Lecture 5: Duality and KKT Conditions

• Lagrange dual function
• Lagrange dual problem
• strong duality and Slater’s condition
• KKT optimality conditions
• sensitivity analysis
• generalized inequalities
Lagrangian

standard form problem, (for now) we don't assume convexity

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

• optimal value \( p^* \), domain \( D \)
• called **primal problem** (in context of duality)

**Lagrangian** \( L : \mathbb{R}^{n+m} \rightarrow \mathbb{R} \)

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

• \( \lambda_i \geq 0 \) and \( \nu_i \) called **Lagrange multipliers** or **dual variables**
• objective is **augmented** with weighted sum of constraint functions
Lagrange dual function

(Lagrange) dual function $g : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$

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

- minimum of augmented cost as function of weights
- can be $-\infty$ for some $\lambda$ and $\nu$
- $g$ is concave (even if $f_i$ not convex!)

**example:** LP

minimize $c^T x$
subject to $a_i^T x - b_i \leq 0, \ i = 1, \ldots, m$

Note that $L(x, \lambda) = c^T x + \sum_{i=1}^m \lambda_i (a_i^T x - b_i) = -b^T \lambda + (A^T \lambda + c)^T x$

hence $g(\lambda) = \begin{cases} 
- b^T \lambda & \text{if } A^T \lambda + c = 0 \\
- \infty & \text{otherwise}
\end{cases}$
Lower bound property

if \( x \) is primal feasible, then
\[
g(\lambda, \nu) \leq f_0(x)
\]

**proof:** if \( f_i(x) \leq 0 \) and \( \lambda_i \geq 0 \),
\[
f_0(x) \geq f_0(x) + \sum \lambda_i f_i(x) + \sum \nu_i h_i(x) \geq \inf_z \left( f_0(z) + \sum \lambda_i f_i(z) + \sum \nu_i h_i(z) \right) = g(\lambda, \nu)
\]
\[
f_0(x) - g(\lambda, \nu)
\]
is called the **duality gap**

minimize over primal feasible \( x \) to get, for any \( \lambda \succeq 0 \) and \( \nu \),
\[
g(\lambda, \nu) \leq p^*
\]

\( \lambda \in \mathbb{R}^m \) and \( \nu \in \mathbb{R}^p \) are **dual feasible** if \( \lambda \succeq 0 \) and \( g(\lambda, \nu) > -\infty \)

dual feasible points yield lower bounds on optimal value!
Lagrange dual problem

let’s find best lower bound on $p^*$:

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

- called (Lagrange) dual problem
  (associated with primal problem)
- always a convex problem, even if primal isn’t!
- optimal value denoted $d^*$
- we always have $d^* \leq p^*$ (called weak duality)
- $p^* - d^*$ is optimal duality gap
Strong duality

for convex problems, we (usually) have strong duality:

\[ d^* = p^* \]

when strong duality holds, dual optimal \( \lambda^* \) serves as certificate of optimality for primal optimal point \( x^* \)

many conditions or constraint qualifications guarantee strong duality for convex problems

**Slater’s condition:** if primal problem is strictly feasible (and convex), \( i.e., \) there exists \( x \in \text{relint} \ D \) with

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

\[ h_i(x) = 0, \ i = 1, \ldots, p \]

then we have \( p^* = d^* \)
Dual of linear program

(primal) LP

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

- \(n\) variables, \(m\) inequality constraints

dual of LP is (after making implicit equality constraints explicit)

\[
\begin{align*}
\text{maximize} & \quad -b^T \lambda \\
\text{subject to} & \quad A^T \lambda + c = 0 \\
& \quad \lambda \succeq 0
\end{align*}
\]

- dual of LP is also an LP (indeed, in std LP format)
- \(m\) variables, \(n\) equality constraints, \(m\) nonnegativity contraints

for LP we have strong duality except in one (pathological) case: primal and dual both infeasible \((p^* = +\infty, d^* = -\infty)\)
(primal) QP

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

we assume \( P \succ 0 \) for simplicity  

Lagrangian is \( L(x, \lambda) = x^T P x + \lambda^T (Ax - b) \)

\( \nabla_x L(x, \lambda) = 0 \) yields \( x = -(1/2) P^{-1} A^T \lambda \), hence dual function is 

\[
g(\lambda) = -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda
\]

• concave quadratic function 
• all \( \lambda \succeq 0 \) are dual feasible

dual of QP is

\[
\begin{align*}
& \text{maximize} & & -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\
& \text{subject to} & & \lambda \succeq 0
\end{align*}
\]

... another QP
Equality constrained least-squares

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

\(A\) is fat, full rank (solution is \(x^* = A^T (AA^T)^{-1}b\))

dual function is

\[
g(\nu) = \inf_x \left( x^T x + \nu^T (Ax - b) \right) = -\frac{1}{4} \nu^T AA^T \nu - b^T \nu
\]

dual problem is

\[
\text{maximize} \quad -\frac{1}{4} \nu^T AA^T \nu - b^T \nu
\]

solution: \(\nu^* = -2( AA^T )^{-1} b \)

can check \(d^* = p^*\)
Introducing equality constraints

**idea:** simple transformation of primal problem can lead to very different dual

**example:** unconstrained geometric programming

**primal problem:**

\[
\text{minimize} \quad \log \sum_{i=1}^{m} \exp(a_i^T x - b_i)
\]

dual function is constant \( g = p^* \) (we have strong duality, but it’s useless)

now **rewrite primal problem** as

\[
\text{minimize} \quad \log \sum_{i=1}^{m} \exp y_i
\]

subject to \( y = Ax - b \)
let us introduce

- \( m \) new variables \( y_1, \ldots, y_m \)
- \( m \) new equality constraints \( y = Ax - b \)

**dual function**

\[
g(\nu) = \inf_{x,y} \left( \log \sum_{i=1}^{m} \exp y_i + \nu^T (Ax - b - y) \right)
\]

- infimum is \(-\infty\) if \( A^T \nu \neq 0 \)
- assuming \( A^T \nu = 0 \), let’s minimize over \( y \):

\[
\frac{e^{y_i}}{\sum_{j=1}^{m} e^{y_j}} = \nu_i
\]

solvable iff \( \nu_i > 0 \), \( 1^T \nu = 1 \)

\[
g(\nu) = -\sum_i \nu_i \log \nu_i - b^T \nu
\]
• same expression if $\nu \succeq 0$, $1^T \nu = 1$ ($0 \log 0 = 0$)

dual problem

maximize $-b^T \nu - \sum_i \nu_i \log \nu_i$

subject to $1^T \nu = 1$, $(\nu \succeq 0)$
$A^T \nu = 0$

moral: trivial reformulation can yield different dual
Duality in algorithms

many algorithms produce at iteration $k$:

- a primal feasible $x^{(k)}$
- a dual feasible $\lambda^{(k)}$ and $\nu^{(k)}$

with $f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \to 0$ as $k \to \infty$

hence at iteration $k$ we know $p^* \in \left[ g(\lambda^{(k)}, \nu^{(k)}), f_0(x^{(k)}) \right]$

- useful for stopping criteria
- algorithms that use dual solution are often more efficient (e.g., LP)
Nonheuristic stopping criteria

**absolute error** = \( f_0(x^{(k)}) - p^* \leq \epsilon \)

stopping criterion: until \( f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon \)

**relative error** = \( \frac{f_0(x^{(k)}) - p^*}{|p^*|} \leq \epsilon \)

stopping criterion:

until \( g(\lambda^{(k)}, \nu^{(k)}) > 0 \ \& \ \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{g(\lambda^{(k)}, \nu^{(k)})} \leq \epsilon \) or \( f_0(x^{(k)}) < 0 \ \& \ \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{-f_0(x^{(k)})} \leq \epsilon \)

achieve target value \( \ell \) or, prove \( \ell \) is unachievable

(\( i.e. \), determine either \( p^* \leq \ell \) or \( p^* > \ell \))

stopping criterion: until \( f_0(x^{(k)}) \leq \ell \) or \( g(\lambda^{(k)}, \nu^{(k)}) > \ell \)
suppose \( x^*, \lambda^*, \) and \( \nu^* \) are primal, dual feasible with zero duality gap (hence, they are primal, dual optimal)

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

hence we have \( \sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0 \), and so

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

- called **complementary slackness** condition
- \( i \)th constraint inactive at optimum \( \implies \lambda_i = 0 \)
- \( \lambda_i^* > 0 \) at optimum \( \implies i \)th constraint active at optimum
KKT optimality conditions

suppose

- $f_i$ are differentiable
- $x^*, \lambda^*$ are (primal, dual) optimal, with zero duality gap

by complementary slackness we have

$$f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) = \inf_x \left( f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* f_i(x) \right)$$

i.e., $x^*$ minimizes $L(x, \lambda^*, \nu^*)$

therefore

$$\nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) + \sum_i \nu_i^* \nabla h_i(x^*) = 0$$
so if $x^{*}$, $\lambda^{*}$, and $\nu^{*}$ are (primal, dual) optimal, with zero duality gap, they satisfy

\[ f_i(x^{*}) \leq 0 \]
\[ h_i(x^{*}) = 0 \]
\[ \lambda_i^{*} \geq 0 \]
\[ \lambda_i^{*} f_i(x^{*}) = 0 \]
\[ \nabla f_0(x^{*}) + \sum_i \lambda_i^{*} \nabla f_i(x^{*}) + \sum_i \nu_i^{*} \nabla h_i(x^{*}) = 0 \]

the Karush-Kuhn-Tucker (KKT) optimality conditions

conversely, if the problem is convex and $x^{*}$, $\lambda^{*}$ satisfy KKT, then they are (primal, dual) optimal
Geometric interpretation of duality

consider set

\[ \mathcal{A} = \{ (u, t) \in \mathbb{R}^{m+1} \mid \exists x \ f_i(x) \leq u_i, \ f_0(x) \leq t \} \]

• \( \mathcal{A} \) is convex if \( f_i \) are
• for \( \lambda \succeq 0 \),

\[
g(\lambda) = \inf \left\{ \left[ \begin{array}{c} \lambda \\ 1 \end{array} \right]^T \left[ \begin{array}{c} u \\ t \end{array} \right] \mid \left[ \begin{array}{c} u \\ t \end{array} \right] \in \mathcal{A} \right\}
\]
(Idea of) proof of Slater’s theorem

Problem convex, strictly feasible $\implies$ strong duality

- $(0, p^*) \in \partial A \implies \exists$ supporting hyperplane at $(0, p^*)$:
  $$ (u, t) \in A \implies \mu_0(t - p^*) + \mu^T u \geq 0 $$
- $\mu_0 \geq 0, \mu \succeq 0, (\mu, \mu_0) \neq 0$
- Strong duality $\iff \exists$ supporting hyperplane with $\mu_0 > 0$: for $\lambda^* = \mu / \mu_0$, we have
  $$ p^* \leq t + \lambda^{*T} u \quad \forall (t, u) \in A, \quad p^* \leq g(\lambda^*) $$
- Slater’s condition: there exists $(u, t) \in A$ with $u < 0$; implies that all supporting hyperplanes at $(0, p^*)$ are non-vertical ($\mu_0 > 0$)
Sensitivity analysis via duality

define $p^*(u)$ as the optimal value of

$$\begin{align*}
\text{minimize} & \quad f_0(x), \\
\text{subject to} & \quad f_i(x) \leq u_i, \quad i = 1, \ldots, m
\end{align*}$$

$\lambda^*$ gives lower bound on $p^*(u)$: $p^*(u) \geq p^* - \sum_{i=1}^{m} \lambda^*_i u_i$

- if $\lambda^*_i$ large: $u_i < 0$ greatly increases $p^*$
- if $\lambda^*_i$ small: $u_i > 0$ does not decrease $p^*$ too much

if $p^*(u)$ is differentiable, $\lambda^*_i = -\left. \frac{\partial p^*(0)}{\partial u_i} \right|_{u=0},$ $\lambda^*_i$ is sensitivity of $p^*$ w.r.t. $i$th constraint
**Generalized inequalities**

minimize \( f_0(x) \)
subject to \( f_i(x) \leq_{K_i} 0, \ i = 1, \ldots, L \)

- \( \leq_{K_i} \) are generalized inequalities on \( \mathbb{R}^{m_i} \)
- \( f_i : \mathbb{R}^n \to \mathbb{R}^{m_i} \) are \( K_i \)-convex

**Lagrangian** \( L : \mathbb{R}^n \times \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_L} \to \mathbb{R}, \)

\[
L(x, \lambda_1, \ldots, \lambda_L) = f_0(x) + \lambda_1^T f_1(x) + \cdots + \lambda_L^T f_L(x)
\]

**Dual function**

\[
g(\lambda_1, \ldots, \lambda_L) = \inf_x \left( f_0(x) + \lambda_1^T f_1(x) + \cdots + \lambda_L^T f_L(x) \right)
\]

\( \lambda_i \) **dual feasible** if \( \lambda_i \succeq_{K_i^*} 0, \ g(\lambda_1, \ldots, \lambda_L) > -\infty \)
**lower bound property:** if $x$ primal feasible and $(\lambda_1, \ldots, \lambda_L)$ is dual feasible, then

$$g(\lambda_1, \ldots, \lambda_L) \leq f_0(x)$$

(hence, $g(\lambda_1, \ldots, \lambda_L) \leq p^*$)

**dual problem**

maximize $g(\lambda_1, \ldots, \lambda_L)$

subject to $\lambda_i \preceq K_0^i, \ i = 1, \ldots, L$

**weak duality:** $d^* \leq p^*$ always

**strong duality:** $d^* = p^*$ usually

**Slater condition:** if primal is strictly feasible, i.e.,

$$\exists x \in \text{relint } D : f_i(x) \prec K_i^0, \ i = 1, \ldots, L$$

then $d^* = p^*$
Example: semidefinite programming

minimize \( c^T x \)
subject to \( F_0 + x_1 F_1 + \cdots + x_n F_n \preceq 0 \)

Lagrangian (multiplier \( Z \succeq 0 \))

\[
L(x, Z) = c^T x + \text{Tr } Z(F_0 + x_1 F_1 + \cdots + x_n F_n)
\]

dual function

\[
g(Z) = \inf_x \left( c^T x + \text{Tr } Z(F_0 + x_1 F_1 + \cdots + x_n F_n) \right)
\]

\[
g(Z) = \begin{cases} 
\text{Tr } F_0 Z & \text{if } \text{Tr } F_i Z + c_i = 0, \ i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases}
\]

dual problem

maximize \( \text{Tr } F_0 Z \)
subject to \( \text{Tr } F_i Z + c_i = 0, \ i = 1, \ldots, n \)
\( Z = Z^T \succeq 0 \)

strong duality holds if there exists \( x \) with \( F_0 + x_1 F_1 + \cdots + x_n F_n < 0 \)