Lecture 7
Duality

- Lagrange dual function
- Lagrange dual problem
- strong duality and Slater's condition
- KKT optimality conditions
- sensitivity analysis
- generalized inequalities

Lagrangian

std form problem

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

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

(for now) we assume

- not necessarily convex
- no equality constraints
- $\mathbf{dom}\, f_i = \mathbf{R}^n$

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

$$L(x, \lambda) = f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)$$

- λ_i called *Lagrange multipliers* or *dual variables*
- objective is *augmented* with weighted sum of constraint fcts

Lagrange dual function

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

$$\begin{aligned} g(\lambda) &= \inf_x L(x, \lambda) \\ &= \inf_x \left(f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)\right) \end{aligned}$$

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

example: LP

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

$$\begin{aligned} L(x, \lambda) &= c^T x + \sum_i \lambda_i (a_i^T x - b_i) \\ &= -b^T \lambda + (A^T \lambda + c)^T x \end{aligned}$$

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

Lower bound property

if $\lambda \succeq 0$ and x is primal feasible, then

$$g(\lambda) \leq f_0(x)$$

proof: if $f_i(x) \leq 0$ and $\lambda_i \geq 0$,

$$\begin{aligned} f_0(x) &\geq f_0(x) + \sum_i \lambda_i f_i(x) \\ &\geq \inf_z \left(f_0(z) + \sum_i \lambda_i f_i(z) \right) \\ &= g(\lambda) \end{aligned}$$

$f_0(x) - g(\lambda)$ is called the *duality gap* of (primal feasible) x and $\lambda \succeq 0$

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

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

$\lambda \in \mathbf{R}^m$ is *dual feasible* if $\lambda \succeq 0$ and $g(\lambda) > -\infty$

dual feasible points yield lower bounds on optimal value!

Lagrange dual problem

let's find **best** lower bound on $p^\star$:

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

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

strong duality: for convex problems we (usually) have $d^\star = p^\star$

- hence, duality is especially important and useful in convex optimization
- strong duality does not hold, in general, for nonconvex problems

Implications of strong duality:

- dual optimal $\lambda^\star$ serves as **certificate of optimality** for primal optimal point $x^\star$

- can solve *constrained* problem

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

  by solving *unconstrained* problem

$$\text{minimize}\ \ f_0(x) + \lambda_1^\star f_1(x) + \cdots + \lambda_m^\star f_m(x)$$

- can express strong duality in symmetric form

$$d^\star = \sup_{\lambda \succeq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \succeq 0} L(x, \lambda) = p^\star$$

  $i.e.$, strong duality $\Longrightarrow$ can swap $\inf$ & $\sup$

many conditions or *constraint qualifications* guarantee strong duality for convex problems

Slater's condition: if primal problem is strictly feasible (and convex) then we have $p^\star = d^\star$

Dual of LP

(primal) LP

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

- n vbles, m inequality constraints

dual of LP is

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

- dual of LP is also an LP (indeed, in std LP format)
- m vbles, 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 QP

(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

Duality in algorithms

many algorithms produce at iteration k

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

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

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

- 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:

$$\textbf{until} \ \left(f_0(x^{(k)}) - g(\lambda^{(k)}) \leq \epsilon\right)$$

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

stopping criterion:

$$\begin{array}{rl} \textbf{until} & \left(g(\lambda^{(k)}) > 0 \ \ \& \ \ \frac{f_0(x^{(k)}) - g(\lambda^{(k)})}{g(\lambda^{(k)})} \leq \epsilon\right) \\ \textbf{or} & \left(f_0(x^{(k)}) < 0 \ \ \& \ \ \frac{f_0(x^{(k)}) - g(\lambda^{(k)})}{f_0(x^{(k)})} \leq \epsilon\right) \end{array}$$

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

stopping criterion:

$$\textbf{until} \ \left(f_0(x^{(k)}) \leq \ell \ \textbf{or} \ g(\lambda^{(k)}) > \ell\right)$$

Complementary slackness

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

$$\begin{array}{rcl} f_0(x^\star) & = & g(\lambda^\star) \\ & = & \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) \right) \\ & \leq & f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) \end{array}$$

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 are differentiable, $x^\star$, $\lambda^\star$ are primal, dual optimal

then we have

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

the Karush-Kuhn-Tucker (KKT) optimality conditions

conversely, any $x^\star$, $\lambda^\star$ that satisfy KKT are primal, dual optimal

for convex problems, KKT are necessary and sufficient optimality conditions, provided

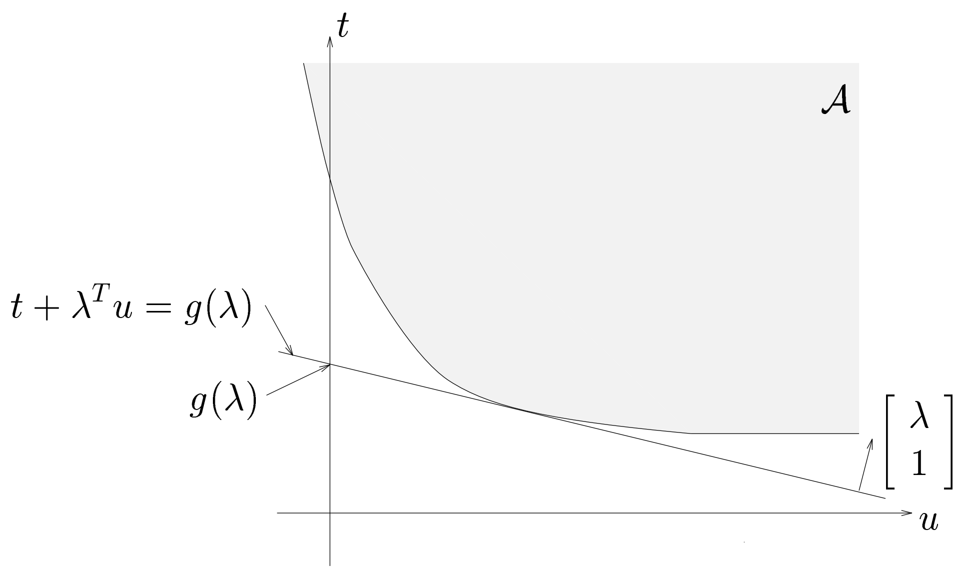
- strong duality holds
- primal & dual optima are attained

Geometric interpretation of dual problem

consider set

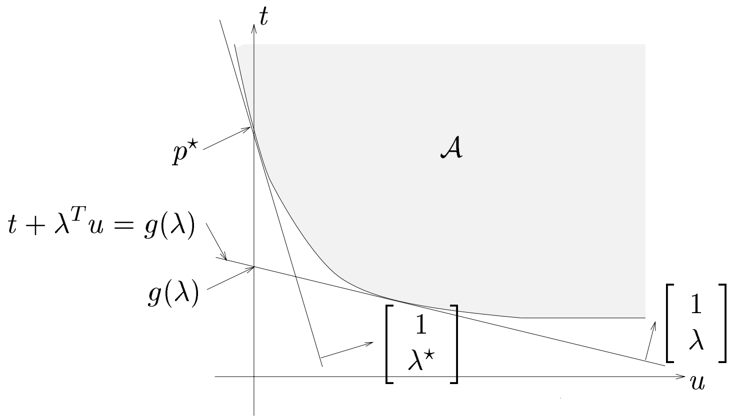
$$\mathcal{A} = \{\ (u,t) \in \mathbf{R}^{m+1} \mid \exists x\ f_i(x) \le u_i, \quad f_0(x) \le t\ \}$$

- $\mathcal{A}$ convex if f_i are
- $g(\lambda) = \inf \left\{ \begin{bmatrix} \lambda \\ 1 \end{bmatrix}^T \begin{bmatrix} u \\ t \end{bmatrix} \;\middle|\; \begin{bmatrix} u \\ t \end{bmatrix} \in \mathcal{A} \right\}$



(Idea of) proof

problem convex, strictly feasible $\Longrightarrow$ strong duality



- $(0, p^\star) \in \partial \mathcal{A}$
- hence $\exists$ supporting hyperplane to $\mathcal{A}$ 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 $\Longleftrightarrow$ $\exists$ supp. hyperplane with $\mu_0 > 0$: for $\lambda^\star = \mu/\mu_0$, we have

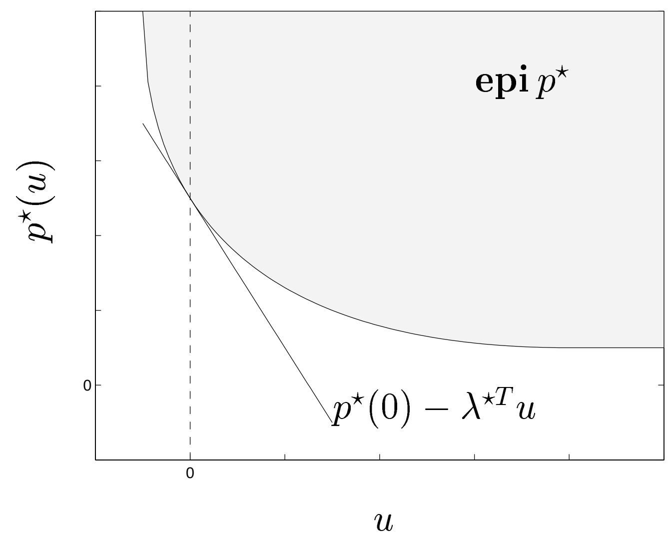
$$\begin{aligned} p^\star &\leq t + \lambda^{\star T} u \;\; \forall (t, u) \in \mathcal{A} \\ p^\star &\leq g(\lambda^\star) \end{aligned}$$

- 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

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



$\lambda^\star$ 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

Equality constraints

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & g_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

optimal value $p^\star$

define **Lagrangian** $L : \mathbf{R}^{n+m+p} \to \mathbf{R}$ as

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

dual function is $g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$

(λ, ν) is dual feasible if $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$
(no sign condition on ν)

lower bound property: if x is primal feasible and (λ, ν) is dual feasible, then $g(\lambda, \nu) \le f_0(x)$

hence, $g(\lambda, \nu) \le p^\star$

dual problem: find best lower bound

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

(note ν unconstrained) optimal value $d^\star$

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

strong duality: if primal is convex then (usually) $d^\star = p^\star$

Slater condition: if primal is strictly feasible (and convex) then $d^\star = p^\star$

KKT conditions:

$$\begin{aligned} f_i(\tilde{x}) &\leq 0 \\ g_i(\tilde{x}) &= 0 \\ \tilde{\lambda}_i &\geq 0 \\ \nabla f_0(\tilde{x}) + \sum_i \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_i \tilde{\nu}_i \nabla g_i(\tilde{x}) &= 0 \\ \tilde{\lambda}_i f_i(\tilde{x}) &= 0 \end{aligned}$$

example: opt cond. for equality constraints only

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b \end{array}$$

$x^\star$ optimal if and only if $\exists \nu^\star$ s.t.

$$\nabla f_0(x^\star) + A^T \nu^\star = 0$$

Example: equality constrained least-squares

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

A is fat, full rank
(soln 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

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

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

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

Example: geometric programming

simple (unconstrained) case

primal problem:

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

dual fct is constant $g = p^\star$

(we have strong duality, but it's useless)

now **rewrite primal problem** as

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

- introduce m new vbles $y_1, \ldots, y_m$
- introduce 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:

$$\exp y_i \Big/ \sum_{j=1}^{n} \exp y_j = \nu_i$$

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

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

dual problem

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

we have strong duality

connection between primal GP and dual entropy problem:

- useful
- not obvious

moral: apparently trivial reformulations of primal yield different duals

Generalized inequalities

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

where

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

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

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

dual function

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

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 $(\lambda_1, \ldots, \lambda_m)$ 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.*,

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

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

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

$Z = Z^T \in \mathbf{R}^{m \times m}$

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

$$F_0 + x_1 F_1 + \cdots + x_n F_n \prec 0$$

Theorem of alternatives

1. there exist x with $f_i(x) < 0$, $i = 1, \ldots, m$
2. there exist $\lambda \neq 0$ with $\lambda \succeq 0$,

$$g(\lambda) = \inf_x \left(\lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)\right) \geq 0$$

- exactly one of these is true
- called **alternatives**
- use in practice: λ that satisfies 2nd condition proves $f_i(x) < 0$ is infeasible

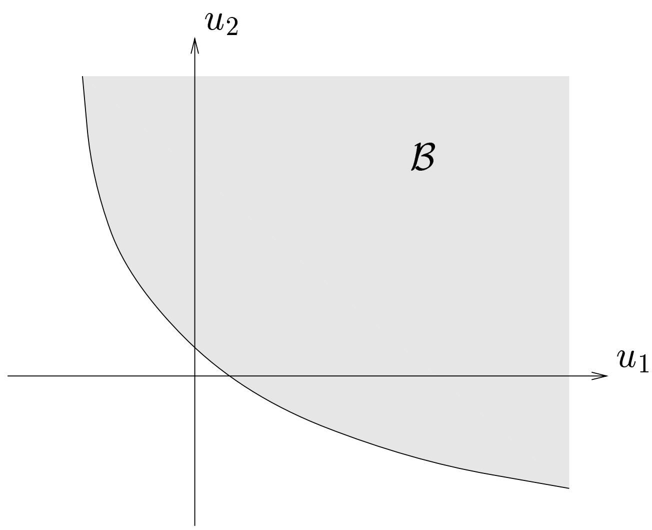
proof

$1 \Rightarrow \neg 2$: by contradiction

$$f_i(x) < 0,\ 0 \neq \lambda \succeq 0 \Longrightarrow \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x) < 0$$

$\neg 1 \Rightarrow 2$:

define $\mathcal{B} = \{u \in \mathbf{R}^m \mid \exists x: \quad f_i(x) \leq u_i\}$



- $\neg 1 \Longleftrightarrow \mathcal{B} \cap \{u \mid u \prec 0\} = \emptyset$
- hence, exists separating hyperplane: $\lambda \neq 0$,

$$u \in \mathcal{B} \Longrightarrow \lambda^T u \geq 0$$

$$u \prec 0 \Longrightarrow \lambda^T u \leq 0$$

- implies $\lambda \succeq 0$ and

$$\lambda_1 f_1(x) + \cdots + \lambda_n f_n(x) \geq 0$$

for all x