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

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{array}$$

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

Lagrangian $L : \mathbf{R}^{n+m} \to \mathbf{R}$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

- $\lambda_i \geq 0$ and ν_i called *Lagrange multipliers* or *dual variables*
- objective is *augmented* with weighted sum of constraint functions

Lagrange dual function

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

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

- minimum of augmented cost as function of weights
- can be $-\infty$ for some λ and ν
- g is concave (even if f_i not convex and h_j no linear!)

example: LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x - b_i \leq 0, \; i = 1, \ldots, m \end{array}$$

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 $f_i(x) \leq 0$ $h_j(x) = 0$, and $\lambda_i \geq 0$,

$$f_0(x) \geq f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \nu_j h_j(x) \geq \inf_z \left(f_0(z) + \sum_i \lambda_i f_i(z) + \sum_j \nu_j h_j(z) \right) = g(\lambda, \nu)$$

$f_0(x) - g(\lambda, \nu)$ is called the **duality gap**

minimize over primal feasible x to get, for any $\lambda \succeq 0$ and ν,

$$g(\lambda, \nu) \leq p^\star$$

$\lambda \in \mathbf{R}^m$ and $\nu \in \mathbf{R}^p$ 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 $p^\star$:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

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

Strong duality

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

$$d^\star = p^\star$$

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

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 $x \in \mathbf{relint}\, D$ with

$$f_i(x) < 0, \ i = 1, \ldots, m$$
$$h_i(x) = 0, \ i = 1, \ldots, p$$

then we have $p^\star = d^\star$

Dual of linear program

(primal) LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

- n variables, m inequality constraints

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

$$\begin{array}{ll} \text{maximize} & -b^T \lambda \\ \text{subject to} & A^T \lambda + c = 0 \\ & \lambda \succeq 0 \end{array}$$

- 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 ($p^\star = +\infty$, $d^\star = -\infty$)

Dual of quadratic program

(primal) QP

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \preceq b \end{array}$$

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

$$\begin{array}{ll} \text{maximize} & -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

. . . another QP

Equality constrained least-squares

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

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

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

$$\text{maximize} \quad -\tfrac{1}{4}\nu^T AA^T \nu - b^T \nu$$

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

can check $d^\star = p^\star$

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 $g = p^\star$ (we have strong duality, but it's useless)

now **rewrite primal problem** as

$$\begin{array}{ll} \text{minimize} & \log \sum_{i=1}^{m} \exp y_i \\ \text{subject to} & y = Ax - b \end{array}$$

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 $\nu_i > 0$, $\mathbf{1}^T\nu = 1$

$$g(\nu) = -\sum_i \nu_i \log \nu_i - b^T\nu$$

- same expression if $\nu \succeq 0$, $\mathbf{1}^T\nu = 1$ $(0 \log 0 = 0)$

dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\nu - \sum_i \nu_i \log \nu_i \\ \text{subject to} & \mathbf{1}^T\nu = 1, \quad (\nu \succeq 0) \\ & A^T\nu = 0 \end{array}$$

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** $p^\star \in \left[g(\lambda^{(k)}, \nu^{(k)}), f_0(x^{(k)})\right]$

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

Nonheuristic stopping criteria

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

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

relative error $= \dfrac{f_0(x^{(k)}) - p^\star}{|p^\star|} \leq \epsilon$

stopping criterion:

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

achieve **target value** ℓ or, prove ℓ is unachievable
($i.e.$, determine either $p^\star \leq \ell$ or $p^\star > \ell$)

stopping criterion: **until** $\left(f_0(x^{(k)}) \leq \ell\right.$ **or** $\left.g(\lambda^{(k)}, \nu^{(k)}) > \ell\right)$

Complementary slackness

suppose $x^\star$, $\lambda^\star$, and $\nu^\star$ are primal, dual feasible with zero duality gap (hence, they are primal, dual optimal)

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

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

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \ldots, m$$

- called **complementary slackness** condition
- ith constraint inactive at optimum $\Longrightarrow \lambda_i = 0$
- $\lambda_i^\star > 0$ at optimum $\Longrightarrow$ ith constraint active at optimum

KKT optimality conditions

suppose

- f_i and h_i are differentiable
- $x^\star$, $\lambda^\star$, and $\nu^\star$ are (primal, dual) optimal, with zero duality gap

by complementary slackness we have

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

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

therefore

$$\nabla f_0(x^\star) + \sum_i \lambda_i^\star \nabla f_i(x^\star) + \sum_i \nu_i^\star \nabla h_i(x^\star) = 0$$

so if $x^\star$, $\lambda^\star$, and $\nu^\star$ are (primal, dual) optimal, with zero duality gap, they satisfy

$$\begin{aligned} f_i(x^\star) &\leq 0 \\ h_i(x^\star) &= 0 \\ \lambda_i^\star &\geq 0 \\ \lambda_i^\star f_i(x^\star) &= 0 \\ \nabla f_0(x^\star) + \textstyle\sum_i \lambda_i^\star \nabla f_i(x^\star) + \sum_i \nu_i^\star \nabla h_i(x^\star) &= 0 \end{aligned}$$

the **Karush-Kuhn-Tucker** (KKT) optimality conditions

conversely, if the problem is convex and $x^\star$, $\lambda^\star$ satisfy KKT, then they are (primal, dual) optimal

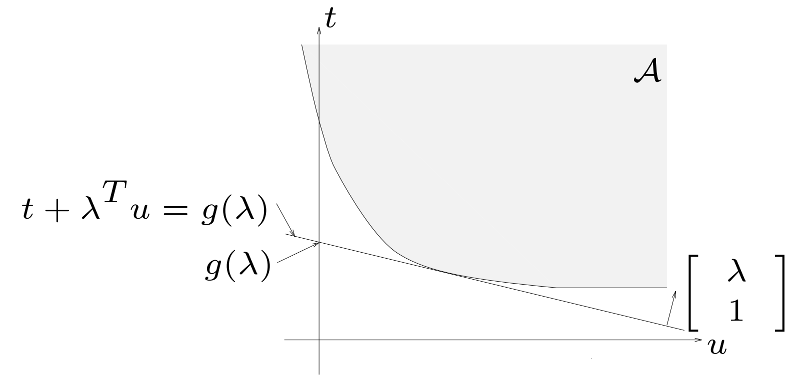
Geometric interpretation of duality

consider set

$$\mathcal{A} = \{\ (u,t) \in \mathbf{R}^{m+1} \mid \exists x\ f_i(x) \le u_i, \quad f_0(x) \le t\ \}$$

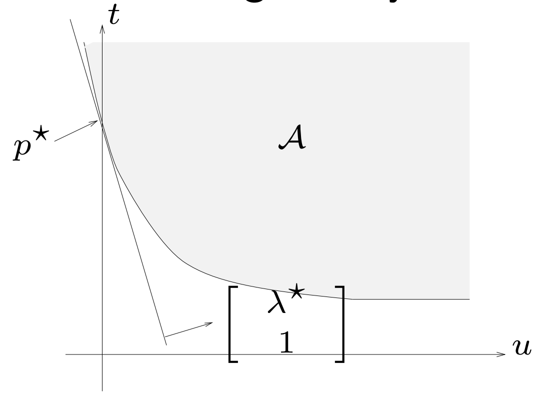
- $\mathcal{A}$ is convex if f_i are
- for $\lambda \succeq 0$,

$$g(\lambda) = \inf \left\{ \left[\begin{array}{c} \lambda \\ 1 \end{array} \right]^T \left[\begin{array}{c} u \\ t \end{array} \right] \ \middle|\ \left[\begin{array}{c} u \\ t \end{array} \right] \in \mathcal{A} \right\}$$



(Idea of) proof of Slater's theorem

problem convex, strictly feasible $\Longrightarrow$ strong duality



- $(0, p^\star) \in \partial\mathcal{A} \Rightarrow \exists$ supporting hyperplane at $(0, p^\star)$:

$$(u, t) \in \mathcal{A} \Longrightarrow \mu_0(t - p^\star) + \mu^T u \geq 0$$

- $\mu_0 \geq 0$, $\mu \succeq 0$, $(\mu, \mu_0) \neq 0$
- strong duality $\Leftrightarrow \exists$ supporting hyperplane with $\mu_0 > 0$: for $\lambda^\star = \mu/\mu_0$, we have

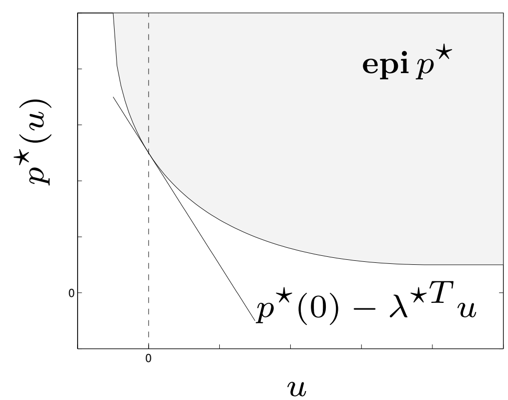
$$p^\star \leq t + \lambda^{\star T} u \;\; \forall (t, u) \in \mathcal{A}, \;\; p^\star \leq g(\lambda^\star)$$

- Slater's condition: there exists $(u, t) \in \mathcal{A}$ with $u \prec 0$; implies that all supporting hyperplanes at $(0, p^\star)$ are non-vertical ($\mu_0 > 0$)

Sensitivity analysis via duality

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

$$\text{minimize} \quad f_0(x), \quad \text{subject to} \quad f_i(x) \le u_i, \quad i = 1, \ldots, m$$



$\lambda^\star$ gives lower bound on $p^\star(u)$: $p^\star(u) \ge p^\star - \sum_{i=1}^m \lambda_i^\star u_i$

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

if $p^\star(u)$ is differentiable, $\lambda_i^\star = -\dfrac{\partial p^\star(0)}{\partial u_i}$, $\lambda_i^\star$ is sensitivity of $p^\star$ w.r.t. ith constraint

Generalized inequalities

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, L \end{array}$$

- $\preceq_{K_i}$ are generalized inequalities on $\mathbf{R}^{m_i}$
- $f_i : \mathbf{R}^n \to \mathbf{R}^{m_i}$ are K_i-convex

Lagrangian $L : \mathbf{R}^n \times \mathbf{R}^{m_1} \times \cdots \times \mathbf{R}^{m_L} \to \mathbf{R}$,

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

dual function

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

λ_i **dual feasible** if $\lambda_i \succeq_{K_i^\star} 0$, $g(\lambda_1, \ldots, \lambda_L) > -\infty$

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

dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_L) \\ \text{subject to} & \lambda_i \succeq_{K_i^\star} 0, \quad i = 1, \ldots, L \end{array}$$

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

strong duality: $d^\star = p^\star$ usually

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

$$\exists x \in \mathbf{relint}\, D : \quad f_i(x) \prec_{K_i} 0, \quad i = 1, \ldots, L$$

then $d^\star = p^\star$

Example: semidefinite programming

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F_0 + x_1 F_1 + \cdots + x_n F_n \preceq 0 \end{array}$$

Lagrangian (multiplier $Z \succeq 0$)

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

dual function

$$\begin{aligned} g(Z) &= \inf_x \left(c^T x + \mathbf{Tr}\, Z(F_0 + x_1 F_1 + \cdots + x_n F_n) \right) \\ &= \begin{cases} \mathbf{Tr}\, F_0 Z & \text{if } \mathbf{Tr}\, F_i Z + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

dual problem

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}\, F_0 Z \\ \text{subject to} & \mathbf{Tr}\, F_i Z + c_i = 0, \quad i = 1, \ldots, n \\ & Z = Z^T \succeq 0 \end{array}$$

strong duality holds if there exists x with $F_0 + x_1 F_1 + \cdots + x_n F_n \prec 0$