2. Convexity and Duality

- Formulation of optimization problems
- Engineering examples
- Convex sets and functions
- Convex optimization problems
- Standard problems: LP and SDP
- Feasibility problems
- Algorithms
- Certificates and separating hyperplanes
- Duality and geometry
- Examples: LP and and SDP
- Theorems of alternatives
Optimization Problems

A familiar problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0 & \text{for all } i = 1, \ldots, m \\
& \quad h_i(x) = 0 & \text{for all } i = 1, \ldots, p
\end{align*}
\]

- \( x \in \mathbb{R}^n \) is the variable
- \( f_0 : \mathbb{R}^n \rightarrow \mathbb{R} \) is the objective function
- \( f_i : \mathbb{R}^n \rightarrow \mathbb{R} \) for \( i = 1, \ldots, m \) define inequality constraints
- \( h_i : \mathbb{R}^n \rightarrow \mathbb{R} \) for \( i = 1, \ldots, p \) define equality constraints
Lyapunov functions

\[
\begin{bmatrix}
\dot{x}(t) \\
\dot{y}(t)
\end{bmatrix} =
\begin{bmatrix}
-0.7 & 0.2 \\
-0.6 & -0.1
\end{bmatrix}
\begin{bmatrix}
x(t) \\
y(t)
\end{bmatrix}
\]

\[
V(x) =
\begin{bmatrix}
x \\
y
\end{bmatrix}^T
\begin{bmatrix}
1.58 & -1.28 \\
-1.28 & 2.4
\end{bmatrix}
\begin{bmatrix}
x \\
y
\end{bmatrix}
\]

- Ellipsoids are the sublevel sets of \( V \)

\[
E = \left\{ x \in \mathbb{R}^2 \mid V(x) < c \right\}
\]

These are invariant sets of the system.

- \( V(x(t)) \) is a decreasing function along every trajectory.
nonlinear systems

Suppose we have the system of ordinary differential equations

\[ \dot{x}(t) = f(x) \]

with \( x(t) \in \mathbb{R}^n \) and \( f(0) = 0 \). (The origin is an equilibrium point.)

**global asymptotic stability**

If there is a function \( V : \mathbb{R}^n \to \mathbb{R} \) (called a Lyapunov function) which satisfies

(i) \( V \) is differentiable, and its gradient is continuous.

(ii) \( V(x) > 0 \) for \( x \neq 0 \), and \( V(0) = 0 \).

(iii) \( \frac{dV}{dt} = \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} f_i(x) < 0 \) for \( x \neq 0 \).

(iv) If \( \{x_0, x_1, \ldots \} \) is a sequence such that \( \|x_k\| \to \infty \), then \( V(x_k) \to \infty \).

then for any initial condition \( x(0) \)

\[
\lim_{t \to \infty} x(t) = 0
\]
example: Lyapunov functions for a nonlinear system

A model for a jet engine (from Moore-Greitzer), with controller

\[
\begin{align*}
\dot{x} &= -y - \frac{3}{2}x^2 - \frac{1}{2}x^3 \\
\dot{y} &= 3x - y
\end{align*}
\]

\[V = 4.5819x^2 - 1.5786xy + 1.7834y^2 - 0.12739x^3 + 2.5189x^2y - 0.34069xy^2 + 0.61188y^3 + 0.47537x^4 - 0.052424x^3y + 0.44289x^2y^2 + 0.0000018868xy^3 + 0.090723y^4\]
computation of lyapunov functions

- the above Lyapunov function was found automatically
- many extensions possible
  - switching or hybrid systems
  - finite-state automata; computation of invariants
  - uncertain systems
  - bounds on cost via solution of Hamilton-Jacobi inequalities
Graph problems

Graph problems appear in many areas: MAX-CUT, independent set, cliques, etc.

MAX CUT partitioning

- Partition the nodes of a graph in two disjoint sets, maximizing the number of edges between sets.
- Practical applications (circuit layout, etc.)
- NP-complete.

How to compute bounds, or exact solutions, for this kind of problems?
discrete problems: LQR with binary inputs

- linear discrete-time system \( x(t + 1) = Ax(t) + Bu(t) \) on interval \( t = 0, \ldots, N \)

- objective is to minimize the quadratic tracking error

\[
\sum_{t=0}^{N-1} (x(t) - r(t))^T Q (x(t) - r(t))
\]

- using binary inputs

\( u_i(t) \in \{-1, 1\} \) for all \( i = 1, \ldots, m \), and \( t = 0, \ldots, N - 1 \)
Basic Nomenclature

A set $S \subset \mathbb{R}^n$ is called

- **affine** if $x, y \in S$ implies $\theta x + (1 - \theta)y \in S$ for all $\theta \in \mathbb{R}$; i.e., the line through $x, y$ is contained in $S$.

- **convex** if $x, y \in S$ implies $\theta x + (1 - \theta)y \in S$ for all $\theta \in [0, 1]$; i.e., the line segment between $x$ and $y$ is contained in $S$.

- **a convex cone** if $x, y \in S$ implies $\lambda x + \mu y \in S$ for all $\lambda, \mu \geq 0$; i.e., the pie slice between $x$ and $y$ is contained in $S$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called

- **affine** if $f(\theta x + (1 - \theta)y) = \theta f(x) + (1 - \theta)f(y)$ for all $\theta \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$; i.e., $f$ is equals a linear function plus a constant $f = Ax + b$.

- **convex** if $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$ for all $\theta \in [0, 1]$ and $x, y \in \mathbb{R}^n$. 
Examples of Convex Functions

- $f(x) = c$
- $f(x) = \lambda^T x + c$
- if $Q \succeq 0$ then $f(x) = x^T Q x$ is convex
- on $\mathbb{R}$, the exponential $f(x) = e^x$ is convex
- any norm $f(x) = \|x\|$ is convex
Properties of Convex Functions

- $f_1 + f_2$ is convex if $f_1$ and $f_2$ are
- $f(x) = \max \{ f_1(x), f_2(x) \}$ is convex if $f_1$ and $f_2$ are
- $g(x) = \sup_y f(x, y)$ is convex if $f(x, y)$ is convex in $x$ for each $y$
- Convex functions are continuous on the interior of their domain
- $f(Ax + b)$ is convex if $f$ is
- $Af(x) + b$ is convex if $f$ is
- $g(x) = \inf_y f(x, y)$ is convex if $f(x, y)$ is jointly convex
- The $\alpha$–sublevel set
  \[
  \{ x \in \mathbb{R}^n \mid f(x) \leq \alpha \}
  \]
  is convex if $f$ is convex; (the converse is not true)
Convex Optimization Problems

minimize \quad f_0(x)

subject to \quad f_i(x) \leq 0 \quad \text{for all } i = 1, \ldots, m
\quad h_i(x) = 0 \quad \text{for all } i = 1, \ldots, p

This problem is called a \textit{convex program} if

- the objective function $f_0$ is convex
- the inequality constraints $f_i$ are convex
- the equality constraints $h_i$ are affine
Linear Programming (LP)

In a *linear program*, the objective and constraint functions are affine.

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad Cx \leq d
\end{align*}
\]

**Example**

\[
\begin{align*}
\text{minimize} & \quad x_1 + x_2 \\
\text{subject to} & \quad 3x_1 + x_2 \geq 3 \\
& \quad x_2 \geq 1 \\
& \quad x_1 \leq 4 \\
& \quad -x_1 + 5x_2 \leq 20 \\
& \quad x_1 + 4x_2 \leq 20
\end{align*}
\]
Linear Programming

Every linear program may be written in the *standard primal form*

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

Here \( x \in \mathbb{R}^n \), and \( x \geq 0 \) means \( x_i \geq 0 \) for all \( i \)

- The *nonnegative orthant* \( \{ x \in \mathbb{R}^n \mid x \geq 0 \} \) is a *convex cone*.
- This convex cone defines the partial ordering \( \geq \) on \( \mathbb{R}^n \)
- Geometrically, the feasible set is the intersection of an affine set with a convex cone.
- Any closed convex cone \( K \) with nonempty interior defines a partial ordering; the dual cone is

\[
K^* = \{ y \mid x^T y \geq 0 \text{ for all } x \in K \}
\]
Semidefinite Programming

minimize $\text{trace } CX$
subject to $\text{trace } A_iX = b_i$ for all $i = 1, \ldots, m$
$X \succeq 0$

- The variable $X$ is in the set of $n \times n$ symmetric matrices
  $$\mathcal{S}^n = \{ A \in \mathbb{R}^{n \times n} \mid A = A^T \}$$
- $X \succeq 0$ means $X$ is positive semidefinite
- As for LP, the feasible set is the intersection of an affine set with a convex cone, in this case the positive semidefinite cone
  $$\{ X \in \mathcal{S}^n \mid X \succeq 0 \}$$
Hence the feasible set is convex.
SDPs with Explicit Variables

We can also explicitly parametrize the affine set to give

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad F_0 + x_1 F_1 + x_2 F_2 + \cdots + x_n F_n \preceq 0
\end{align*}
\]

where \( F_0, F_1, \ldots, F_n \) are symmetric matrices.

The inequality constraint is called a \textit{linear matrix inequality}; e.g.,

\[
\begin{bmatrix}
x_1 - 3 & x_1 + x_2 & -1 \\
x_1 + x_2 & x_2 - 4 & 0 \\
-1 & 0 & x_1
\end{bmatrix} \preceq 0
\]

which is equivalent to

\[
\begin{bmatrix}
-3 & 0 & -1 \\
0 & -4 & 0 \\
-1 & 0 & 0
\end{bmatrix} + x_1 \begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix} + x_2 \begin{bmatrix}
0 & 1 & 0 \\
1 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix} \preceq 0
\]
The Feasible Set is Semialgebraic

The *(basic closed) semialgebraic set* defined by polynomials $f_1, \ldots, f_m$ is

$$\left\{ x \in \mathbb{R}^n \mid f_i(x) \geq 0 \text{ for all } i = 1, \ldots, m \right\}$$

The feasible set of an SDP is a semialgebraic set.

Because a matrix $A \succeq 0$ if and only if

$$\det(A_k) > 0 \text{ for } k = 1, \ldots, n$$

where $A_k$ is the submatrix of $A$ consisting of the first $k$ rows and columns.
The Feasible Set

For example

\[
0 < \begin{bmatrix}
3 - x_1 & -(x_1 + x_2) & 1 \\
-(x_1 + x_2) & 4 - x_2 & 0 \\
1 & 0 & -x_1
\end{bmatrix}
\]

is equivalent to the polynomial inequalities

\[
0 < 3 - x_1 \\
0 < (3 - x_1)(4 - x_2) - (x_1 + x_2)^2 \\
0 < -x_1((3 - x_1)(4 - x_2) - (x_1 + x_2)^2) - (4 - x_2)
\]
Intersection of Feasible Sets

The intersection of the feasible sets

\[
\begin{bmatrix}
2x_1 + x_2 + 2 & 0 \\
0 & -x_1 - 5
\end{bmatrix} < 0
\]

and

\[
\begin{bmatrix}
x_1 - 3 & x_1 + x_2 & -1 \\
x_1 + x_2 & x_2 - 4 & 0 \\
-1 & 0 & x_1
\end{bmatrix} < 0
\]

is given by

\[
\begin{bmatrix}
x_1 - 3 & x_1 + x_2 & -1 & 0 & 0 & 0 \\
x_1 + x_2 & x_2 - 4 & 0 & 0 & 0 & 0 \\
-1 & 0 & x_1 & 0 & 0 & 0 \\
0 & 0 & 0 & 2x_1 + x_2 + 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -x_1 - 5
\end{bmatrix} < 0
\]
Optimal Points

Since SDPs are convex, if the feasible set is closed then the optimal is always achieved on the boundary.
LMIs with Matrix Variables

The inequality
\[
\begin{bmatrix}
A^T Y + Y^T A & Y^T B \\
B^T Y & -I
\end{bmatrix} < 0
\]
is a linear matrix inequality in the variable \(Y\). Here
\[
A \in \mathbb{R}^{n \times n} \quad B \in \mathbb{R}^{n \times m} \quad Y \in \mathbb{R}^{n \times n}
\]
This form is accepted by some software, e.g. YALMIP.

To convert this to standard form, write
\[
Y(x) = \begin{bmatrix}
x_1 & x_{n+1} & \cdots \\
x_2 & \vdots & \ddots \\
x_n & x_{2n} & \cdots & x_{n^2}
\end{bmatrix}
\]
equivalently \(x = \text{vec}(Y)\)

Then
\[
\begin{bmatrix}
A^T Y(x) + Y^T(x)A & Y^T(x)B \\
B^T Y(x) & -I
\end{bmatrix}
is affine in \(x\).
Convex Optimization Problems

For a convex optimization problem, the \textit{feasible set}

\[ S = \{ x \in \mathbb{R}^n \mid f_i(x) \leq 0 \text{ and } h_j(x) = 0 \text{ for all } i, j \} \]

is convex. So we can write the problem as

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad x \in S
\end{align*}
\]

This approach emphasizes the \textit{geometry} of the problem.

For a convex optimization problem, any local minimum is also a global minimum.
Feasibility Problems

We are also interested in \textit{feasibility problems} as follows. Does there exist $x \in \mathbb{R}^n$ which satisfies

$$f_i(x) \leq 0 \quad \text{for all } i = 1, \ldots, m$$
$$h_i(x) = 0 \quad \text{for all } i = 1, \ldots, p$$

If there does not exist such an $x$, the problem is described as \textit{infeasible}. 
Feasibility Problems

We can always convert an optimization problem into a feasibility problem; does there exist $x \in \mathbb{R}^n$ such that

\[
\begin{align*}
    f_0(x) &\leq t \\
    f_i(x) &\leq 0 \\
    h_i(x) &= 0
\end{align*}
\]

Bisection search over the parameter $t$ finds the optimal.
Feasibility Problems

Conversely, we can convert feasibility problems into optimization problems.

e.g. the feasibility problem of finding $x$ such that

$$f_i(x) \leq 0 \quad \text{for all } i = 1, \ldots, m$$

can be solved as

$$\begin{align*}
\text{minimize} & \quad y \\
\text{subject to} & \quad f_i(x) \leq y \quad \text{for all } i = 1, \ldots, m
\end{align*}$$

where there are $n + 1$ variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}$

This technique may be used to find an initial feasible point for optimization algorithms
Algorithms

For convex optimization problems, there are several good algorithms

- interior-point algorithms work well in theory and practice
- for certain classes of problems, (e.g. LP and SDP) there is a worst-case time-complexity bound
- conversely, some convex optimization problems are known to be NP-hard
- problems are specified either in standard form, for LPs and SDPs, or via an oracle

Some Matlab Software

- SeDuMi http://fewcal.kub.nl/sturm/software/sedumi.html
- YALMIP http://www.control.isy.liu.se/~johanl/yalmip.html
Certificates

Consider the feasibility problem

\[
\begin{align*}
\text{Does there exist } x \in \mathbb{R}^n \text{ which satisfies} \\
& f_i(x) \leq 0 \text{ for all } i = 1, \ldots, m \\
& h_i(x) = 0 \text{ for all } i = 1, \ldots, p
\end{align*}
\]

There is a fundamental asymmetry between establishing that

- There exists at least one feasible \( x \)
- The problem is infeasible

To show existence, one needs a \textit{feasible point} \( x \in \mathbb{R}^n \).
To show emptiness, one needs a \textit{certificate of infeasibility}; a mathematical proof that the problem is infeasible.
Certificates and Separating Hyperplanes

The simplest form of certificate is a separating hyperplane. The idea is that a hyperplane $L \subset \mathbb{R}^n$ breaks $\mathbb{R}^n$ into two half-spaces,

$$H_1 = \left\{ x \in \mathbb{R}^n \mid b^T x \leq a \right\} \quad \text{and} \quad H_2 = \left\{ x \in \mathbb{R}^n \mid b^T x > a \right\}$$

If two closed, bounded and convex sets are disjoint, there is a hyperplane that separates them.

So to prove infeasibility of

$$f_i(x) \leq 0 \quad \text{for } i = 1, 2$$

we show that

$$\left\{ x \in \mathbb{R}^n \mid f_1(x) \leq 0 \right\} \subset H_1 \quad \text{and} \quad \left\{ x \in \mathbb{R}^n \mid f_2(x) \leq 0 \right\} \subset H_2$$
Duality

We’d like to solve

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0 \quad \text{for all } i = 1, \ldots, m \\
& \quad h_i(x) = 0 \quad \text{for all } i = 1, \ldots, p
\end{align*}
\]

define the Lagrangian for \( x \in \mathbb{R}^n \), \( \lambda \in \mathbb{R}^m \) and \( \nu \in \mathbb{R}^p \) by

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

and the Lagrange dual function

\[
g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)
\]

We allow \( g(\lambda, \nu) = -\infty \) when there is no finite infimum
Duality

the dual problem is

maximize \( g(\lambda, \nu) \)
subject to \( \lambda \geq 0 \)

we call \( \lambda, \nu \) dual feasible if \( \lambda \geq 0 \) and \( g(\lambda, \nu) \) is finite.

- The dual function \( g \) is always concave, even if the primal problem is not convex
Weak Duality

For any primal feasible $x$ and dual feasible $\lambda, \nu$ we have

$$g(\lambda, \nu) \leq f_0(x)$$

because

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

$$\leq f_0(x)$$

- A feasible $\lambda, \nu$ provides a certificate that the primal optimal is greater than $g(\lambda, \nu)$
- many interior-point methods simultaneously optimize the primal and the dual problem; when $f_0(x) - g(\lambda, \nu) \leq \varepsilon$ we know that $x$ is $\varepsilon$—suboptimal
**Strong Duality**

- $p^*$ is the optimal value of the primal problem,
- $d^*$ is the optimal value of the dual problem

Weak duality means $p^* \geq d^*$

If $p^* = d^*$ we say *strong duality* holds. Equivalently, we say the *duality gap* $p^* - d^*$ is zero.

*Constraint qualifications* give sufficient conditions for strong duality.

An example is *Slater’s condition*; strong duality holds if the primal problem is convex and strictly feasible.
Geometric Interpretations: The Lagrangian

consider the optimization problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_1(x) \leq 0
\end{align*}
\]

The value of the Lagrangian \( L(x, \lambda) \) is the intersection of the hyperplane \( H_\lambda \) with the vertical axis

\[
H_\lambda = \{ (z, y) \mid \lambda^Tz + y = L(x, \lambda) \}
\]
The Lagrange Dual Function

The Lagrange dual function is

\[ g(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda) \]

i.e., the minimum intersection for a given slope \(-\lambda\)
Sensitivity

consider the perturbed problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq y_i \quad \text{for all } i = 1, \ldots, m
\end{align*}
\]

and let \( p^*(y) \) be the optimal value parametrized by \( y \). Then for any optimal \( \lambda^* \) we have

\[
\lambda^* = -\nabla p^*(0)
\]
Complementary Slackness

For \( \lambda^* \) dual optimal, and \( x^* \) primal optimal, we have

\[
\lambda_i^* f_i(x^*) = 0 \quad \text{for all } i = 1, \ldots, m
\]

whenever strong duality holds; i.e., if the \( i \)'th constraint is active, then \( \lambda_i^* = 0 \)

\[
\{ (f_1(x), f_0(x)) \mid x \in R \}
\]

\[
H_\lambda = \{ (z, y) \mid \lambda^T z + y = L(x, \lambda) \}
\]
Example: Linear Programming

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

The Lagrange dual function is

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

\[
= \begin{cases} 
  b^T \nu & \text{if } c - A^T \nu - \lambda = 0 \\
  -\infty & \text{otherwise}
\end{cases}
\]

So the dual problem is

\[
\begin{align*}
\text{maximize} & \quad b^T \nu \\
\text{subject to} & \quad A^T \nu \leq c
\end{align*}
\]
Example: Semidefinite Programming

minimize \( \text{trace } CX \)
subject to \( \text{trace } A_i X = b_i \) for all \( i = 1, \ldots, m \)
\( X \succeq 0 \)

The Lagrange dual is

\[
g(Z, \nu) = \inf_X \left( \text{trace } CX - \text{trace } ZX + \sum_{i=1}^{m} \nu_i (b_i - \text{trace } A_i X) \right)
\]

\[
= \begin{cases}
  b^T \nu & \text{if } C - Z - \sum_{i=1}^{m} \nu_i A_i = 0 \\
  -\infty & \text{otherwise}
\end{cases}
\]

So the dual problem is to maximize \( b^T \nu \) subject to

\[
C - Z - \sum_{i=1}^{m} \nu_i A_i = 0 \quad \text{and} \quad Z \succeq 0
\]
Semidefinite Programming Duality

The primal problem is

\[
\begin{align*}
\text{minimize} & \quad \text{trace } CX \\
\text{subject to} & \quad \text{trace } A_i X = b_i \quad \text{for all } i = 1, \ldots, m \\
& \quad X \succeq 0
\end{align*}
\]

The dual problem is

\[
\begin{align*}
\text{maximize} & \quad b^T \nu \\
\text{subject to} & \quad \sum_{i=1}^{m} \nu_i A_i \preceq C
\end{align*}
\]
The Fourfold Way

There are several ways of formulating an SDP for its numerical solution.

Because *subspaces* can be described

- Using *generators* or a *basis*; Equivalently, the subspace is the range of a linear map
\[
\{ x \mid x = B \lambda \text{ for some } \lambda \} \]

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix} = \lambda_1 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_1 - \lambda_2 \\ 2\lambda_1 + 2\lambda_2 \end{bmatrix}
\]

- Through the defining equations; i.e, as the *kernel* \( \{ x \mid Ax = 0 \} \)

\[
\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid 4x_1 - 2x_2 - x_3 = 0\}
\]

Depending on which description we use, and whether we write a primal or dual formulation, we have *four* possibilities (two primal-dual pairs).
Example: Two Primal-Dual Pairs

maximize \( 2x + 2y \)
subject to \[
\begin{bmatrix}
1 + x & y \\
y & 1 - x
\end{bmatrix} \succeq 0
\]

minimize \( \text{trace} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} W \)
subject to \[
\text{trace} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} W = 2
\]
\[
\text{trace} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} W = 2
\]
\( W \succeq 0 \)

Another, more efficient formulation which solves the same problem:

maximize \( \text{trace} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} Z \)
subject to \[
\text{trace} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} Z = 2
\]
\( Z \succeq 0 \)

minimize \( 2t \)
subject to \[
\begin{bmatrix} t - 1 & -1 \\ -1 & t + 1 \end{bmatrix} \succeq 0
\]
Duality

- Duality has many interpretations; via economics, game-theory, geometry.
- e.g., one may interpret Lagrange multipliers as a price for violating constraints, which may correspond to resource limits or capacity constraints.
- Often physical problems associate specific meaning to certain Lagrange multipliers, e.g. pressure, momentum, force can all be viewed as Lagrange multipliers.
Example: Mechanics

- Spring under compression
- Mass at horizontal position $x$, equilibrium at $x = \frac{1}{2}$

minimize \[ \frac{k}{2}(x - 2)^2 \]
subject to \[ x \leq 1 \]

The Lagrangian is \[ L(x, \lambda) = \frac{k}{2}(x - 2)^2 + \lambda(x - 1) \]

If $\lambda$ is dual optimal and $x$ is primal optimal, then \[ \frac{\partial}{\partial x}L(x, \lambda) = 0, \text{ i.e.,} \]

\[ k(x - 2) + \lambda = 0 \]

so we can interpret $\lambda$ as a force
Feasibility of Inequalities

The *primal feasibility problem* is

\[
\text{does there exist } x \in \mathbb{R}^n \text{ such that } f_i(x) \geq 0 \quad \text{for all } i = 1, \ldots, m
\]

The *dual function* \( g : \mathbb{R}^m \rightarrow \mathbb{R} \) is

\[
g(\lambda) = \sup_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \lambda_i f_i(x)
\]

The *dual feasibility problem* is

\[
\text{does there exist } \lambda \in \mathbb{R}^m \text{ such that } g(\lambda) < 0 \\
\lambda \geq 0
\]
Theorem of Alternatives

If the dual problem is feasible, then the primal problem is infeasible.

Proof

Suppose the primal problem is feasible, and let \( \tilde{x} \) be a feasible point. Then

\[
g(\lambda) = \sup_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \lambda_i f_i(x)
\]

\[
\geq \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) \quad \text{for all } \lambda \in \mathbb{R}^m
\]

and so \( g(\lambda) \geq 0 \) for all \( \lambda \geq 0 \).
Geometric Interpretation

$S = \left\{ z \in \mathbb{R}^m \mid z \geq 0 \right\}$

$H_\lambda = \left\{ z \in \mathbb{R}^m \mid \lambda^T z = g(\lambda) \right\}$

If $g(\lambda) < 0$ and $\lambda \geq 0$ then the hyperplane $H_\lambda$ separates $S$ from $T$, where

$$T = \left\{ \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \mid x \in \mathbb{R}^n \right\}$$
Certificates

- A dual feasible point gives a *certificate* of infeasibility of the primal problem.
- If the Lagrange dual function $g$ is easy to compute, and we can show $g(\lambda) < 0$, then this is a *proof* that the primal is infeasible.
- One way to do this is to have an explicit expression for

$$g(\lambda) = \sup_x L(x, \lambda)$$

where for feasibility problems, the Lagrangian is $L(x, \lambda) = \sum_{i=1}^{m} \lambda_i f_i(x)$

- Alternatively, given $\lambda$, we may be able to show directly that

$$L(x, \lambda) < -\varepsilon \quad \text{for all } x \in \mathbb{R}^n$$

for some $\varepsilon > 0$. 