ELE539A: Optimization of Communication Systems
Lecture 2: Convex Optimization and Lagrange Duality

Professor M. Chiang
Electrical Engineering Department, Princeton University

February 7, 2007
Lecture Outline

- Convex optimization
- Optimality condition
- Lagrange dual problem
- Interpretations
- KKT optimality condition
- Sensitivity analysis

Thanks: Stephen Boyd (some materials and graphs from Boyd and Vandenberghe)
Convex Optimization

A convex optimization problem with variables $x$:

minimize $f_0(x)$

subject to $f_i(x) \leq 0, \ i = 1, 2, \ldots, m$

$a_i^T x = b_i, \ i = 1, 2, \ldots, p$

where $f_0, f_1, \ldots, f_m$ are convex functions.

- Minimize convex objective function (or maximize concave objective function)

- Upper bound inequality constraints on convex functions ($\Rightarrow$ Constraint set is convex)

- Equality constraints must be affine
Convex Optimization

- Epigraph form:

  minimize $t$
  subject to $f_0(x) - t \leq 0$
  $f_i(x) \leq 0, \ i = 1, 2, \ldots, m$
  $a_i^T x = b_i, \ i = 1, 2, \ldots, p$

- Not in convex optimization form:

  minimize $x_1^2 + x_2^2$
  subject to $\frac{x_1}{1+x_2} \leq 0$
  $(x_1 + x_2)^2 = 0$

Now transformed into a convex optimization problem:

  minimize $x_1^2 + x_2^2$
  subject to $x_1 \leq 0$
  $x_1 + x_2 = 0$
Locally Optimal $\Rightarrow$ Globally Optimal

Given $x$ is locally optimal for a convex optimization problem, i.e., $x$ is feasible and for some $R > 0$,

$$f_0(x) = \inf\{f_0(z)|z \text{ is feasible}, \|z - x\|_2 \leq R\}$$

Suppose $x$ is not globally optimal, i.e., there is a feasible $y$ such that $f_0(y) < f_0(x)$

Since $\|y - x\|_2 > R$, we can construct a point $z = (1 - \theta)x + \theta y$ where $\theta = \frac{R}{2\|y - x\|_2}$. By convexity of feasible set, $z$ is feasible. By convexity of $f_0$, we have

$$f_0(z) \leq (1 - \theta)f_0(x) + \theta f_0(y) < f_0(x)$$

which contradicts locally optimality of $x$

Therefore, there exists no feasible $y$ such that $f_0(y) < f_0(x)$
Optimality Condition for Differentiable $f_0$

$x$ is optimal for a convex optimization problem iff $x$ is feasible and for all feasible $y$:

$$\nabla f_0(x)^T(y - x) \geq 0$$

$-\nabla f_0(x)$ is supporting hyperplane to feasible set

Unconstrained convex optimization: condition reduces to:

$$\nabla f_0(x) = 0$$

Proof: take $y = x - t\nabla f_0(x)$ where $t \in \mathbb{R}_+$. For small enough $t$, $y$ is feasible, so $\nabla f_0(x)^T(y - x) = -t\|\nabla f_0(x)\|^2_2 \geq 0$. Thus $\nabla f_0(x) = 0$
Unconstrained Quadratic Optimization

Minimize $f_0(x) = \frac{1}{2} x^T P x + q^T x + r$

$P$ is positive semidefinite. So it's a convex optimization problem

$x$ minimizes $f_0$ iff $(P, q)$ satisfy this linear equality:

$$\nabla f_0(x) = Px + q = 0$$

• If $q \notin \mathcal{R}(P)$, no solution. $f_0$ unbounded below
• If $q \in \mathcal{R}(P)$ and $P \succ 0$, there is a unique minimizer $x^* = -P^{-1} q$
• If $q \in \mathcal{R}(P)$ and $P$ is singular, set of optimal $x$: $-P^\dagger q + \mathcal{N}(P)$
Duality Mentality

Bound or solve an optimization problem via a different optimization problem!

We’ll develop the basic Lagrange duality theory for a general optimization problem, then specialize for convex optimization
**Lagrange Dual Function**

An optimization problem in standard form:

minimize \( f_0(x) \)

subject to \( f_i(x) \leq 0, \ i = 1, 2, \ldots, m \)
\( h_i(x) = 0, \ i = 1, 2, \ldots, p \)

Variables: \( x \in \mathbb{R}^n \). Assume nonempty feasible set

Optimal value: \( p^* \). Optimizer: \( x^* \)

Idea: augment objective with a weighted sum of constraints

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

**Lagrange multipliers (dual variables):** \( \lambda \succeq 0, \nu \)

**Lagrange dual function:** \( g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \)
Lower Bound on Optimal Value

Claim: \( g(\lambda, \nu) \leq p^*, \ \forall \lambda \succeq 0, \nu \)

Proof: Consider feasible \( \tilde{x} \):

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

since \( f_i(\tilde{x}) \leq 0 \) and \( \lambda_i \geq 0 \)

Hence, \( g(\lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}) \) for all feasible \( \tilde{x} \)

Therefore, \( g(\lambda, \nu) \leq p^* \)
Lagrange Dual Function and Conjugate Function

- Lagrange dual function $g(\lambda, \nu)$
- Conjugate function: $f^*(y) = \sup_{x \in \text{dom}} f(y^T x - f(x))$

Consider linearly constrained optimization:

$$\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad Ax \preceq b \\
& \quad Cx = d
\end{align*}$$

$$
\begin{align*}
g(\lambda, \nu) &= \inf_x \left( f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \right) \\
&= -b^T \lambda - d^T \nu + \inf_x \left( f_0(x) + (A^T \lambda + C^T \nu)^T x \right) \\
&= -b^T \lambda - d^T \nu - f_0^*(-A^T \lambda - C^T \nu)
\end{align*}
$$
Example

We’ll use the simplest version of entropy maximization as our example for the rest of this lecture on duality. Entropy maximization is an important basic problem in information theory:

\[
\begin{align*}
\text{minimize} & \quad f_0(x) = \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & \quad Ax \preceq b \\
& \quad 1^T x = 1
\end{align*}
\]

Since the conjugate function of \( u \log u \) is \( e^{y-1} \), by independence of the sum, we have

\[
f_0^*(y) = \sum_{i=1}^{n} e^{y_i-1}
\]

Therefore, dual function of entropy maximization is

\[
g(\lambda, \nu) = -b^T \lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T \lambda}
\]

where \( a^i \) are columns of \( A \)
Lagrange Dual Problem

Lower bound from Lagrange dual function depends on \((\lambda, \nu)\). What’s the best lower bound that can be obtained from Lagrange dual function?

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
\]

This is the Lagrange dual problem with dual variables \((\lambda, \nu)\)

Always a convex optimization! (Dual objective function always a concave function since it’s the infimum of a family of affine functions in \((\lambda, \nu)\))

Denote the optimal value of Lagrange dual problem by \(d^*\)
Weak Duality

What’s the relationship between $d^*$ and $p^*$?

Weak duality always hold (even if primal problem is not convex):

$$d^* \leq p^*$$

Optimal duality gap:

$$p^* - d^*$$

Efficient generation of lower bounds through (convex) dual problem
Strong Duality

**Strong duality** (zero optimal duality gap):

\[ d^* = p^* \]

If strong duality holds, solving dual is ‘equivalent’ to solving primal. But strong duality does not always hold.

Convexity and constraint qualifications \( \Rightarrow \) Strong duality

A simple constraint qualification: **Slater’s condition** (there exists strictly feasible primal variables \( f_i(x) < 0 \) for non-affine \( f_i \))

Another reason why convex optimization is ‘easy’
Example

Primal optimization problem (variables $x$):

minimize $f_0(x) = \sum_{i=1}^{n} x_i \log x_i$

subject to $Ax \preceq b$

$1^T x = 1$

Dual optimization problem (variables $\lambda, \nu$):

maximize $-b^T \lambda - \nu - e^{-\nu} - 1 \sum_{i=1}^{n} e^{-a_i^T \lambda}$

subject to $\lambda \succeq 0$

Analytically maximize over the unconstrained $\nu \Rightarrow$ Simplified dual optimization problem (variables $\lambda$):

maximize $-b^T \lambda - \log \sum_{i=1}^{n} \exp(-a_i^T \lambda)$

subject to $\lambda \succeq 0$

Strong duality holds
Saddle Point Interpretation

Assume no equality constraints. We can express primal optimal value as

\[ p^* = \inf_x \sup_{\lambda \geq 0} L(x, \lambda) \]

By definition of dual optimal value:

\[ d^* = \sup_{\lambda \geq 0} \inf_x L(x, \lambda) \]

Weak duality (max min inequality):

\[ \sup_{\lambda \geq 0} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda \geq 0} L(x, \lambda) \]

Strong duality (saddle point property):

\[ \sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda) \]
Economics Interpretation

- Primal objective: cost of operation
- Primal constraints: can be violated
- Dual variables: price for violating the corresponding constraint (dollar per unit violation). For the same price, can sell ‘unused violation’ for revenue
- Lagrangian: total cost
- Lagrange dual function: optimal cost as a function of violation prices
- Weak duality: optimal cost when constraints can be violated is less than or equal to optimal cost when constraints cannot be violated, for any violation prices
- Duality gap: minimum possible arbitrage advantage
- Strong duality: can price the violations so that there is no arbitrage advantages
Complementary Slackness

Assume strong duality holds:

\[ f_0(x^*) = g(\lambda^*, \nu^*) \]

\[ = \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \]

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

\[ \leq f_0(x^*) \]

So the two inequalities must hold with equality. This implies:

\[ \lambda_i^* f_i(x^*) = 0, \quad i = 1, 2, \ldots, m \]

Complementary Slackness Property:

\[ \lambda_i^* > 0 \quad \Rightarrow \quad f_i(x^*) = 0 \]

\[ f_i(x^*) < 0 \quad \Rightarrow \quad \lambda_i^* = 0 \]
KKT Optimality Conditions

Since $x^*$ minimizes $L(x, \lambda^*, \nu^*)$ over $x$, we have

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda^*_i \nabla f_i(x^*) + \sum_{i=1}^{p} \nu^*_i \nabla h_i(x^*) = 0$$

Karush-Kuhn-Tucker optimality conditions:

$$f_i(x^*) \leq 0, \ h_i(x^*) = 0, \ \lambda^*_i \geq 0$$

$$\lambda^*_i f_i(x^*) = 0$$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda^*_i \nabla f_i(x^*) + \sum_{i=1}^{p} \nu^*_i \nabla h_i(x^*) = 0$$

- Any optimization (with differentiable objective and constraint functions) with strong duality, KKT condition is necessary condition for primal-dual optimality
- Convex optimization (with differentiable objective and constraint functions) with Slater’s condition, KKT condition is also sufficient condition for primal-dual optimality (useful for theoretical and numerical purposes)
Waterfilling

maximize \( \sum_{i=1}^{n} \log(\alpha_i + x_i) \)

subject to \( x \succeq 0, \ 1^T x = 1 \)

Variables: \( x \) (powers). Constants: \( \alpha \) (noise)

KKT conditions:

\[
x^* \succeq 0, \ 1^T x^* = 1, \ \lambda^* \succeq 0 \\
\lambda_i^* x_i^* = 0, \ -1/(\alpha_i + x_i) - \lambda_i^* + \nu^* = 0
\]

Since \( \lambda^* \) are slack variables, reduce to

\[
x^* \succeq 0, \ 1^T x^* = 1 \\
x_i^* (\nu^* - 1/(\alpha_i^* + x_i^*)) = 0, \ \nu^* \geq 1/(\alpha_i + x_i^*)
\]

If \( \nu^* < 1/\alpha_i \), \( x_i^* > 0 \). So \( x_i^* = 1/\nu^* - \alpha_i \). Otherwise, \( x_i^* = 0 \)

Thus, \( x_i^* = [1/\nu^* - \alpha_i]^+ \) where \( \nu^* \) is such that \( \sum_i x_i^* = 1 \)
Global Sensitivity Analysis

Perturbed optimization problem:

minimize $f_0(x)$
subject to $f_i(x) \leq u_i$, $i = 1, 2, \ldots, m$
$h_i(x) = v_i$ $i = 1, 2, \ldots, p$

Optimal value $p^*(u, v)$ as a function of parameters $(u, v)$

Assume strong duality and that dual optimum is attained:

$p^*(0, 0) = g(\lambda^*, \nu^*) \leq f_0(x) + \sum_i \lambda^*_i f_i(x) + \sum_i \nu^*_i h_i(x) \leq f_0(x) + \lambda^T u + \nu^T v$
$p^*(u, v) \geq p^*(0, 0) - \lambda^T u - \nu^T v$

• If $\lambda^*_i$ is large, tightening $i$th constraint $(u_i < 0)$ will increase optimal value greatly
• If $\lambda^*_i$ is small, loosening $i$th constraint $(u_i > 0)$ will reduce optimal value only slightly
Local Sensitivity Analysis

Assume that $p^*(u, v)$ is differentiable at $(0, 0)$:

$$
\lambda^*_i = -\frac{\partial p^*(0, 0)}{\partial u_i}, \quad \nu^*_i = -\frac{\partial p^*(0, 0)}{\partial v_i}
$$

Shadow price interpretation of Lagrange dual variables

Small $\lambda^*_i$ means tightening or loosening $i$th constraint will not change optimal value by much.
Lecture Summary

- Convexity mentality. Convex optimization is ‘nice’ for several reasons: local optimum is global optimum, zero optimal duality gap (under technical conditions), KKT optimality conditions are necessary and sufficient

- Duality mentality. Can always bound primal through dual, sometimes solve primal through dual

- Primal-dual: where is the optimum, how sensitive it is to perturbations

Readings: Sections 4.1-4.2 and 5.1-5.6 in Boyd and Vandenberghe