5. Duality

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions

- examples
- generalized inequalities

Lagrangian

standard form problem (not necessarily convex)

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

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

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

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

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

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

x~ feasible when f_i(x~) ≤ 0 and h_i(x~) = 0, also λ_i ≥ 0
then Σ_i λ_i f_i(x~) + Σ_i v_i h_i(x~) ≤ 0

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda, \nu)$

Least-norm solution of linear equations

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

Ax - b = 0

dual function

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2) A^T \nu$$

- plug in in L to obtain g:

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$$

a concave function of ν

lower bound property: $p^\star \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

Ax - b = 0   - x ≤ 0

dual function

- Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

- L is affine in x, hence

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

for any nonzero vector y, we can make y'x arbitrarily small

g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

lower bound property: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

Recall A'v - λ + c = 0
Then A'v + c = λ
But λ ≥ 0
Then A'v + c ≥ 0

Equality constrained norm minimization

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

- A x + b = 0

dual function

$$g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T\nu) = \begin{cases} b^T\nu & \|A^T\nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

= b'v + inf_x(|x| - v'Ax)

where $\|v\|_* = \sup_{\|u\|\leq 1} u^T v$ is dual norm of $\| \cdot \|$

Let y = A'v, proof: follows from $\inf_x(\|x\| - y^T x) = 0$ if $\|y\|_* \leq 1$, $-\infty$ otherwise

- if $\|y\|_* \leq 1$, then $\|x\| - y^T x \geq 0$ for all x, with equality if $x = 0$

  Cauchy-Schwarz: y'x ≤ |y|_* |x| ≤ |x|

- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \leq 1$, $u^T y = \|y\|_* > 1$:

  since |y|_* = sup_{|u|≤1} u'y > 1

  |x| - y'x = t |u| - t y'u = t |u| - t |y|_*

$$= t(\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty$$

< 0

lower bound property: $p^\star \geq b^T\nu$ if $\|A^T\nu\|_* \leq 1$

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

x_i is -1 or + 1

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

  (one set is all i's where x_i = -1, the second set is all i's where x_i = +1)

dual function

$$\begin{aligned} g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

if W+diag(v) has at least one negative eigenvalue we can make x'(W+diag(v))x arbitrarily small

lower bound property: $p^\star \geq -\mathbf{1}^T \nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

The dual problem

Lagrange dual problem

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

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual (page 5–5)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad\qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

A nice example of why we care about dual problems

~~A nonconvex problem with strong duality~~

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

x'x - 1 ≤ 0

$A \not\succeq 0$, hence nonconvex

Range of a matrix A in R^{m*n}:
R(A) = { Ax | x in R^n }

* The span of columns of A
* The set of vectors y for which Ax = y has a solution

dual function: $g(\lambda) = \inf_x (x^T(A+\lambda I)x + 2b^T x - \lambda)$

- unbounded below if $A + \lambda I \not\succeq 0$ or if $A + \lambda I \succeq 0$ and $b \notin \mathcal{R}(A + \lambda I)$
- minimized by $x = -(A+\lambda I)^\dagger b$ otherwise: $g(\lambda) = -b^T(A+\lambda I)^\dagger b - \lambda$

dual problem

$$\begin{array}{ll} \text{maximize} & -b^T(A+\lambda I)^\dagger b - \lambda \\ \text{subject to} & A + \lambda I \succeq 0 \\ & b \in \mathcal{R}(A + \lambda I) \end{array}$$

For simplicity assume (A+λI) > 0

L(x,λ) = x'Ax + 2 b'x + λ(x'x - 1) = x'(A+λI)x + 2 b'x - λ
g(λ) = inf_x L(x,λ)
dL/dx = 2(A+λI)x + 2b = 0 => x^* = -(A+λI)^-1 b

Then g(λ) = L(x^*,λ) = -b' (A+λI)^-1 b - λ
Lagrange dual: max g(λ) s.t. λ ≥ 0

Let A = UDU', then A+λI = U(D+λI)U' = U S(λ) U', where s_ii(λ) = d_ii + λ
Then (A+λI)^-1 = U S^-1(λ) U', where s_ii^-1(λ) = 1/(d_ii + λ)

Let U = [u_1 … u_n], where u_i are column eigenvectors
g(λ) = -b' U S^-1(λ) U' b - λ = -Σ_i b' u_i s_ii^-1(λ) u'_i b - λ
= -Σ_i s_ii^-1(λ) (b' u_i)^2 - λ
dg/dλ = Σ_i (b' u_i)^2 / (d_ii + λ)^2 - 1
It is easy to use a ONE-DIMENSIONAL gradient ascent or Newton method!

Lagrange dual and conjugate function

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

dual function

$$g(\lambda, \nu) = \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right)$$

= inf_x{ fo(x) + (A'λ+C'v)'x } - b'λ - d'v
= - sup_x{ (-A'λ-C'v)'x - fo(x) } - b'λ - d'v
= - fo*(-A'λ-C'v) - b'λ - d'v

- recall definition of conjugate $f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^T x - f(x))$
- simplifies derivation of dual if conjugate of f_0 is known

example: entropy maximization

$$f_0(x) = \sum_{i=1}^{n} x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^{n} e^{y_i - 1}$$

Weak and strong duality

weak duality: $d^\star \le p^\star$

Remember the lower bound property: if λ ≥ 0 then g(λ,v) ≤ p^*
By taking the optimal λ^* and v^*, d^* = g(λ^*,v^*) ≤ p^*

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

  for example, solving the SDP

  Duality gap: p^* - d^*

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T\nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

  gives a lower bound for the two-way partitioning problem on page 5–7

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

- 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 qualification

strong duality holds for a convex problem

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

if it is strictly feasible, *i.e.*,

$$\exists x \qquad f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b$$

strict inequality

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened:

  Assume f_1(x) … f_k(x) are affine and dom(fo) open, then the REFINED Slater's condition is
  there is an x, f_i(x) ≤ 0 for i = 1…k f_i(x) < 0 for i = k+1…m Ax = b

  Thus, if all inequalities are affine (k=m) then strict inequality is not necessary!
- there exist many other types of constraint qualifications

Inequality form LP

primal problem

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

dual function

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

dual problem

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

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ except when primal and dual are infeasible **(refined Slater's)**

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

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

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

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

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always **(refined Slater's)**

Geometric interpretation

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

min fo(x)
s.t. f1(x) ≤ 0

interpretation of dual function:

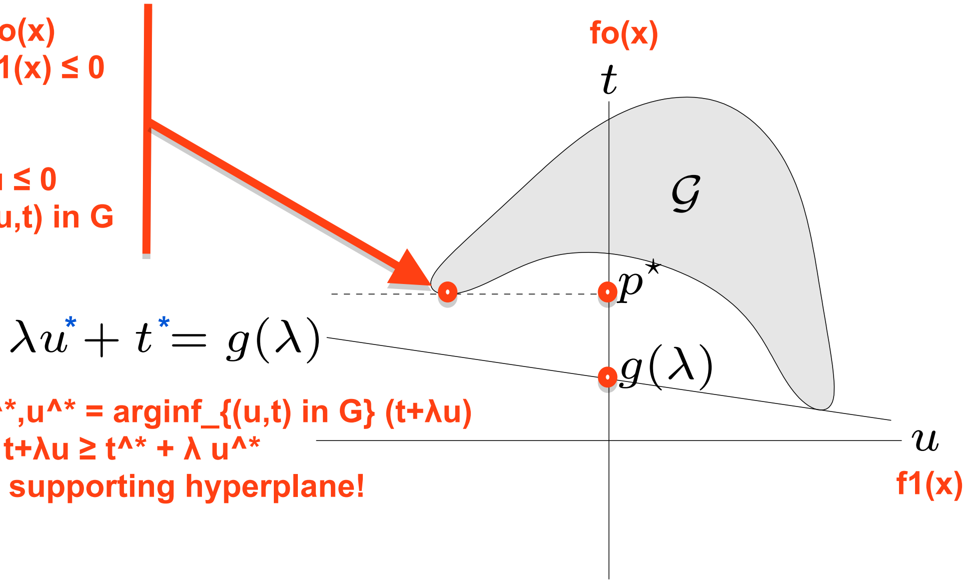
g(λ) = inf_{x in D} (fo(x) + λ f1(x))   equivalent to:

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

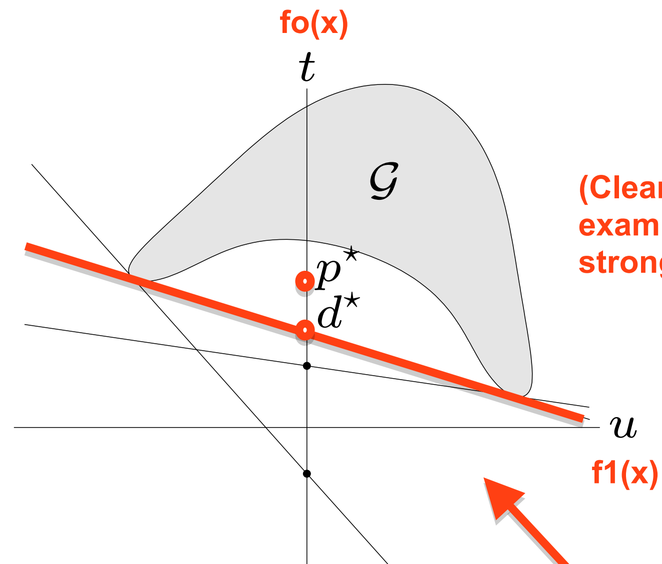
t

min fo(x)
s.t. f1(x) ≤ 0

min t
s.t. u ≤ 0
(u,t) in G



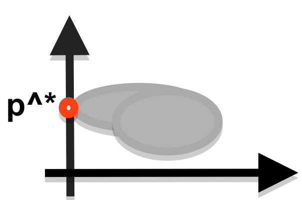
Let t^*,u^* = arginf_{(u,t) in G} (t+λu)
Then t+λu ≥ t^* + λ u^*
It is a supporting hyperplane!



(Clearly in this example there is not strong duality)

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

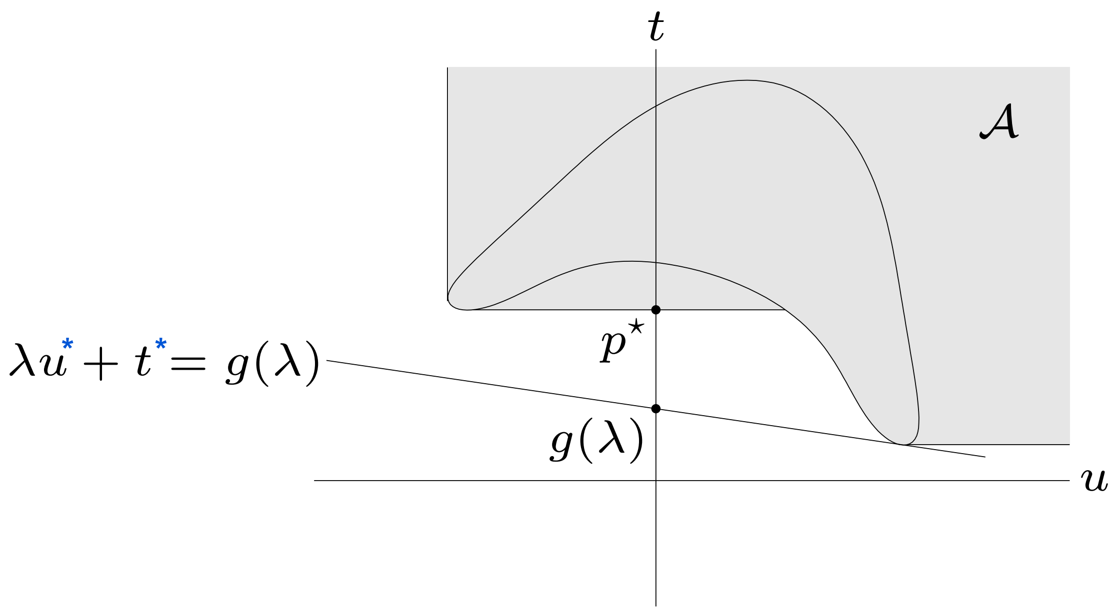
what if all u ≥ 0?
(i.e. f1(x) ≥ 0 for all x)
Constraint is f1(x) ≤ 0
Then solution has f1(x) = 0
VERTICAL supporting hyperplane
λ^* = infinity



Dual: λ^* = argmax_{λ≥0} g(λ)
d^* = g(λ^*) is the “tightest” supporting hyperplane
(you cannot go up without violating the definition
of supporting hyperplane)

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

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



strong duality

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

(explained in previous slide)

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$ **(Primal feasibility)**
2. dual constraints: $\lambda \succeq 0$ **(Dual feasiblity)**
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$ **if λ_i > 0 then f_i(x) = 0**
   if f_i(x) < 0 then λ_i = 0
4. gradient of Lagrangian with respect to x vanishes: **(Stationarity)**

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

from page 5–17: if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

General idea, for general possibly nonconvex primal problem: OPTIMAL => KKT satisfied.
(subject to some technical conditions)

Complementary slackness

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

fo(x^*) = g(λ^*,v^*) = inf_x L(x, λ^*, v^*)

$$
\begin{aligned}
&= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\
&\leq f_0(x^\star)
\end{aligned}
$$

hence, the two inequalities hold with equality

since h_i(x^*) = 0 given that x^* is feasible:
Σ_i λ_i^* f_i(x^*) = 0
but each term in sum is nonpositive (none of the terms can be negative because there will not be a positive to make sum = 0)

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

KKT conditions for convex problem

General idea, for convex primal problem: KKT satisfied => OPTIMAL and thus KKT satisfied <=> OPTIMAL
(subject to some technical conditions)

if $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

Since λ~_i f_i(x~) = 0, and also h_i(x~)=0
then Σ_i λ~_i f_i(x~) + Σ_i v~_i h_i(x~) = 0

- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

Stationarity: gradient of L(x, λ~, v~) w.r.t. x vanishes,
=> x~ minimizes L ... (this is why we assumed convexity
otherwise stationarity does not
imply that x~ is the minimizer of L)
=> L(x~,λ~,v~) = inf_x L(x, λ~, v~)
= g(λ~, v~)

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

zero duality gap since x~ = x^*, λ~ = λ^*, v~ = v^*
fo(x^*) = p^* = d^* = g(λ^*,v^*)

if **Slater's condition** is satisfied:

x is optimal if and only if there exist λ, ν that satisfy KKT conditions

Slide 5-11: Slater => strong duality
Slide 5-18: Strong duality + OPTIMAL => KKT satisfied
Here so far: KKT satisfied => OPTIMAL
Therefore assume Slater: KKT satistied <=> OPTIMAL

Lagrangian:
L(x,λ,v) = Σ_i { - log(x_i+a_i) } - λ'x + v(1'x - 1)
= Σ_i { - log(x_i+a_i) - λ_i x_i + v x_i } - v
Then:
dL/dx_i = -1/(x_i+a_i) - λ_i + v = 0

example: water-filling (assume $\alpha_i > 0$)

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

-x ≤ 0   1'x - 1 = 0

Primal feasibility

x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

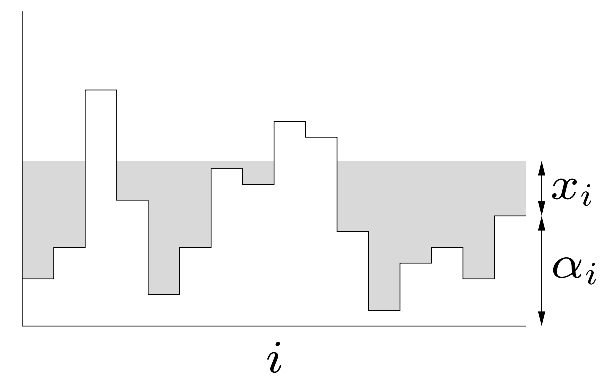
Dual feasibility   Complementary slackness   Stationarity

$$\lambda \succeq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$ (because λ_i cannot be negative)
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$ (because λ_i x_i = 0)
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- n patches; level of patch i is at height α_i



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 ϕ convex, increasing

Introducing new variables and equality constraints

$$\text{minimize} \quad f_0(Ax+b)$$

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

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax + b - y = 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

dual function follows from

$$g(\nu) = \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu)$$

= inf_y { fo(y) - v'y } + inf_x { v'Ax } + b'v
= - sup_y { -fo(y) + v'y} + inf_x { v'Ax } + b'v
= | -fo*(v) + b'v   if A'v = 0
   | -infinity   otherwise

Note: if A'v ≠ 0, we can pick x so that v'Ax is arbitrarily small

norm approximation problem: minimize $\|Ax - b\|$

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = Ax - b \end{array}$$

can look up conjugate of $\|\cdot\|$, or derive dual directly

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T \nu) \\ &= \begin{cases} b^T\nu + \inf_y(\|y\| + \nu^T y) & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} b^T\nu & A^T\nu = 0, \quad \|\nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

(see page 5–4)

dual of norm approximation problem

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & A^T\nu = 0, \quad \|\nu\|_* \le 1 \end{array}$$

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_{-\mathbf{1} \preceq x \preceq \mathbf{1}} (c^T x + \nu^T(Ax - b))$$

= inf_{|x|_infty ≤ 1} {(A’v+c)’x} - b’v
= - sup_{|x|_infty ≤ 1} {(-A’v-c)’x} - b’v
= - |A’v+c|_1 - b’v … by norm duality

dual problem: maximize $-b^T\nu - \|A^T\nu + c\|_1$

Problems with generalized inequalities

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

$\preceq_{K_i}$ is generalized inequality on $\mathbf{R}^{k_i}$

definitions are parallel to scalar case:

- Lagrange multiplier for $f_i(x) \preceq_{K_i} 0$ is vector $\lambda_i \in \mathbf{R}^{k_i}$
- Lagrangian $L : \mathbf{R}^n \times \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, is defined as

$$L(x, \lambda_1, \cdots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- dual function $g : \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, is defined as

$$g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \cdots, \lambda_m, \nu)$$

lower bound property: if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda_i \succeq_{K_i^*} 0$, then

Slide 2-21:
If λ_i ≥ 0 with respect to dual cone K_i*
and f_i(x~) ≤ 0 with respect to cone K_i
then λ_i ' f_i(x~) ≤ 0

$$\begin{aligned} f_0(\tilde{x}) &\geq f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \\ &\geq \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\ &= g(\lambda_1, \ldots, \lambda_m, \nu) \end{aligned}$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda_1, \ldots, \lambda_m, \nu)$

dual problem

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

- weak duality: $p^\star \geq d^\star$ always
- strong duality: $p^\star = d^\star$ for convex problem with constraint qualification (for example, Slater's: primal problem is strictly feasible)

Semidefinite program

primal SDP $(F_i, G \in \mathbf{S}^k)$

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

Remember tr(A'B) is the inner-product of matrices A and B

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

  = -tr(ZG) + Σ_i x_i (c_i + tr(Z F_i))

  Note: if c_i + tr(Z F_i) ≠ 0, we can pick x_i so that x_i (c_i + tr(Z F_i)) is arbitrarily small
- dual function

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

dual SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)