Basic convex problems
Linear Programming

\[ \min \quad c^T x \]
\[ \text{s.t.} \quad Ax \leq b, \quad Cx = d. \]
Linear Programming

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad Ax \leq b, \quad Cx = d.
\end{align*}
\]

Piecewise linear minimization is an LP

\[
\min f(x) = \max_{1 \leq i \leq m} (a_i^T x + b_i)
\]
Linear Programming

\[
\min \ c^T x \\
\text{s.t. } Ax \leq b, \quad Cx = d.
\]

**Piecewise linear minimization is an LP**

\[
\min \ f(x) = \max_{1 \leq i \leq m} (a_i^T x + b_i)
\]

\[
\begin{align*}
\min_{x,t} & \quad t \\
\text{s.t. } & \quad a_i^T x + b_i \leq t, \quad i = 1, \ldots, m.
\end{align*}
\]
Exercises

Formulate $\min_x \|Ax - b\|_1$ as an LP ($\|x\|_1 = \sum_i |x_i|$)

Formulate $\min_x \|Ax - b\|_\infty$ as an LP ($\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$)
Exercises

Formulate $\min_x \|Ax - b\|_1$ as an LP ($\|x\|_1 = \sum_i |x_i|$)

Formulate $\min_x \|Ax - b\|_\infty$ as an LP ($\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$)

Explore: LP formulations for Markov Decision Processes (MDPs). MDPs are frequently used models in Reinforcement Learning, and in some cases admit nice LP formulations.
Exercises

Formulate $\min_x \|Ax - b\|_1$ as an LP ($\|x\|_1 = \sum_i |x_i|$)

Formulate $\min_x \|Ax - b\|_\infty$ as an LP ($\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$)

Explore: LP formulations for Markov Decision Processes (MDPs). MDPs are frequently used models in Reinforcement Learning, and in some cases admit nice LP formulations.

Explore: Integer LP: $\min_x c^T x, Ax \leq b, x \in \mathbb{Z}^n$. 

N. Meggido, On the complexity of linear programing: Click here!
Exercises

Formulate $\min_x \|Ax - b\|_1$ as an LP ($\|x\|_1 = \sum_i |x_i|$)

Formulate $\min_x \|Ax - b\|_\infty$ as an LP ($\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$)

Explore: LP formulations for Markov Decision Processes (MDPs). MDPs are frequently used models in Reinforcement Learning, and in some cases admit nice LP formulations.

Explore: Integer LP: $\min_x c^T x, Ax \leq b, x \in \mathbb{Z}^n$.

Open Problem. Can we solve the system of inequalities $Ax \leq b$ in strongly polynomial time in the dimensions of the system, independent of the magnitudes of the coefficients? Best known result (Tardos, 1984) depends on coefficients of $A$ but permits independence on magnitudes of $b$ and the cost vector $c$.

N. Meggido, *On the complexity of linear programing*: Click here!
Quadratic Programming

\[
\min \quad \frac{1}{2} x^T A x + b^T x + c \\
\text{s.t.} \quad Gx \leq h.
\]

We assume \( A \succeq 0 \) (semidefinite).

Exercise: Suppose no constraints; does QP always have solutions?

Nonnegative least squares (NNLS)

\[
\min \quad \|Ax - b\|_2 \\
\text{s.t.} \quad x \geq 0
\]

Exercise: Prove that NNLS always has a solution.
Quadratic Programming

\[
\min \; \frac{1}{2} x^T Ax + b^T x + c \quad \text{s.t.} \quad Gx \leq h.
\]

We assume \( A \succeq 0 \) (semidefinite).

**Exercise:** Suppose no constraints; does QP always have solutions?
Quadratic Programming

\[ \min \quad \frac{1}{2} x^T Ax + b^T x + c \quad \text{s.t.} \quad Gx \leq h. \]

We assume \( A \succeq 0 \) (semidefinite).

**Exercise:** Suppose no constraints; does QP always have solutions?

**Nonnegative least squares (NNLS)**

\[ \min \quad \frac{1}{2} \| Ax - b \|^2 \quad \text{s.t.} \quad x \geq 0. \]

**Exercise:** Prove that NNLS always has a solution.
Regularized least-squares

Lasso

\[
\min \quad \frac{1}{2} \| Ax - b \|_2^2 + \lambda \| x \|_1.
\]

**Exercise:** How large must \( \lambda > 0 \) so that \( x = 0 \) is the optimum?
Regularized least-squares

**Lasso**

\[ \min \quad \frac{1}{2} \| Ax - b \|_2^2 + \lambda \| x \|_1. \]

**Exercise:** How large must \( \lambda > 0 \) so that \( x = 0 \) is the optimum?

**Total-variation denoising**

\[ \min \quad \frac{1}{2} \| Ax - b \|_2^2 + \lambda \sum_{i=1}^{n-1} |x_{i+1} - x_i|. \]

**Exercise:** Is the total-variation term a norm? Prove or disprove.
Regularized least-squares

**Lasso**

\[
\min \quad \frac{1}{2} \| Ax - b \|_2^2 + \lambda \| x \|_1.
\]

**Exercise:** How large must \( \lambda > 0 \) so that \( x = 0 \) is the optimum?

**Total-variation denoising**

\[
\min \quad \frac{1}{2} \| Ax - b \|_2^2 + \lambda \sum_{i=1}^{n-1} |x_{i+1} - x_i|.
\]

**Exercise:** Is the total-variation term a norm? Prove or disprove.

**Group Lasso**

\[
\min_{x_1, \ldots, x_T} \quad \frac{1}{2} \left\| b - \sum_{j=1}^{T} A_j x_j \right\|_2^2 + \lambda \sum_{j=1}^{T} \| x_j \|_2.
\]

**Exercise:** What is the dual norm of the regularizer above?
Robust LP as an SOCP

\[
\begin{align*}
\min & \quad c^T x, \\
\text{s.t.} & \quad a_i^T x \leq b_i \quad \forall a_i \in \mathcal{E}_i \\
\mathcal{E}_i & := \{ \bar{a}_i + P_i u \mid \|u\|_2 \leq 1 \}
\end{align*}
\]

Constraints are uncertain but with bounded uncertainty.
Robust LP as an SOCP

\[
\begin{align*}
\min \quad & c^T x, \\
\text{s.t.} \quad & a_i^T x \leq b_i \quad \forall a_i \in \mathcal{E}_i \\
\mathcal{E}_i := & \{ \bar{a}_i + P_i u \mid \|u\|_2 \leq 1 \}
\end{align*}
\]

Constraints are uncertain but with bounded uncertainty.

(Adversarially) Robust LP formulation

\[
\min_x \max_{\|u\|_2 \leq 1} \left\{ c^T x \mid a_i^T x \leq b_i, \quad a_i \in \mathcal{E}_i \right\}
\]
Robust LP as an SOCP

\[
\min c^T x, \quad \text{s.t.} \quad a_i^T x \leq b_i \quad \forall a_i \in \mathcal{E}_i
\]

\[
\mathcal{E}_i := \{\bar{a}_i + P_i u \mid \|u\|_2 \leq 1\}
\]

Constraints are **uncertain** but with bounded uncertainty.

**(Adversarially) Robust LP formulation**

\[
\min_x \max_{\|u\|_2 \leq 1} \left\{c^T x \mid a_i^T x \leq b_i, \quad a_i \in \mathcal{E}_i\right\}
\]

**Second Order Cone Program**

\[
\min \quad c^T x, \quad \text{s.t.} \quad \|P_i^T x\|_2 \leq -\bar{a}_i^T x + b_i, \quad i = 1, \ldots, m.
\]

SOCP constraint comes from:

\[
\max_{\|u\|_2 \leq 1} (\bar{a}_i + P_i u)^T x = \bar{a}_i^T x + \|P_i^T x\|_2
\]

**Exercise:** Give a quick argument for above equality.
Semidefinite Program (SDP)

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad c^T x \\
\text{s.t.} & \quad A(x) := A_0 + x_1A_1 + x_2A_2 + \ldots + x_nA_n \succeq 0.
\end{align*}
\]
Semidefinite Program (SDP)

\[
\min_{x \in \mathbb{R}^n} c^T x \\
\text{s.t. } A(x) := A_0 + x_1A_1 + x_2A_2 + \ldots + x_nA_n \succeq 0.
\]

- \(A_0, \ldots, A_n\) are real, symmetric matrices
- Inequality \(A \preceq B\) means \(B - A\) is **semidefinite**
- Also a **cone program** (conic optimization problem)
Semidefinite Program (SDP)

\[
\min_{x \in \mathbb{R}^n} \quad c^T x \\
\text{s.t.} \quad A(x) := A_0 + x_1A_1 + x_2A_2 + \ldots + x_nA_n \succeq 0.
\]

- \(A_0, \ldots, A_n\) are real, symmetric matrices
- Inequality \(A \preceq B\) means \(B - A\) is \textit{semidefinite}
- Also a \textit{cone program} (conic optimization problem)
- SDP \supset SOCP \supset QP \supset LP
- **Exercise:** Write LPs, QPs, and SOCPs as SDPs
Semidefinite Program (SDP)

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad c^T x \\
\text{s.t.} & \quad A(x) := A_0 + x_1 A_1 + x_2 A_2 + \ldots + x_n A_n \succeq 0.
\end{align*}
\]

- \(A_0, \ldots, A_n\) are real, symmetric matrices
- Inequality \(A \preceq B\) means \(B - A\) is semidefinite
- Also a cone program (conic optimization problem)
- SDP \(\supset\) SOCP \(\supset\) QP \(\supset\) LP
- **Exercise:** Write LPs, QPs, and SOCPs as SDPs
- Feasible set of SDP is \(\{\text{semidefinite cone } \cap \text{hyperplanes}\}\)

**Explore:** Which convex problems **representable** as SDPs? (This is an important topic in optimization theory).
Examples

♠ Eigenvalue optimization: $\min_x \lambda_{\max}(A(x))$

$$\min \ t \ \text{s.t.} \ A(x) \preceq tI.$$
Examples

♠ Eigenvalue optimization: \( \min_x \lambda_{\text{max}}(A(x)) \)

\[
\begin{align*}
\min & \quad t \\
\text{s.t.} & \quad A(x) \preceq tI.
\end{align*}
\]

♠ Norm minimization: \( \min_x \|A(x)\| \)

\[
\begin{align*}
\min & \quad t \\
\text{s.t.} & \quad \begin{bmatrix}
  tI & A(x)^T \\
  A(x) & tI
\end{bmatrix} \preceq 0.
\end{align*}
\]
Examples

♠ Eigenvalue optimization: \( \min_x \lambda_{\text{max}}(A(x)) \)
\[
\min t \quad \text{s.t.} \quad A(x) \preceq tI.
\]

♠ Norm minimization: \( \min_x \|A(x)\| \)
\[
\min t \quad \text{s.t.} \quad \begin{bmatrix} tI & A(x)^T \\ A(x) & tI \end{bmatrix} \succeq 0.
\]

♠ More examples – see CVX documentation and BV book

Explore: SDP relaxations of nonconvex probs: important technique, starting with MAXCUT SDP (Goemans-Williamson).

Explore: Sum-of-squares (SOS) optimization, Lasserre hierarchy of relaxations; see also: https://www.sumofsquares.org
Duality

(Weak duality, strong duality)
Primal problem

Let \( f_i : \mathbb{R}^n \rightarrow \mathbb{R} \ (1 \leq i \leq m) \). Generic nonlinear program

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad f_i(x) \leq 0, \quad 1 \leq i \leq m, \\
& \quad x \in \{ \text{dom} f \cap \text{dom} f_1 \cap \cdots \cap \text{dom} f_m \}.
\end{align*}
\]
Primal problem

Let \( f_i : \mathbb{R}^n \to \mathbb{R} \) (\( 1 \leq i \leq m \)). Generic nonlinear program

\[
\begin{align*}
\text{min} & \quad f(x) \\
\text{s.t.} & \quad f_i(x) \leq 0, \quad 1 \leq i \leq m, \\
& \quad x \in \{\text{dom} f \cap \text{dom} f_1 \cap \cdots \cap \text{dom} f_m\}.
\end{align*}
\]

(P)

Domain: The set \( \mathcal{X} := \{\text{dom} f \cap \text{dom} f_1 \cap \cdots \cap \text{dom} f_m\} \)

- We call (P) the **primal problem**
- The variable \( x \) is the **primal variable**
Primal problem

Let $f_i : \mathbb{R}^n \to \mathbb{R}$ ($1 \leq i \leq m$). Generic nonlinear program

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad f_i(x) \leq 0, \quad 1 \leq i \leq m, \\
& \quad x \in \{\text{dom } f \cap \text{dom } f_1 \cap \ldots \cap \text{dom } f_m\}.
\end{align*}
\]

\[(P)\]

**Domain:** The set $\mathcal{X} := \{\text{dom } f \cap \text{dom } f_1 \cap \ldots \cap \text{dom } f_m\}

- We call $(P)$ the **primal problem**
- The variable $x$ is the **primal variable**

---

Lagrangians and Duality
The reader will find no figures in this work. The methods which I set forth do not require either constructions or geometrical or mechanical reasonings: but only algebraic operations, subject to a regular and uniform rule of procedure.

—Joseph-Louis Lagrange
Preface to Mécanique Analytique
To primal, associate **Lagrangian** $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}_+^m \to (-\infty, \infty)$,

$$\mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x).$$
Lagrangian

To primal, associate **Lagrangian** $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}_+^m \rightarrow (-\infty, \infty)$,

$$\mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x).$$

♣ Variables $\lambda \in \mathbb{R}_+^m$ called **Lagrange multipliers**
To primal, associate **Lagrangian** $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_+ \to (-\infty, \infty]$,

$$\mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x).$$

♠ Variables $\lambda \in \mathbb{R}^m_+$ called **Lagrange multipliers**

♠ Suppose $x$ feasible, and $\lambda \geq 0$. Lower-bound holds:

$$f(x) \geq \mathcal{L}(x, \lambda) \quad \forall x \in \mathcal{X}, \; \lambda \in \mathbb{R}^m_+.$$
To primal, associate **Lagrangian** \( \mathcal{L} : \mathbb{R}^n \times \mathbb{R}_+^m \rightarrow (-\infty, \infty] \),

\[
\mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^m \lambda_i f_i(x).
\]

♠ Variables \( \lambda \in \mathbb{R}_+^m \) called **Lagrange multipliers**

♠ Suppose \( x \) feasible, and \( \lambda \geq 0 \). Lower-bound holds:

\[
f(x) \geq \mathcal{L}(x, \lambda) \quad \forall x \in \mathcal{X}, \lambda \in \mathbb{R}_+^m.
\]

♠ In other words,

\[
\sup_{\lambda \in \mathbb{R}_+^m} \mathcal{L}(x, \lambda) = \begin{cases} f(x), & \text{if } x \text{ feasible,} \\ +\infty, & \text{otherwise.} \end{cases}
\]

Proof on next slide
Lagrangian – proof

\[ \mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \]

\[ \Rightarrow f(x) \geq \mathcal{L}(x, \lambda), \forall x \in \mathcal{X}, \lambda \in \mathbb{R}_+^m; \text{ so primal optimal (value)} \]

\[ p^* = \inf_{x \in \mathcal{X}} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda). \]
Lagrangian – proof

\[ \mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \]

- \[ f(x) \geq \mathcal{L}(x, \lambda), \forall x \in X, \lambda \in \mathbb{R}_+^m \]; so **primal optimal** (value)

\[ p^* = \inf_{x \in X} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda). \]

- If \( x \) is not feasible, then some \( f_i(x) > 0 \)
Lagrangian – proof

\[ \mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \]

- \( f(x) \geq \mathcal{L}(x, \lambda) \), \( \forall x \in \mathcal{X}, \lambda \in \mathbb{R}^m_+ \); so **primal optimal** (value)

\[ p^* = \inf_{x \in \mathcal{X}} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda). \]

- If \( x \) is not feasible, then some \( f_i(x) > 0 \)
- In this case, inner sup is \( +\infty \), so claim true by definition
\[
\mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x).
\]

\[\nabla \, f(x) \geq \mathcal{L}(x, \lambda), \forall x \in \mathcal{X}, \lambda \in \mathbb{R}_+^m; \text{ so } \text{primal optimal} \text{ (value)}
\]

\[p^* = \inf_{x \in \mathcal{X}} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda).
\]

\[\nabla \text{ If } x \text{ is not feasible, then some } f_i(x) > 0
\]
\[\nabla \text{ In this case, inner sup is } +\infty, \text{ so claim true by definition}
\]
\[\nabla \text{ If } x \text{ is feasible, each } f_i(x) \leq 0, \text{ so } \sup_{\lambda} \sum_i \lambda_i f_i(x) = 0
\]
Dual value

\[ \mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \]

Primal value \( \in [\neg\infty, +\infty] \)

\[ p^* = \inf_{x \in \mathcal{X}} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda). \]

Observe that \( g(\lambda) \) is always concave!
Dual value

\[ \mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \]

Primal value \( \in [\mathbb{R}, +\infty] \)

\[ p^* = \inf_{x \in \mathcal{X}} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda). \]

Dual value \( \in [\mathbb{R}, +\infty] \)

\[ d^* = \sup_{\lambda \geq 0} \inf_{x \in \mathcal{X}} \mathcal{L}(x, \lambda). \]

Observe that \( g(\lambda) \) is always concave!
Dual value

\[ \mathcal{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \]

**Primal value** \( \in \mathbb{R} \)

\[ p^* = \inf_{x \in X} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda). \]

**Dual value** \( \in \mathbb{R} \)

\[ d^* = \sup_{\lambda \geq 0} \inf_{x \in X} \mathcal{L}(x, \lambda). \]

**Dual function**

\[ g(\lambda) := \inf_{x \in X} \mathcal{L}(x, \lambda). \]

Observe that \( g(\lambda) \) is always concave!
**Weak duality theorem**

**Theorem.** (*Weak duality*). $p^* \geq d^*$. (i.e., WD always holds)

**Proof:**

1. $f(x') \geq \mathcal{L}(x', \lambda) \quad \forall x' \in \mathcal{X}$

2. Thus, for any $x \in \mathcal{X}$, we have $f(x) \geq \inf_{x'} \mathcal{L}(x', \lambda) = g(\lambda)$

3. Now minimize over $x$ on lhs to obtain

$$\forall \lambda \in \mathbb{R}^m_+ \quad p^* \geq g(\lambda).$$

4. Thus, taking sup over $\lambda \in \mathbb{R}^m_+$ we obtain $p^* \geq d^*$. 
\[ \begin{align*}
\min \quad & f(x) \\
\text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\
& h_i(x) = 0, \quad i = 1, \ldots, p.
\end{align*} \]

**Exercise:** Show that we get the Lagrangian dual

\[ g : \mathbb{R}_+^m \times \mathbb{R}^p : (\lambda, \nu) \mapsto \inf_x \mathcal{L}(x, \lambda, \nu), \]

Lagrange variable \( \nu \) corresponds to the equality constraints.

**Exercise:** Prove that \( p^* \geq \sup_{\lambda \geq 0, \nu \in \mathbb{R}^p} g(\lambda, \nu) = d^*. \)
Exercises: Some duals

Derive Lagrangian duals for the following problems

- Least-norm solution of linear equations: $\min x^T x \text{ s.t. } Ax = b$
- Dual of an LP
- Dual of an SOCP
- Dual of an SDP
- Study example (5.7) in BV (binary QP)
Strong duality
Duality gap

\[ p^* - d^* \]
Duality gap

\[ p^* - d^* \]

*Strong duality* holds if duality gap is zero: \( p^* = d^* \)
Duality gap

\[ p^* - d^* \]

*Strong duality* holds if duality gap is zero: \( p^* = d^* \)

Several *sufficient* conditions known!
Duality gap

\[ p^* - d^* \]

*Strong duality* holds if duality gap is zero: \( p^* = d^* \)

Several *sufficient* conditions known!

“Easy” necessary and sufficient conditions: *unknown*
Abstract duality gap theorem*

**Theorem.** Let $v : \mathbb{R}^m \to \mathbb{R}$ be the primal value function

$$v(u) := \inf \{ f(x) \mid f_i(x) \leq u_i, \ 1 \leq i \leq m \}.$$ 

The following relations hold:

1. $p^* = v(0)$
2. $v^*(-\lambda) = \begin{cases} -g(\lambda) & \lambda \geq 0 \\ +\infty & \text{otherwise.} \end{cases}$
3. $d^* = v^{**}(0)$

**So if** $v(0) = v^{**}(0)$ **we have strong duality**

**Remark:** Conditions such as Slater’s ensure $\partial v(0) \neq \emptyset$, which ensures $v$ is finite and lsc at $0$, whereby $v(0) = v^{**}(0)$ holds.
Slater’s sufficient conditions

\[
\begin{aligned}
&\min f(x) \\
&\text{s.t. } f_i(x) \leq 0, \quad 1 \leq i \leq m, \\
&\quad Ax = b.
\end{aligned}
\]
Slater’s sufficient conditions

\[ \min \ f(x) \]
\[ \text{s.t. } f_i(x) \leq 0, \quad 1 \leq i \leq m, \]
\[ Ax = b. \]

**Constraint qualification:** There exists \( x \in \text{ri } \mathcal{X} \) s.t.

\[ f_i(x) < 0, \quad Ax = b. \]

In words: there is a **strictly feasible** point.
Slater’s sufficient conditions

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad f_i(x) \leq 0, \quad 1 \leq i \leq m, \\
& \quad Ax = b.
\end{align*}
\]

Constraint qualification: There exists \(x \in \text{ri } \mathcal{X}\) s.t.

\[
\begin{align*}
f_i(x) < 0, \quad Ax = b.
\end{align*}
\]

In words: there is a strictly feasible point.

**Theorem.** Let the primal problem be convex. If there is a point that is strictly feasible for the non-affine constraints (merely feasible for affine), then strong duality holds. Moreover, in this case, the dual optimal is attained (i.e., \(\partial v(0) \neq \emptyset\)).

See BV §5.3.2 for a proof; (above, \(v\) is the primal value function)
Example with positive duality-gap

\[
\min_{x, y} e^{-x} \quad x^2/y \leq 0,
\]

over the domain \( \mathcal{X} = \{(x, y) \mid y > 0\} \).
Example with positive duality-gap

\[
\min_{x,y} e^{-x} \quad x^2 / y \leq 0,
\]

over the domain \( \mathcal{X} = \{(x, y) \mid y > 0\} \).

Clearly, only feasible \( x = 0 \). So \( p^* = 1 \)
Example with positive duality-gap

\[
\min_{x,y} e^{-x} \quad \frac{x^2}{y} \leq 0,
\]

over the domain \( \mathcal{X} = \{(x, y) \mid y > 0\} \).

Clearly, only feasible \( x = 0 \). So \( p^* = 1 \)

\[
\mathcal{L}(x, y, \lambda) = e^{-x} + \lambda \frac{x^2}{y},
\]

so dual function is

\[
g(\lambda) = \inf_{x,y>0} e^{-x} + \lambda x^2 y = \begin{cases} 
0 & \lambda \geq 0 \\
-\infty & \lambda < 0.
\end{cases}
\]
Example with positive duality-gap

\[
\min_{x, y} e^{-x} \quad x^2 / y \leq 0,
\]
over the domain \( \mathcal{X} = \{(x, y) \mid y > 0\} \).
Clearly, only feasible \( x = 0 \). So \( p^* = 1 \)

\[
\mathcal{L}(x, y, \lambda) = e^{-x} + \lambda x^2 / y,
\]
so dual function is

\[
g(\lambda) = \inf_{x, y > 0} e^{-x} + \lambda x^2 y = \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0 \end{cases}
\]

**Dual problem**

\[
d^* = \max_\lambda 0 \quad \text{s.t. } \lambda \geq 0.
\]
Example with positive duality-gap

\[ \min_{x,y} e^{-x} x^2 / y \leq 0, \]

over the domain \( \mathcal{X} = \{(x, y) \mid y > 0\} \).

Clearly, only feasible \( x = 0 \). So \( p^* = 1 \)

\[ \mathcal{L}(x, y, \lambda) = e^{-x} + \lambda x^2 / y, \]

so dual function is

\[ g(\lambda) = \inf_{x,y>0} e^{-x} + \lambda x^2 y = \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0. \end{cases} \]

**Dual problem**

\[ d^* = \max_{\lambda} 0 \quad \text{s.t. } \lambda \geq 0. \]

Thus, \( d^* = 0 \), and gap is \( p^* - d^* = 1 \).
Example with positive duality-gap

\[
\min_{x,y} e^{-x} \quad x^2/y \leq 0,
\]

over the domain \( \mathcal{X} = \{(x, y) \mid y > 0\} \).

Clearly, only feasible \( x = 0 \). So \( p^* = 1 \)

\[
\mathcal{L}(x, y, \lambda) = e^{-x} + \lambda x^2/y,
\]

so dual function is

\[
g(\lambda) = \inf_{x, y > 0} e^{-x} + \lambda x^2 y = \begin{cases} 
0 & \lambda \geq 0 \\
-\infty & \lambda < 0.
\end{cases}
\]

**Dual problem**

\[
d^* = \max_{\lambda} 0 \quad \text{s.t. } \lambda \geq 0.
\]

Thus, \( d^* = 0 \), and gap is \( p^* - d^* = 1 \).

Here, we had no strictly feasible solution.
Example: Support Vector Machine (SVM)

\[
\min_{x, \xi} \quad \frac{1}{2} \|x\|^2 + C \sum_i \xi_i \\
\text{s.t.} \quad Ax \geq 1 - \xi, \quad \xi \geq 0.
\]
Example: Support Vector Machine (SVM)

\[
\begin{align*}
\min_{x, \xi} & \quad \frac{1}{2} \|x\|_2^2 + C \sum_i \xi_i \\
\text{s.t.} & \quad Ax \geq 1 - \xi, \quad \xi \geq 0.
\end{align*}
\]

\[
L(x, \xi, \lambda, \nu) = \frac{1}{2} \|x\|_2^2 + C^T \xi - \lambda^T (Ax - 1 + \xi) - \nu^T \xi
\]

\[d^* = \max_{\lambda \geq 0, \nu \geq 0} g(\lambda, \nu) = \inf L(x, \xi, \lambda, \nu) = \begin{cases} 
\lambda^T 1 - \frac{1}{2} \|A^T \lambda\|_2^2, & \lambda^T \nu = C^T + \infty \\
\text{otherwise} & \end{cases} \]

Exercise: Using \(\nu \geq 0\), eliminate \(\nu\) from above dual and obtain the canonical dual SVM formulation.
Example: Support Vector Machine (SVM)

\[
\begin{align*}
\min_{x, \xi} & \quad \frac{1}{2} \|x\|^2 + C \sum_i \xi_i \\
\text{s.t.} & \quad Ax \geq 1 - \xi, \quad \xi \geq 0.
\end{align*}
\]

\[
L(x, \xi, \lambda, \nu) = \frac{1}{2} \|x\|^2 + C \lambda^T \xi - \lambda^T (Ax - 1 + \xi) - \nu^T \xi
\]

\[
g(\lambda, \nu) := \inf L(x, \xi, \lambda, \nu)
\]

\[
g(\lambda, \nu) = \begin{cases} 
\lambda^T 1 - \frac{1}{2} \|A^T \lambda\|^2 & \lambda + \nu = C 1 \\
+\infty & \text{otherwise}
\end{cases}
\]

\[
d^* = \max_{\lambda \geq 0, \nu \geq 0} g(\lambda, \nu)
\]

Exercise: Using \(\nu \geq 0\), eliminate \(\nu\) from above dual and obtain the canonical dual SVM formulation.
Example: norm regularized problems

\[
\min f(x) + \|Ax\|
\]
Example: norm regularized problems

\[
\min f(x) + \|Ax\|
\]

**Dual problem**

\[
\min_{y} f^{*}(-A^{T}y) \quad \text{s.t. } \|y\|_{*} \leq 1.
\]
Example: norm regularized problems

\[ \min f(x) + \|Ax\| \]

**Dual problem**

\[ \min_y f^*(-A^T y) \quad \text{s.t. } \|y\|_* \leq 1. \]

Say \( \|\bar{y}\|_* < 1 \), such that \( A^T \bar{y} \in \text{ri}(\text{dom } f^*) \), then we have strong duality—for instance if \( 0 \in \text{ri}(\text{dom } f^*) \)

**Exercise.** Write the constrained form of *group-lasso*:

\[ \min_{x_1, \ldots, x_T} \frac{1}{2} \left\| b - \sum_{j=1}^T A_j x_j \right\|_2^2 + \lambda \sum_{j=1}^T \|x_j\|_2. \]
Example: Dual via Fenchel conjugates

\[
\min_x f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0 \ (1 \leq i \leq m), \quad Ax = b.
\]

Introduce \( \nu \) and \( \lambda \) as dual variables; consider Lagrangian

\[
\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b)
\]

\( F^* \) seems rather opaque...
Example: Dual via Fenchel conjugates

\[
\min_x f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0 \quad (1 \leq i \leq m), \quad Ax = b.
\]

Introduce \(\nu\) and \(\lambda\) as dual variables; consider Lagrangian

\[
\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b)
\]

\[
g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)
\]
Example: Dual via Fenchel conjugates

$$\min_x f_0(x) \quad \text{s.t. } f_i(x) \leq 0 \ (1 \leq i \leq m), \ Ax = b.$$ 

Introduce $\nu$ and $\lambda$ as dual variables; consider Lagrangian

$$L(x, \lambda, \nu) := f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b)$$
$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$
$$g(\lambda, \nu) = -\nu^T b + \inf_x x^T A^T \nu + F(x)$$
Example: Dual via Fenchel conjugates

\[
\min_x f_0(x) \quad \text{s.t. } f_i(x) \leq 0 \ (1 \leq i \leq m), \ Ax = b.
\]

Introduce \( \nu \) and \( \lambda \) as dual variables; consider Lagrangian

\[
\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b)
\]

\[
g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)
\]

\[
g(\lambda, \nu) = -\nu^T b + \inf_x x^T A^T \nu + F(x)
\]

\[
F(x) := f_0(x) + \sum_i \lambda_i f_i(x)
\]

\( F^* \) seems rather opaque...

Suvrit Sra (suvrit@mit.edu) 6.881 Optimization for Machine Learning (2/23/21; Lecture 3)
Example: Dual via Fenchel conjugates

\[
\begin{align*}
\min_x f_0(x) \quad \text{s.t. } f_i(x) &\leq 0 \ (1 \leq i \leq m), \ Ax = b. \\
\end{align*}
\]

Introduce \( \nu \) and \( \lambda \) as dual variables; consider Lagrangian

\[
\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b)
\]

\[
g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)
\]

\[
g(\lambda, \nu) = -\nu^T b + \inf_x x^T A^T \nu + F(x)
\]

\[
F(x) := f_0(x) + \sum_i \lambda_i f_i(x)
\]

\[
g(\lambda, \nu) = -\nu^T b - \sup_x \langle x, -A^T \nu \rangle - F(x)
\]

\[F^*\] seems rather opaque...
Example: Dual via Fenchel conjugates

\[
\min_x f_0(x) \quad \text{s.t. } f_i(x) \leq 0 \ (1 \leq i \leq m), \ Ax = b.
\]

Introduce \( \nu \) and \( \lambda \) as dual variables; consider Lagrangian

\[
\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b)
\]

\[
g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)
\]

\[
g(\lambda, \nu) = -\nu^T b + \inf_x x^T A^T \nu + F(x)
\]

\[
F(x) := f_0(x) + \sum_i \lambda_i f_i(x)
\]

\[
g(\lambda, \nu) = -\nu^T b - \sup_x \langle x, -A^T \nu \rangle - F(x)
\]

\[
g(\lambda, \nu) = -\nu^T b - F^*(-A^T \nu).
\]

\( F^* \) seems rather opaque...
Example: Dual via Fenchel conjugates

Important trick: “variable splitting”

\[
\begin{align*}
\min_x f_0(x) & \quad \text{s.t.} \quad f_i(x_i) \leq 0, Ax = b \\
& \quad \quad \quad x = x_i, i = 1, \ldots, m.
\end{align*}
\]
Example: Dual via Fenchel conjugates

**Important trick:** “variable splitting”

\[
\begin{align*}
\min_x & \quad f_0(x) \quad \text{s.t.} \quad f_i(x_i) \leq 0, Ax = b \\
& \quad x = x_i, i = 1, \ldots, m. \\
\end{align*}
\]

\[
L(x, x_i, \lambda, \nu, \pi_i) = f_0(x) + \sum_i \lambda_i f_i(x_i) + \nu^T (Ax - b) + \sum_i \pi_i^T (x_i - x)
\]
Example: Dual via Fenchel conjugates

Important trick: “variable splitting”

\[
\min_x f_0(x) \quad \text{s.t.} \quad f_i(x_i) \leq 0, \ Ax = b
\]
\[
x = x_i, \ i = 1, \ldots, m.
\]

\[
\mathcal{L}(x, x_i, \lambda, \nu, \pi_i) := f_0(x) + \sum_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum_i \pi_i^T(x_i - x)
\]

\[
g(\lambda, \nu, \pi_i) = \inf_{x, x_i} \mathcal{L}(x, x_i, \lambda, \nu, \pi_i)
\]
Example: Dual via Fenchel conjugates

Important trick: “variable splitting”

\[
\min_x f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0, Ax = b
\]
\[
x = x_i, i = 1, \ldots, m.
\]

\[
\mathcal{L}(x, x_i, \lambda, \nu, \pi_i)
:= f_0(x) + \sum_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum_i \pi_i^T(x_i - x)
\]
\[
g(\lambda, \nu, \pi_i) = \inf_{x, x_i} \mathcal{L}(x, x_i, \lambda, \nu, \pi_i)
= -\nu^Tb + \inf_x \left( f_0(x) + \nu^T Ax - \sum_i \pi_i^T x \right)
\]
\[
+ \sum_i \inf_{x_i} \left( \pi_i^T x_i + \lambda_i f_i(x_i) \right),
\]
Example: Dual via Fenchel conjugates

Important trick: “variable splitting”

\[
\begin{align*}
& \min_x f_0(x) \quad \text{s.t.} \quad f_i(x_i) \leq 0, Ax = b \\
& \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad x = x_i, i = 1, \ldots, m.
\end{align*}
\]

\[
\mathcal{L}(x, x_i, \lambda, \nu, \pi_i) := f_0(x) + \sum_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum_i \pi_i^T(x_i - x)
\]

\[
g(\lambda, \nu, \pi_i) = \inf_{x, x_i} \mathcal{L}(x, x_i, \lambda, \nu, \pi_i)
\]

\[
= -\nu^T b + \inf_x \left( f_0(x) + \nu^T Ax - \sum_i \pi_i^T x \right)
\]

\[
+ \sum_i \inf_{x_i} \left( \pi_i^T x_i + \lambda_i f_i(x_i) \right),
\]

\[
= -\nu^T b - f^* \left( -A^T \nu + \sum_i \pi_i \right) - \sum_i (\lambda_i f_i)^*(-\pi_i).
\]

(\text{you may want to write } \sum_i \pi_i = s)
Exercise: the variable splitting trick

\[
\min_x \quad f(x) + h(x).
\]

**Exercise:** Fill in the details for the following steps

\[
\min_{x, z} \quad f(x) + h(z) \quad \text{s.t.} \quad x = z \\
L(x, z, \nu) = f(x) + h(z) + \nu^T(x - z) \\
g(\nu) = \inf_{x, z} L(x, z, \nu)
\]
Strong duality: nonconvex example

Trust region subproblem (TRS)

\[
\min \ x^T Ax + 2b^T x \quad x^T x \leq 1.
\]

A is symmetric but not necessarily semidefinite!
Strong duality: nonconvex example

Trust region subproblem (TRS)

\[ \min_x x^T A x + 2b^T x \quad x^T x \leq 1. \]

*Theorem.* TRS always has zero duality gap.

*Proof:* Read Section 5.2.4 of BV.

See the challenge problems on pg 18, Lect1
von Neumann minmax theorem

\textbf{(Simplified.)} Let $A$ be linear, $C$, $D$ be compact convex sets.

\[
\min_{x \in C} \max_{y \in D} \langle Ax, y \rangle = \max_{y \in D} \min_{x \in C} \langle Ax, y \rangle.
\]
von Neumann minmax theorem

(Simplified.) Let $A$ be linear, $C, D$ be compact convex sets.

\[
\min_{x \in C} \max_{y \in D} \langle Ax, y \rangle = \max_{y \in D} \min_{x \in C} \langle Ax, y \rangle.
\]

von Neumann proved this via fixed-point theory. By considering the Fenchel problem

\[
\min_{x} \ 1_C(x) + 1_D^*(Ax),
\]

we can conclude the theorem (some work required).