*}
\text{min} & \quad c^\top x \\
\text{(SDP)}: & \quad \text{subject to: } x_1 A_1 + x_2 A_2 + \cdots x_n A_n \leq B
\end{align*}
\]

where the matrices \((A_1, \ldots, A_n, B)\) are all symmetric.

- Assume we are interested in the optimal value \(p^*\) of (SDP)
- Can we construct a lower bound for \(p^*\), i.e. \(d^* \leq p^*\), by solving another problem?
- This problem, called dual, might sometimes be easier to solve

To do this we first need some machinery – Duality Theory
The Lagrangian function

Recall our standard form (primal) optimization problem:

\[
\begin{align*}
\min_{x \in \mathcal{X}} & \quad f_0(x) \\
\text{(P)} : & \quad \text{subject to: } f_i(x) \leq 0 \quad i = 1 \ldots m \\
& \quad h_i(x) = 0 \quad i = 1 \ldots p
\end{align*}
\]

with (primal) decision variable \( x \), domain \( \mathcal{X} \) and optimal value \( p^* \).

**Lagrangian Function:** \( L : \mathcal{X} \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \)

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

- \( \lambda_i \) : inequality Lagrange multiplier for \( f_i(x) \leq 0 \).
- \( \nu_i \) : equality Lagrange multiplier for \( h_i(x) = 0 \).
- Lagrangian: weighted sum of the objective and constraint functions.
The dual function \( g : \mathbb{R}^m \times \mathbb{R}^p \) is

\[
g(\lambda, \nu) = \inf_{x \in X} L(x, \lambda, \nu)
= \inf_{x \in X} \left[ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right]
\]

The dual function \( g(\lambda, \nu) \) is always a \textbf{concave} function.

\( g(\lambda, \nu) \) is the pointwise infimum of affine functions.

Do you recall pointwise maximum?
The **dual function** $g : \mathbb{R}^m \times \mathbb{R}^p$ is

$$g(\lambda, \nu) = \inf_{x \in \mathcal{X}} L(x, \lambda, \nu)$$

$$= \inf_{x \in \mathcal{X}} \left[ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right]$$

The dual function generates lower bounds for the primal optimal value, i.e. $g(\lambda, \nu) \leq p^*$ for $\lambda \geq 0$:

**Proof:**

For any primal feasible solution $\bar{x}$: $\sum_{i=1}^{m} \lambda_i f_i(\bar{x}) + \sum_{i=1}^{p} \nu_i h_i(\bar{x}) \leq 0$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{X}} L(x, \lambda, \nu) \leq L(\bar{x}, \lambda, \nu) \leq f_0(\bar{x}) \text{ for all } \bar{x}$$

$$g(\lambda, \nu) \leq \inf_{x \in \mathcal{X}} f_0(x) = p^*$$

- $g(\lambda, \nu)$ might be $-\infty$; Non-trivial if $\text{dom } g := \{\lambda, \nu \mid g(\lambda, \nu) > -\infty\}$
The dual problem

Every $\nu \in \mathbb{R}^p$, $\lambda \geq 0$ produces a lower bound for $p^*$ using the dual function.
Which is the best?

$$\begin{align*}
\max_{\lambda, \nu} & \quad g(\lambda, \nu) \\
\text{subject to:} & \quad \lambda \geq 0
\end{align*}$$

- Problem (D) is **convex**, even if (P) is not.
- Problem (D) has optimal value $d^* \leq p^*$.
- The point $(\lambda, \nu)$ is **dual feasible** if $\lambda \geq 0$ and $(\lambda, \nu) \in \text{dom } g$.
- Often impose the constraint $(\lambda, \nu) \in \text{dom } g$ explicitly in (D).
Example: Dual of LPs

\[
(P) : \quad \begin{align*}
\min_{x \in \mathbb{R}^n} & \quad c^T x \\
\text{subject to:} & \quad Ax = b \\
& \quad Cx \leq d
\end{align*}
\]

The dual function is

\[
g(\lambda, \nu) = \min_{x \in \mathbb{R}^n} \left[ c^T x + \nu^T (Ax - b) + \lambda^T (Cx - d) \right]
\]

\[
= \min_{x \in \mathbb{R}^n} \left[ (A^T \nu + C^T \lambda + c)^T x - b^T \nu - d^T \lambda \right]
\]

\[
= \begin{cases} 
-b^T \nu - d^T \lambda & \text{if } A^T \nu + C^T \lambda + c = 0 \\
-\infty & \text{otherwise}
\end{cases}
\]
Example: Dual of LPs – (cont’d)

\[
\begin{align*}
(P) : \quad \min_{x \in \mathbb{R}^n} & \quad c^T x \\
\text{subject to:} & \quad Ax = b \\
& \quad Cx \leq d
\end{align*}
\]

The dual problem is

\[
\begin{align*}
(D) : \quad \max_{\lambda, \nu} & \quad -b^T \nu - d^T \lambda \\
\text{subject to:} & \quad A^T \nu + C^T \lambda + c = 0 \\
& \quad \lambda \geq 0
\end{align*}
\]

- Lower bound property:
  \[-b^T \nu - d^T \lambda \leq p^* \text{ whenever } \lambda \geq 0.\]
- The dual of a linear program is also a linear program.
Weak and strong duality

Weak Duality

- It is **always** true that $d^* \leq p^*$.

- Sometimes the dual is much easier to solve than the primal (or vice-versa).

- Example: The dual of an MILP (difficult to solve) is a standard LP (easy to solve).

Strong Duality

- It is **sometimes** true that $d^* = p^*$.

- Strong duality usually holds for convex problems.

- Strong duality usually does not hold for non-convex problems.

- Can impose conditions on convex problems to guarantee that $d^* = p^*$.
Strong duality for convex problems

An optimization problem with $f_0$ and all $f_i$ convex:

$$\min \ f_0(x)$$

$$\text{(P)} : \ \text{subject to: } f_i(x) \leq 0 \quad i = 1 \ldots m$$

$$Ax = b \quad A \in \mathbb{R}^{p \times n}$$

**Slater Condition**

If there is at least one strictly feasible point, i.e.

$$\left\{ x \ \bigg| \ Ax = b, \ f_i(x) < 0, \ \forall i \in \{1, \ldots, m\} \right\} \neq \emptyset$$

Then $p^* = d^*$.

- Stronger version: Only the nonlinear functions $f_i(x)$ must be strictly satisfiable (non-empty interior).
- Other constraint qualification conditions exist.
Primal and dual solution properties

Assume that strong duality holds, with optimal solution $x^*$ and $(\lambda^*, \nu^*)$.

- From strong duality, $d^* = p^* \Rightarrow g(\lambda^*, \nu^*) = f_0(x^*)$.

- From the definition of the dual function:

$$f_0(x^*) = g(\lambda^*, \nu^*) = \min_x \left\{ f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right\}$$

$$\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \quad \text{[weak duality]}$$

$$\implies f_0(x^*) = g(\lambda^*, \nu^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*)$$

$$\implies \lambda_i^* = 0 \text{ for every } f_i(x^*) < 0.$$  

$$\implies f_i(x^*) = 0 \text{ for every } \lambda_i^* > 0. \quad \text{Complementary slackness}$$
Karush-Kuhn-Tucker (KKT) optimality conditions

Assume that all $f_i$ and $h_i$ are differentiable. **Necessary** conditions for optimality:

1) **Primal Feasibility:**

$$f_i(x^*) \leq 0 \quad i = 1, \ldots, m$$
$$h_i(x^*) = 0 \quad i = 1 \ldots, p$$

2) **Dual Feasibility:**

$$\lambda^* \geq 0$$

3) **Complementary Slackness:**

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

4) **Stationarity:**

$$\nabla_x L(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$
KKT optimality conditions – Geometric view

Assume inequality constraints only.

Rewrite stationarity condition as:

\[-\nabla f_0(x) = \sum_{i=1}^{m} \lambda_i \nabla f_i(x)\]

*Direction of steepest descent is in convex cone spanned by constraint gradients \(\nabla g_i\).*
KKT optimality conditions – Convex Programs

For a convex optimization problem:

1) If \((x^*, \lambda^*, \nu^*)\) satisfy the KKT conditions, then \(p^* = d^*\).
   - \(p^* = f_0(x^*) = L(x^*, \lambda^*, \nu^*)\) (due to complementary slackness)
   - \(d^* = g(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)\) (due to convexity of the functions and stationarity)

2) If the Slater condition holds, then
   - \(x^*\) is optimal if and only if there exist \((\lambda^*, \nu^*)\) satisfying the KKT conditions.
Example: KKT optimality conditions for QPs

Consider a (convex) quadratic program with $Q \succeq 0$:

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad \frac{1}{2} x^\top Q x + c^\top x \\
\text{(P)}: \quad & \text{subject to: } A x = b \\
& \quad x \geq 0
\end{align*}
\]

The Lagrangian is $L(x, \lambda, \nu) = \frac{1}{2} x^\top Q x + c^\top x + \nu^\top (A x - b) - \lambda^\top x$.

The KKT conditions are:

\[
\nabla_x L(x, \lambda, \nu) = Q x + A^\top \nu - \lambda + c = 0 \quad \text{[stationarity]}
\]

\[
A x = b \quad \text{[primal feasibility]}
\]

\[
x \geq 0 \quad \text{[primal feasibility]}
\]

\[
\lambda \geq 0 \quad \text{[dual feasibility]}
\]

\[
x_i \lambda_i = 0 \quad i = 1 \ldots n \quad \text{[complementarity]}
\]
Game theoretic view

Assume inequality constraints only.

We have that for all $x$

$$
\max_{\lambda \geq 0} L(x, \lambda) = \max_{\lambda \geq 0} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)
$$

$$
= \begin{cases} 
  f_0(x) & \text{if } f_i(x) \leq 0 \text{ for all } i; \\
  \infty & \text{otherwise.}
\end{cases}
$$

Since this holds for all $x$, we then have that

$$
p^* = \min_{x \in \mathcal{X}} \max_{\lambda \geq 0} L(x, \lambda)
$$

$$
d^* = \max_{\lambda \geq 0} \min_{x \in \mathcal{X}} L(x, \lambda)
$$
Game theoretic view

- Game between **primal (Peter)** and **dual (Debbie)** variables:

  \[ p^* = \min_x \max_\lambda L(x, \lambda) \]
  \[ d^* = \max_\lambda \min_x L(x, \lambda) \]

- Consider the \( d^* \) game – **Debbie** plays first, **Peter** plays second

  \[ d^* = \max_\lambda \min_x L(x, \lambda) \leq \text{any value} \]
  \[ = \forall \lambda \ \exists x \ L(x, \lambda) \leq \text{any value} \]
  \[ = \exists x(\lambda) \ \forall \lambda \ L(x, \lambda) \leq \text{any value} \quad [x(\cdot) \text{ is a strategy}] \]
  \[ \leq \exists x \ \forall \lambda \ L(x, \lambda) \leq \text{any value} \]
  \[ = \min_x \max_\lambda L(x, \lambda) \]
  \[ = p^* \]
Game theoretic view

- Game between primal (Peter) and dual (Debbie) variables:
  
  \[
  p^* = \min_x \max_\lambda L(x, \lambda) \\
  d^* = \max_\lambda \min_x L(x, \lambda)
  \]

- If Peter plays second ⇒
  
  \[
  d^* \leq p^* \quad \text{[weak duality]}
  \]

- Duality gap corresponds to the advantage of Peter

- Strong duality = Zero duality gap
  
  ⇒ No advantage for any of the players
Primal SDP problem:

$$\min \ c^\top x$$

subject to: \( x_1 A_1 + x_2 A_2 + \cdots x_n A_n \preceq B \)

where the matrices \((A_1, \ldots, A_n, B)\) are all symmetric.

Lagrangian:

$$\mathcal{L}(x, \Lambda) = c^\top x + \sum_i \langle \Lambda, A_i \rangle x_i - \langle \Lambda, B \rangle,$$

where \( \langle X, Y \rangle = \text{trace}(X^\top Y) = \sum_{i,j} X_{ij} Y_{ij} \).

This fact relies on “dual cone” arguments, and the fact that trace is the inner product for matrices.
Semidefinite programming

**Primal SDP problem:**

\[
\begin{aligned}
\min & \quad c^\top x \\
\text{subject to:} & \quad x_1 A_1 + x_2 A_2 + \cdots x_n A_n \preceq B
\end{aligned}
\]

where the matrices \((A_1, \ldots, A_n, B)\) are all symmetric.

**Lagrangian:**

\[
\mathcal{L}(x, \Lambda) = c^\top x + \sum_i \langle \Lambda, A_i \rangle x_i - \langle \Lambda, B \rangle \\
= \sum_i \left(c_i + \langle \Lambda, A_i \rangle \right) x_i - \langle \Lambda, B \rangle
\]

**Dual function:**

\[
g(\lambda) = \begin{cases} 
-\langle \Lambda, B \rangle & \text{if } c_i + \langle \Lambda, A_i \rangle = 0 \text{ for } i = 1 \ldots n \\
-\infty & \text{otherwise}
\end{cases}
\]
Semidefinite programming

**Primal SDP problem:**

\[
\begin{align*}
\min & \quad c^\top x \\
\text{subject to:} & \quad x_1 A_1 + x_2 A_2 + \cdots + x_n A_n \preceq B
\end{align*}
\]

where the matrices \((A_1, \ldots, A_n, B)\) are all symmetric.

**Dual function:**

\[
g(\lambda) = \begin{cases} 
-\langle \Lambda, B \rangle & \text{if } c_i + \langle \Lambda, A_i \rangle = 0 \text{ for } i = 1 \ldots n \\
-\infty & \text{otherwise}
\end{cases}
\]

**The dual problem:**

\[
\begin{align*}
\max & \quad -\langle B, \Lambda \rangle \\
\text{subject to:} & \quad \langle A_i, \Lambda \rangle = -c_i, \text{ for all } i \\
& \quad \Lambda \succeq 0
\end{align*}
\]
Semidefinite programming

Primal SDP problem:
\[ \min c^T x \]
subject to: \( x_1 A_1 + x_2 A_2 + \cdots + x_n A_n \leq B \)

The dual problem:
\[ \max -\langle B, \Lambda \rangle \]
subject to: \( \langle A_i, \Lambda \rangle = -c_i, \text{ for all } i \)
\[ \Lambda \succeq 0 \]

Weak duality:
\[ p^* - d^* = c^T x + \langle B, \Lambda \rangle \quad \text{[primal feasibility]} \]
\[ \geq c^T x + \sum_i \langle A_i, \Lambda \rangle x_i \quad \text{[dual feasibility]} \]
\[ = \sum_i c_i x_i - \sum_i c_i x_i = 0 \]
Semidefinite programming

Primal SDP problem:

$$\min \ c^\top x$$

subject to: $x_1A_1 + x_2A_2 + \cdots x_nA_n \preceq B$

The dual problem:

$$\max \ -\langle B, \Lambda \rangle$$

subject to: $\langle A_i, \Lambda \rangle = -c_i$, for all $i$

$$\Lambda \succeq 0$$

Weak duality: $p^* - d^* \geq 0$

Strong duality:

Also true under Slater's condition (constraint qualification). Constraints in the primal need to be satisfied with $\prec$ instead of $\preceq$. 
Summary

1. Duality Theory
   - Construct $d^* \leq p^*$ in three steps
     1. Construct the Lagrangian (lift and weight constraints in the objective)
     2. Construct dual function and “eliminate” primal variables
     3. Formulate dual problem (don’t forget constraints on dual variables)
   - Optimality conditions
   - Geometric and gaming interpretation of duality

2. LMIs in optimization
   - Semidefinite programming (SDP)
   - Construct the dual of an SDP (similar procedure with linear programs)
   - Weak duality, strong duality under Slater’s condition
Reformulation in LMIs

- The Schur complement
  - Non-obvious LMIs
  - From nonlinear constraints to LMIs
- The S-procedure
  - From quadratic implications to LMIs
  - Turning set containment arguments in LMIs

LMIs for stability & controller synthesis

- Recap of stability theorems
- Lyapunov matrix inequality
- Example for nonlinear system stability
- Controller synthesis by means of an example
Non-obvious LMIs

Some cases (like the QP) are harder to write as LMIs.
The Schur complement provides the means to do so

**Schur complement:** Turns a nonlinear constraint into an LMI

**Theorem (Schur complement)**

Assume that $Q(x) = Q(x)^T$, $R(x) = R(x)^T$: affine functions of $x$. We then have that

\[
R(x) \succ 0 \text{ and } Q(x) - S(x)R(x)^{-1}S(x)^T \succ 0
\]

\[
\iff \begin{bmatrix} Q(x) & S(x) \\ S(x)^T & R(x) \end{bmatrix} \succ 0
\]
Schur complement

Schur complement: The non-strict case

Assume that \( Q(x) = Q(x)^\top, R(x) = R(x)^\top \succ 0 \): affine functions of \( x \)

We then have that

\[
Q(x) - S(x)R(x)^{-1}S(x)^\top \succeq 0 \iff \begin{bmatrix} Q(x) & S(x) \\ S(x)^\top & R(x) \end{bmatrix} \succeq 0
\]

Example 1:

\[
\|A\|_2 \leq t \iff A^\top A \preceq t^2 I, \quad t \geq 0 \iff \begin{bmatrix} tl & A^\top \\ A & tl \end{bmatrix} \succeq 0
\]

Example 2: The QP (recall from earlier)

\[
x^\top Qx + b^\top x + c \geq 0, \quad Q \succ 0 \iff \begin{bmatrix} c + b^\top x & x^\top \\ x & -Q^{-1} \end{bmatrix} \succeq 0
\]
Schur complement – Proof for the strict case

Proof of (\(\Leftrightarrow\)):

Assume \(\begin{bmatrix} Q(x) & S(x) \\ S(x)^\top & R(x) \end{bmatrix} \succ 0\). For all \([u \ v] \neq 0\) we have

\[
F(u, v) = \begin{bmatrix} u \\ v \end{bmatrix}^\top \begin{bmatrix} Q(x) & S(x) \\ S(x)^\top & R(x) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} > 0
\]

Considering \(u = 0\) we have

\[
F(0, v) = v^\top R(x) v > 0, \text{ for all } v \neq 0 \Rightarrow R(x) \succ 0
\]

Consider now \(v = -R(x)^{-1} S(x)^\top u\), with \(u \neq 0\)

\[
F(u, v) = u^\top (Q(x) - S(x)R(x)^{-1}S(x)^\top) u > 0, \text{ for all } u \neq 0
\]

\(\Rightarrow Q(x) - S(x)R(x)^{-1}S(x)^\top \succ 0\)
Schur complement – Proof for the strict case

Proof of \((\Rightarrow)\):

Now assume \(R(x) \succ 0\) and \(Q(x) - S(x)R(x)^{-1}S(x)^\top \succ 0\), and as before

\[
F(u, v) = \begin{bmatrix} u \\ v \end{bmatrix}^\top \begin{bmatrix} Q(x) & S(x) \\ S(x)^\top & R(x) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}
\]

Fix \(u\) and minimize over \(v\): \(\nabla_v F(u, v) = 2R(x)v + 2S(x)^\top u = 0\). Since \(R(x) \succ 0\), we have that \(v^* = -R(x)^{-1}S(x)^\top u\). Substitute it in the expression of \(F(u, v)\) to obtain

\[
F(u) = u^\top (Q(x) - S(x)R(x)^{-1}S(x)^\top) u
\]

Since \(Q(x) - S(x)R(x)^{-1}S(x)^\top \succ 0\), \(u^* = 0\) minimizes \(F(u)\). As a result, \((u^*, v^*) = (0, 0)\) and \(F(u^*, v^*) = 0\).

Hence, \(F(u, v) \succ 0\) for all \(u, v \neq 0\) \(\Rightarrow \begin{bmatrix} Q(x) & S(x) \\ S(x)^\top & R(x) \end{bmatrix} \succ 0\).
Schur complement – Ellipsoidal inequality

Assume that $Q(x) = Q(x)^\top$, $R(x) = R(x)^\top \succeq 0$: affine functions of $x$. We then have that

$$Q(x) - S(x)R(x)^{-1}S(x)^\top \succeq 0 \iff \begin{bmatrix} Q(x) & S(x) \\ S(x)^\top & R(x) \end{bmatrix} \succeq 0$$

Consider the ellipsoid

$$(x - x_c)^\top A^{-1}(x - x_c) \leq 1, \quad A = A^\top \succ 0$$

(... and recall that it is convex).

Setting $Q(x) = 1$, $R(x) = A$ and $S(x) = (x - x_c)^\top$:

$$\begin{bmatrix} 1 & (x - x_c)^\top \\ (x - x_c) & A \end{bmatrix} \succeq 0$$
Schur complement – Maximum singular value

Assume that \( Q(x) = Q(x)^\top \), \( R(x) = R(x)^\top \succeq 0 \): affine functions of \( x \). We then have that

\[
Q(x) - S(x)R(x)^{-1}S(x)^\top \succeq 0 \iff \begin{bmatrix} Q(x) & S(x) \\ S(x)^\top & R(x) \end{bmatrix} \succeq 0
\]

Let \( A(x) \): affine in \( x \) and real valued. Let also \( \bar{\sigma}[A(x)] \) be the maximum singular value of \( A(x) \), i.e. the square root of the largest eigenvalue of \( A(x)A(x)^\top \), i.e. \( \bar{\lambda}[A(x)A(x)^\top]^{1/2} \).

\[
\bar{\sigma}(A(x)) \leq 1 \iff \bar{\lambda}[A(x)A(x)^\top] \leq 1 \\
\iff A(x)A(x)^\top \preceq I \\
\iff I - A(x)A(x)^{-1}A(x)^\top \succeq 0 \\
\iff \begin{bmatrix} I & A(x) \\ A(x)^\top & I \end{bmatrix} \succeq 0
\]
S-procedure: Turns quadratic implications to LMIs

Consider two quadratic functions

\[
f_0(x) = x^\top A_0 x + 2x^\top b_0 + c_0
\]
\[
f(x) = x^\top A x + 2x^\top b + c,
\]

where all matrices/vectors are given, and \( A_0 = A_0^\top, A = A^\top \).

Problem: When is it true that one quadratic inequality implies another? In other words, when does

\[
f(x) \geq 0, \ x \neq 0 \Rightarrow f_0(x) \geq 0
\]
Theorem

The following implication holds

\[ f(x) \geq 0, \ x \neq 0 \Rightarrow f_0(x) \geq 0 \]

if there exists

\[ \tau \geq 0 \text{ such that } f_0(x) - \tau f(x) \geq 0 \]

Still not an LMI ... but \( f_0(x), f(x), \) are quadratic in \( x. \)
Theorem

The following implication holds

\[ f(x) \geq 0, \ x \neq 0 \ \Rightarrow \ f_0(x) \geq 0 \]

if there exists

\[ \tau \geq 0 \text{ such that } f_0(x) - \tau f(x) \geq 0 \]

For a quadratic function \( f(x) = x^\top Ax + 2x^\top b + c \)

\[
\begin{bmatrix} x \end{bmatrix}^\top \begin{bmatrix} A & b \\ b^\top & c \end{bmatrix} \begin{bmatrix} x \end{bmatrix} \geq 0, \ \forall x \iff \begin{bmatrix} \xi x \end{bmatrix}^\top \begin{bmatrix} A & b \\ b^\top & c \end{bmatrix} \begin{bmatrix} \xi x \end{bmatrix} \geq 0, \ \forall x, \xi
\]

\[ \iff \begin{bmatrix} A \\ b^\top \\ c \end{bmatrix} \geq 0 \]
Theorem

The following implication holds

\[ f(x) \geq 0, \ x \neq 0 \Rightarrow f_0(x) \geq 0 \]

if there exists

\[ \tau \geq 0 \text{ such that } f_0(x) - \tau f(x) \geq 0 \]

Since \( f_0(x), f(x), \) are quadratic in \( x \), the condition above is equivalent to an LMI in \( \tau \)

\[
\begin{bmatrix}
A_0 & b_0 \\
b_0^T & c_0
\end{bmatrix} - \tau
\begin{bmatrix}
A & b \\
b^T & c
\end{bmatrix} \succeq 0
\]
Theorem

The following implication holds

\[ f(x) \geq 0, \ x \neq 0 \implies f_0(x) \geq 0 \]

if there exists \( \tau \geq 0 \) such that

\[ f_0(x) - \tau f(x) \geq 0 \]

Since \( f_0(x), f(x) \), are quadratic in \( x \), this is equivalent to an LMI in \( \tau \)

\[
\begin{bmatrix}
  A_0 & b_0 \\
  b_0^T & c_0
\end{bmatrix} - \tau
\begin{bmatrix}
  A & b \\
  b^T & c
\end{bmatrix} \succeq 0
\]

The only if part also holds true (though non-obvious) if \( \exists \bar{x} \) such that \( f(\bar{x}) > 0 \), i.e. the “ellipsoids” have non-empty interior condition. In that case we get equivalence!
Stability analysis – Linear systems

Consider the linear, time-invariant (LTI) dynamical system

\[ \dot{x}(t) = Ax(t), \quad x(0) = x_0 \]

where \( x(t) \in \mathbb{R}^n \) is the system state and \( A \in \mathbb{R}^{n\times n} \). It is called *autonomous* since there are no inputs.

**Definition:** The autonomous LTI system is *asymptotically stable* if, for all \( x(0) \in \mathbb{R}^n \),

\[ \lim_{t \to \infty} x(t) = 0. \]

In the scalar case (\( n = 1 \) and \( A = a \in \mathbb{R} \)), we can solve the ODE:

\[ x(t) = e^{at} x_0 \]

If \( a < 0 \), then the system is asymptotically stable.
Stability analysis recap – Linear systems

Consider the linear, time-invariant (LTI) dynamical system

\[ \dot{x}(t) = Ax(t), \quad x(0) = x_0 \]

where \( x(t) \in \mathbb{R}^n \) is the system state and \( A \in \mathbb{R}^{n \times n} \). It is called \textit{autonomous} since there are no inputs.

**Definition:** The autonomous LTI system is \textit{asymptotically stable} if, for all \( x(0) \in \mathbb{R}^n \),

\[ \lim_{t \to \infty} x(t) = 0. \]

What if \( n > 1 \)? Can we work the same way? The ODE solution is then

\[ x(t) = e^{At}x_0, \]

where \( e^{At} \) is the matrix exponential, i.e.

\[ e^{At} = I + At + \frac{1}{2}A^2t^2 + \frac{1}{3}A^3t^3 + \ldots \]
Consider the linear, time-invariant (LTI) dynamical system

\[ \dot{x}(t) = Ax(t), \quad x(0) = x_0 \]

where \( x(t) \in \mathbb{R}^n \) is the system state and \( A \in \mathbb{R}^{n \times n} \). It is called *autonomous* since there are no inputs.

**Definition:** The autonomous LTI system is *asymptotically stable* if, for all \( x(0) \in \mathbb{R}^n \),

\[
\lim_{t \to \infty} x(t) = 0.
\]

What if \( n > 1 \)? Can we work the same way? The ODE solution is then

\[ x(t) = e^{At}x_0, \]

where \( e^{At} \) is the matrix exponential. Can we do without computing \( e^{At} \)?
Stability analysis recap – Linear systems

**Theorem**

An autonomous LTI system is asymptotically stable, i.e. \( \lim_{t \to \infty} x(t) = 0 \), if and only if \( A \) is Hurwitz, i.e. all its eigenvalues have negative real part.

Moved from matrix exponential to eigenvalue computation – there must be some connection with LMIs.

**Theorem**

*Given some matrix \( Q = Q^\top \succ 0 \), a matrix \( A \) is Hurwitz if and only if there exists \( X = X^\top \succ 0 \) that satisfies the Lyapunov Matrix Equation*

\[
A^\top X +XA = -Q
\]

Equivalently, since \( Q \succ 0 \) and it is arbitrary ...
Stability analysis recap – Linear systems

For asymptotic stability \( A \) has to be Hurwitz, i.e.

**Theorem**

Given some matrix \( Q = Q^\top \succ 0 \), a matrix \( A \) is Hurwitz **if and only if** there exists \( X = X^\top \succ 0 \) that satisfies the Lyapunov Matrix Equation

\[
A^\top X + XA = -Q
\]

Equivalently, since \( Q \succ 0 \) and it is arbitrary ...

**Theorem**

A matrix \( A \) is Hurwitz **if and only if** there exists \( X = X^\top \succ 0 \) that satisfies the Lyapunov Matrix Inequality

\[
A^\top X + XA \prec 0
\]

This is an LMI in \( X \)!
Summary

1. Reformulation in LMI constraints
   - Schur complement
     - Commonly used “trick”
     - Appears in quadratic problems, and many others
   - The S-procedure
     - Turns quadratic implications in LMI constraints
     - Useful in set containment problems

2. LMIs for stability & controller synthesis
Thank you! Questions?

Contact at:

kostas.margellos@eng.ox.ac.uk