Lecture 5: Duality

Xiugang Wu

University of Delaware

Fall 2019
Outline

• Lagrange Dual Function
• Lagrange Dual Problem
• Geometric Interpretation
• Saddle-point Interpretation
• Optimality Conditions
• Perturbation and Sensitivity Analysis
• Examples
• Generalized Inequalities
Outline

• Lagrange Dual Function
• Lagrange Dual Problem
• Geometric Interpretation
• Saddle-point Interpretation
• Optimality Conditions
• Perturbation and Sensitivity Analysis
• Examples
• Generalized Inequalities
Lagrangian

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \ i = 1, 2, \ldots, m \\
& \quad h_i(x) = 0, \ i = 1, 2, \ldots, p
\end{align*}
\]

Lagrangian: \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \), with \( \text{dom}L = D \times \mathbb{R}^m \times \mathbb{R}^p \)

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

- weighted sum of objective and constraint functions
- \( \lambda_i \) is Lagrange multiplier associated with \( i \)th inequality constraint
- \( \nu_i \) is Lagrange multiplier associated with \( i \)th equality constraint
- \( \lambda, \nu \) are called dual variables or Lagrange multiplier vectors
Lagrange Dual Function

Lagrange dual function: $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$

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

- For each $x$ the Lagrangian $L(x, \lambda, \nu)$ is affine in $(\lambda, \nu)$; thus, pointwise infimum over $x \in \mathcal{D}$ yields concave $g(\lambda, \nu)$

- For any $\lambda \geq 0$ and $\nu$, $g(\lambda, \nu) \leq p^*$

- Linear approximation interpretation: Consider the following unconstrained problem with the same optimal point and optimal value as the original one:

  $$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x))$$

  where $I_-$ and $I_0$ are equal to 0 if the argument satisfies the subscript condition and infinity otherwise. For any $\lambda \geq 0$ and $\nu$, $L(x, \lambda, \nu)$ is simply an under-estimator of the above formulation, and therefore minimizing $L(x, \lambda, \nu)$ yields a lower bound $g$ of the original optimal value $p^*$. 
Least-Norm Solution of Linear Equations

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

- Lagrangian is

\[
L(x, \nu) = x^T x + \nu^T (Ax - b)
\]

which is a convex quadratic function of \(x\). Taking derivative of \(L(x, \nu)\) yields

\[
\nabla_x L(x, \nu) = 2x + A^T \nu
\]

which vanishes when \(x = -\frac{1}{2} A^T \nu\).

- Therefore, the Lagrange dual function is

\[
g(\nu) = L \left(-\frac{1}{2} A^T \nu, \nu\right) = (-1/4)\nu^T AA^T \nu - \nu^T b
\]

which is concave in \(\nu\)

- Lower bound property:

\[
p^* \geq (-1/4)\nu^T AA^T \nu - \nu^T b, \quad \forall \nu
\]
Standard Form LP

minimize $c^T x$
subject to $Ax = b$
$x \geq 0$

- Lagrangian is

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b) = (c^T - \lambda^T + \nu^T A)x - \nu^T b$$

- Lagrange dual function is

$$g(\lambda, \nu) = \begin{cases} -\nu^T b & c^T - \lambda^T + \nu^T A = 0 \\ -\infty & \text{otherwise} \end{cases}$$

which is linear on affine domain and hence concave

- Lower bound property:

$$p^* \geq -\nu^T b, \text{ for any } \nu \text{ such that } c^T + \nu^T A \geq 0$$
Lagrange Dual and Conjugate Function

minimize \( f_0(x) \)

subject to \( Ax \leq b, Cx = d \)

Dual function:

\[
g(\lambda, \nu) = \inf_{x \in \text{dom} f_0} (f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d))
\]

\[
= \inf_{x \in \text{dom} f_0} (f_0(x) + (A^T \lambda + C^T \nu)^T x - \lambda^T b - \nu^T d)
\]

\[
= -f_0^*(-A^T \lambda - C^T \nu) - \lambda^T b - \nu^T d
\]

- recall definition of conjugate: \( f^*(y) = \sup_{x \in \text{dom} f} (y^T x - f(x)) \)

- simplifies derivation of dual if conjugate of \( f_0 \) is known

- example: entropy maximization

\[
f_0(x) = \sum_{i=1}^{n} x_i \log x_i, \quad f_0^*(y) = \sum_{i=1}^{n} e^{y_i-1}
\]
Outline

• Lagrange Dual Function
• **Lagrange Dual Problem**
• Geometric Interpretation
• Saddle-point Interpretation
• Optimality Conditions
• Perturbation and Sensitivity Analysis
• Examples
• Generalized Inequalities
Lagrange Dual Problem

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

- called the Lagrange dual problem, associated with the primal problem
- finds best lower bound on \( p^* \), obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted by \( d^* \)
- \( \lambda, \nu \) are said to be dual feasible if \( \lambda \geq 0 \) and \( (\lambda, \nu) \in \text{dom} \ g \)
- often simplified by making implicit constraints \( (\lambda, \nu) \in \text{dom} \ g \) explicit

example: standard form LP and its dual

minimize \( c^T x \) \hspace{2cm} \text{maximize} \hspace{0.5cm} -b^T \nu \)
subject to \( Ax = b \) \hspace{2cm} \text{subject to} \hspace{0.5cm} A^T \nu + c \geq 0 \)
\( x \geq 0 \)
Weak and Strong Duality

weak duality: $d^* \leq p^*$
- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bound for difficult primal problems
- $p^* - d^*$ is referred to as the optimal duality gap, which is always nonnegative

strong duality: $d^* = p^*$
- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called constraint qualifications
Slater’s Constraint Qualifications

strong duality holds for a convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, i = 1, 2, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

if it is strictly feasible, i.e.,

\[
\exists x \in \text{int} D : f_i(x) < 0, i = 1, 2, \ldots, m, Ax = b
\]

- also guarantees that the dual optimum is attained if \( d^* = p^* > -\infty \), i.e. there exists a dual feasible \((\lambda^*, \nu^*)\) such that \( g(\lambda^*, \nu^*) = d^* = p^* \)

- can be sharpened; there also exist many other types of constraint qualifications

- sufficient but not necessary condition; strong duality can hold for convex problems not satisfying Slater’s condition, or for nonconvex problems
Outline

• Lagrange Dual Function
• Lagrange Dual Problem
• Geometric Interpretation
• Saddle-point Interpretation
• Optimality Conditions
• Perturbation and Sensitivity Analysis
• Examples
• Generalized Inequalities
**Geometric Interpretation**

for simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function:

$$g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u), \quad \mathcal{G} = \{(f_1(x), f_0(x)) | x \in \mathcal{D}\}$$

- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$

- hyperplane intersects $t$-axis at $t = g(\lambda)$
Geometric Interpretation

epigraph variation: same interpretation if $\mathcal{G}$ is replaced with

$$\mathcal{A} = \{(u, t)|f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$$

strong duality
- holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^*)$
- for convex problem, $\mathcal{A}$ is convex, hence has supp. hyperplane at $(0, p^*)$
- Slater’s condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^*)$ must be non-vertical
Outline

• Lagrange Dual Function
• Lagrange Dual Problem
• Geometric Interpretation
• Saddle-point Interpretation
• Optimality Conditions
• Perturbation and Sensitivity Analysis
• Examples
• Generalized Inequalities
Max-Min Characterization of Duality

- for simplicity, assume no equality constraints

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

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

- strong duality: \[ \sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda) \]

- In general, we have max-min inequality

\[ \sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z) \]

When the equality holds in the above, we say \( f \) (and \( W, Z \)) satisfy the strong max-min property or the saddle-point property. Game interpretation.

- strong max-min property holds only in special cases, e.g., when \( f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) is the Lagrangian of a problem for which strong duality attains, \( W = \mathbb{R}^n \) and \( Z = \mathbb{R}^m_+ \)
Outline

• Lagrange Dual Function
• Lagrange Dual Problem
• Geometric Interpretation
• Saddle-point Interpretation
• **Optimality Conditions**
• Perturbation and Sensitivity Analysis
• Examples
• Generalized Inequalities
Complementary Slackness

assume strong duality holds, $x^*$ is primal optimal, $(\lambda^*, \nu^*)$ is dual optimal

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

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

$$\leq f_0(x^*)$$

hence the two inequalities hold with equality
- $x^*$ minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i \in [1 : m]$ (known as complementary slackness)

$$\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0, \quad f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$$
Karush-Kuhn-Tucker (KKT) Conditions

KKT conditions (for differentiable $f_i, h_i$):

1. primal constraints: $f_i(x) \leq 0, i \in [1 : m]$, $h_i(x) = 0, i \in [1 : p]$

2. dual constraints: $\lambda \succeq 0$

3. complementary slackness: $\lambda_i f_i(x) = 0, i \in [1 : m]$

4. gradient of Lagrangian w.r.t. $x$ vanishes

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

If strong duality holds and $x, \lambda, \nu$ are optimal, then they must satisfy the KKT condition
if $x, \lambda, \nu$ satisfy KKT for a convex problem, then they are optimal:
- from complementary slackness: $f_0(x) = L(x, \lambda, \nu)$
- from 4th condition (and convexity): $g(\lambda, \nu) = L(x, \lambda, \nu)$

hence, $f_0(x) = g(\lambda, \nu)$

if Slater’s condition is satisfied:
$x$ is optimal if and only if there exist $\lambda, \nu$ that satisfy KKT conditions
- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem
Example: Waterfilling

Consider the problem of allocating a total power of one to a set of \( n \) communication channels with noise levels \( \alpha_1, \alpha_2, \ldots, \alpha_n \). The goal is to maximize the total communication rate \( \sum_{i=1}^{n} \log(1 + x_i/\alpha_i) \), i.e.,

\[
\text{minimize } - \sum_{i=1}^{n} \log(\alpha_i + x_i)
\]

subject to \( x \geq 0, \mathbf{1}^T x = 1 \)

By KKT conditions, \( x \) is optimal iff \( x \geq 0, \mathbf{1}^T x = 1 \), and \( \exists \lambda, \nu \) such that

\[
\lambda \geq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu
\]

- if \( \nu < 1/\alpha_i \): \( \lambda_i = 0 \) and \( x_i = 1/\nu - \alpha_i \)
- if \( \nu \geq 1/\alpha_i \): \( \lambda_i = \nu - 1/\alpha_i \) and \( x_i = 0 \)
- determine \( \nu \) from \( \mathbf{1}^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1 \)

interpretation
- \( n \) patches; level of patch is at height \( \alpha_i \)
- flood area with unit amount of water
- resulting level is \( 1/\nu^* \)
Outline

• Lagrange Dual Function
• Lagrange Dual Problem
• Geometric Interpretation
• Saddle-point Interpretation
• Optimality Conditions
• **Perturbation and Sensitivity Analysis**
• Examples
• Generalized Inequalities
Perturbation and Sensitivity Analysis

unperturbed problem and its dual

\begin{align*}
\text{minimize} & \quad f_0(x) & \text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad f_i(x) \leq 0, i = 1, 2, \ldots, m & \text{subject to} & \quad \lambda \geq 0 \\
& \quad h_i(x) = 0, i = 1, 2, \ldots, p & & \\
\end{align*}

perturbed problem and its dual

\begin{align*}
\text{minimize} & \quad f_0(x) & \text{maximize} & \quad g(\lambda, \nu) - u^T \lambda - v^T \nu \\
\text{subject to} & \quad f_i(x) \leq u_i, i = 1, 2, \ldots, m & \text{subject to} & \quad \lambda \geq 0 \\
& \quad h_i(x) = v_i, i = 1, 2, \ldots, p & & \\
\end{align*}

- $x$ is primal variable; $u, v$ are parameters
- $p^*(u, v)$ is optimal value as a function of $u, v$
- what can we say about $p^*(u, v)$ based on the solution of the unperturbed problem and its dual?
Global Sensitivity Result

assume strong duality holds for unperturbed problem, and $\lambda^*, \nu^*$ are dual optimal for unperturbed problem. Apply weak duality to perturbed problem:

$$p^*(u, v) \geq g(\lambda^*, \nu^*) - u^T \lambda^* - v^T \nu^*$$

$$= p^*(0, 0) - u^T \lambda^* - v^T \nu^*$$

- $\lambda_i^*$ large: $p^*$ increases greatly if we tighten constraint $i$ (choose $u_i < 0$)
- $\lambda_i^*$ small: $p^*$ doesn’t decrease much if we loosen constraint $i$ (choose $u_i > 0$)
- $\nu_i^*$ large and positive: $p^*$ increases greatly if we take $v_i < 0$
- $\nu_i^*$ large and negative: $p^*$ increases greatly if we take $v_i > 0$
- $\nu_i^*$ small and positive: $p^*$ doesn’t decrease much if we take $v_i > 0$
- $\nu_i^*$ small and negative: $p^*$ doesn’t decrease much if we take $v_i < 0$
Local Sensitivity Result

if (in addition) $p^*(u,v)$ is differentiable at $(0,0)$, then

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i}, \quad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

This gives us a quantitative measure of how active a constraint is at the optimum $x^*$:

- $f_i(x^*) < 0$ and $\lambda_i^* = 0$ (complementary slackness): constraint $i$ is inactive and can be tightened or loosened a small amount without affecting the optimal value

- $f_i(x^*) = 0$ and $\lambda_i^*$ is small: constraint $i$ is active, but can be tightened or loosened a small amount without much effect on the optimal value

- $f_i(x^*) = 0$ and $\lambda_i^*$ is large: constraint $i$ is active, and loosening or tightening it a bit will have great effect on the optimal value
Outline

• Lagrange Dual Function
• Lagrange Dual Problem
• Geometric Interpretation
• Saddle-point Interpretation
• Optimality Conditions
• Perturbation and Sensitivity Analysis
• Examples
• Generalized Inequalities
Duality and Problem Reformulations

equivalent formulations of a problem can lead to very different duals

reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

common reformulations
- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions, e.g., replace \( f_0(x) \) by \( \phi(f_0(x)) \) with \( \phi \) convex, increasing
Introducing New Variables and Equality Constraints

minimize \( f_0(Ax + b) \)

- dual function is constant: \( g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^* \)
- we have strong duality, but dual is quite useless

reformulated problem and its dual

minimize \( f_0(y) \)  
subject to \( Ax + b - y = 0 \)

maximize \( b^T \nu - f_0^*(\nu) \) 
subject to \( A^T \nu = 0 \)

dual function follows from

\[
g(\nu) = \inf_{x,y} (f_0(y) - \nu^T y + \nu^T Ax + b^T \nu) \\
= \begin{cases} 
- f_0^*(\nu) + b^T \nu & A^T \nu = 0 \\
-\infty & \text{otherwise}
\end{cases}
\]
Implicit Constraints

LP with box constraints: primal and dual problem

minimize \( c^T x \) \hspace{1cm} \text{maximize} \hspace{1cm} -b^T \nu - 1^T \lambda_1 - 1^T \lambda_2

subject to \( Ax = b \) \hspace{1cm} \text{subject to} \hspace{1cm} c + A^T \nu + \lambda_1 - \lambda_2 = 0

\(-1 \leq x \leq 1\) \hspace{1cm} \lambda_1 \geq 0, \lambda_2 \geq 0

Reformulation with box constraints made implicit

\[
\begin{align*}
\text{minimize} \quad & f_0(x) = \begin{cases} 
  c^T x & -1 \leq x \leq 1 \\
  \infty & \text{otherwise}
\end{cases} \\
\text{subject to} \quad & Ax = b
\end{align*}
\]

dual function \( g(\nu) = \inf_{-1 \leq x \leq -1} (c^T x + \nu^T (Ax - b)) \)

\[
= -b^T \nu - \|A^T \nu + c\|_1
\]

dual problem: maximize \(-b^T \nu - \|A^T \nu + c\|_1\)
Outline

- Lagrange Dual Function
- Lagrange Dual Problem
- Geometric Interpretation
- Saddle-point Interpretation
- Optimality Conditions
- Perturbation and Sensitivity Analysis
- Examples
- Generalized Inequalities
Semidefinite Program

Primal SDP \((F_i, G \in S^k)\)

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

- Lagrange multiplier is matrix \(Z \in S^k\)
- Lagrangian \(L(x, Z) = c^T x + \text{tr}(Z(x_1 F_1 + \cdots + x_n F_n - G))\)
- dual function

\[
g(Z) = \inf_x L(x, Z) = \begin{cases} 
-\text{tr}(GZ) & \text{tr}(F_i Z) + c_i = 0, i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases}
\]

dual SDP:

\[
\begin{align*}
& \text{maximize} \quad -\text{tr}(GZ) \\
& \text{subject to} \quad Z \succeq 0, \text{tr}(F_i Z) + c_i = 0, i = 1, \ldots, n
\end{align*}
\]

\(p^* = d^*\) if primal SDP is strictly feasible \((\exists x \text{ with } x_1 F_1 + \cdots + x_n F_n < G)\)