DSA3102: Lecture 9 (Reading Sec 5.5, 5.6, 5.7)

Optimization Problem:

$$\min_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad \begin{array}{ll} f_i(x) \le 0 & \forall i \in [m] \\ h_i(x) = 0 & \forall i \in [p]. \end{array}$$

$a_i^T x - b_i = 0$

Convex opt. Problem

$$p^* = \min_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad \begin{array}{ll} f_i(x) \le 0 & \forall i \in [m] \\ Ax = b, & A \in \mathbb{R}^{p \times n} \end{array}$$

f_i: convex $\forall\, i = 0, 1, \ldots, m.$

p affine constraints

Lagrangian:

Lagrange multipliers

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

primal decision variable

dual variables

Lagrange dual function:

D: domain of primal.

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

Dual Problem: $d^* = \max_{\lambda \ge 0, \nu \in \mathbb{R}^p} g(\lambda, \nu)$

Convex optimization problem.

(even if original problem the primal is not convex)

Weak duality: $d^* \leq p^*$

Thm: If the optimization problem is convex and $\exists\ \bar{x} \in \text{relint}(D)$ s.t.

relative interior

$$f_i(\bar{x}) < 0 \quad \forall i = 1, \dots, m$$
$$A\bar{x} = b$$

then strong duality holds, i.e., $p^* = d^*$.

Rmk: If some of the f_i's are affine, they don't have to be satisfied strictly by the Slater vector.

Min-max characterization of weak and strong duality

Assume that there are no equality constraints.

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

fix primal variable

Fix $x \in \mathbb{R}^n$, consider

$$\sup_{\substack{\lambda \geq 0 \\ \lambda \in \mathbb{R}^m_+}} L(x, \lambda) = \sup_{\lambda \geq 0} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right\}$$

x either feasible or non feasible

x feasible (all inequality constraints satisfied)

$$= \begin{cases} f_0(x) & f_i(x) \le 0, \ \forall i=1,\dots,m \\ +\infty & \text{else} \end{cases}$$

Primal optimal value:

$$p^* = \inf_{x \in D} \{ f_0(x) : f_i(x) \le 0 \quad \forall i=1,\dots,m \}$$

$$p^* = \inf_{x \in D} \sup_{\lambda \ge 0} L(x,\lambda).$$

min max problem

Dual function: $g(\lambda) = \inf_{x \in D} L(x,\lambda).$

Dual Optimal value: $d^* = \sup_{\lambda \ge 0} g(\lambda)$

$$d^* = \sup_{\lambda \ge 0} \inf_{x \in D} L(x,\lambda).$$

no constraints required

Weak duality: always true $\max_b \min_a h(a,b) \le \min_a \max_b h(a,b)$

$$d^* = \sup_{\lambda \ge 0} \inf_{x \in D} L(x,\lambda) \le p^* = \inf_{x \in D} \sup_{\lambda \ge 0} L(x,\lambda).$$

Strong duality: Under constraint qualifications,

$$d^* = \sup_{\lambda \ge 0} \inf_{x \in D} L(x,\lambda) = p^* = \inf_{x \in D} \sup_{\lambda \ge 0} L(x,\lambda).$$

Strong duality implies that the inf and sup can be interchanged.
Sion's minimax theorem.

Optimality Conditions.

If we can find a dual feasible (λ, ν) $(\lambda \geq 0, \nu \in \mathbb{R}^p)$
then $g(\lambda, \nu) \leq p^*$ by weak duality
(λ, ν) provides a certificate that $p^* \geq g(\lambda, \nu)$.

Given a primal decision variable x, we also know how suboptimal it is without knowing p^*.

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu)$$

Can be used to design "primal-dual" algorithms.

sequence of decision variables in $\mathbb{R}^n$

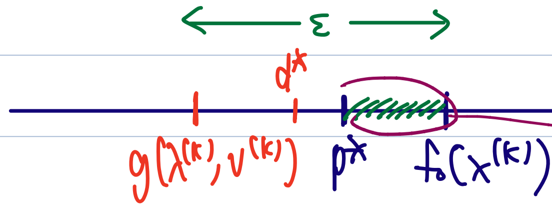
Algorithm: $\{x^{(k)}\}_{k=1}^{\infty} \subset \mathbb{R}^n$, $\{(\lambda^{(k)}, \nu^{(k)})\}_{k=1}^{\infty}$ dual
primal feasibles. feasible sequence

Stop whenever $f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \varepsilon$. — (+)
This guarantees that $x^{(k)}$ is at most ε-suboptimal

iteration index

$$f_0(x^{(k)}) - p^* \leq \varepsilon$$

want sequence of primal variables $x^{(k)}$ s.t $f_0(x^{(k)})$ comes extremely close to p^*



gap cannot be more than ε (by picture)

Guarantees that $x^{(k)}$, at the stopping time $\oplus$ holds, $x^{(k)}$ is ε-optimal.

Complementary Slackness

Suppose that the primal and dual optimal values (p^* and d^*) are attained by x^* and (λ^*, ν^*) respectively.

$$p^* = f_0(x^*) \qquad x^* \text{ feasible.}$$

$$d^* = g(\lambda^*, \nu^*) \qquad \lambda^* \geq 0 .$$

Assume strong duality holds. $(p^* = d^*)$

If slater's condition holds, dual optimal values are attained

$$f_0(x^*) = g(\lambda^*, \nu^*)$$

$$= \inf_{x \in D} L(x, \lambda^*, \nu^*) \quad (\text{def}^{\text{n}} \text{ of } L).$$

$$= \inf_{x \in D} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right\}$$

$$\leq f_0(x^*) + \sum_{i=1}^{m} \underbrace{\overset{\geq 0}{\lambda_i^*} \overset{\leq 0}{f_i(x^*)}}_{\leq 0} + \sum_{i=1}^{p} \underbrace{\nu_i^* h_i(x^*)}_{=0}$$

inf is smaller than anything evaluated at any point

$h_i(x) = 0$ since x^* is feasible

$\leq f_0(x^*)$ left same as right so everything is equal

All inequalities are equalities.

$$\sum_{i=1}^{m} \lambda_i f_i(x^*) = 0 \implies \lambda_i f_i(x^*) = 0 \quad \forall i \in [m]$$

because $\lambda_i f_i(x^*) \leq 0$ (in the same direction)

$\Downarrow$

Complementary slackness. (C.S.)

C.S. holds for all primal optimal x^* and dual optimal (λ^*, ν^*) when strong duality holds.

C.S. can be expressed as:

$$\lambda_i f_i(x^*) = 0 \quad \forall i \in [m]$$

or

i) $\lambda_i^* > 0$ for some $i \in [m]$, then $f_i(x^*) = 0$ (ith constraint is active)

ii) $f_i(x^*) < 0$ for some $i \in [m]$, then $\lambda_i^* = 0$.
(ith primal constraint is slack) (dual is 0)

The ith Lagrange multiplier is zero unless the ith constraint is active at the optimum. ($f_i(x^*) = 0$)

KKT optimality conditions:

Assume f_i, $i = 0, 1, \ldots, m$ & h_i, $i = 1, \ldots, p$ are diff^ble.

We do not assume that they are convex.

KKT conditions for nonconvex problems

f_i's may not be convex (h_i not affine too)

$$f_0(x^*) = g(\lambda^*, \nu^*)$$
$$= \inf_{x \in D} L(x, \lambda^*, \nu^*) \quad (\text{def}^{n} \text{ of } L).$$
$$= \inf_{x \in D} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right\}$$
$$\overset{=}{\leq} f_0(x^*) + \sum_{i=1}^{m} \underbrace{\overset{\geq 0}{\lambda_i^*} \overset{\leq 0}{f_i(x^*)}}_{\leq 0} + \sum_{i=1}^{p} \underbrace{\nu_i^* h_i(x^*)}_{=0}$$

equal to 0

$$\overset{=}{\leq} f_0(x^*)$$

Due to this equality x^* (primal feasible decision variable) also minimizes $L(x, \lambda^*, \nu^*)$.

due to equality

This means that $\nabla_x L(x, \lambda^*, \nu^*)\big|_{x=x^*} = 0$.

$$\nabla_x \left[f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right] = 0$$

Stationarity.

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

Primal feasibility: $f_i(x^*) \leq 0 \quad \forall i = 1, \dots, m$

$h_i(x^*) = 0 \quad \forall i = 1, \dots, p$

Dual feasibility: $\lambda_i^* \geq 0 \quad \forall i = 1, \dots, m$

Complementary Slackness: $\lambda_i^* f_i(x^*) = 0 \quad \forall i = 1, \ldots, m.$

These are called the Karush-Kuhn-Tucker (KKT) conditions.
(set of 4 different conditions)

Any optimization problem with diff^{ble} f_i, h_i for which strong duality holds, for any primal-dual optimal points (x^*, λ^*, v^*) must satisfy 4 KKT conditions.

KKT conditions are necessary for asserting that (x^*, λ^*, v^*) are primal-dual optimal. If (x^*, λ^*, v^*) are p-d optimal, they satisfy KKT. But given they satisfy KKT, may not be p-d optimal

---

KKT conditions for convex optimization problems.

convex opt prob

When the primal convex, the KKT conditions are also sufficient to say that (x^*, λ^*, v^*) is primal-dual optimal.

use 4 conditions to find p-d points

f_i: convex. $\qquad h_i$: affine.

$\tilde{x}$ is a primal point & $(\tilde{\lambda}, \tilde{v})$ is a dual point and they satisfy

$$\text{Primal feas: } \overset{\text{convex}}{f_i}(\tilde{x}) \le 0 \quad \forall i \in [m], \quad \overset{\text{affine}}{h_i}(\tilde{x}) = 0 \quad \forall i \in [p].$$

$$\text{Dual feas: } \tilde{\lambda}_i \ge 0 \quad \forall i \in [m]$$

$$\text{C.S. } \tilde{\lambda}_i f_i(\tilde{x}) = 0 \quad \forall i \in [m].$$

Stationarity:

$$\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{v}_i \nabla h_i(\tilde{x}) = 0$$

when KKT satisfied,

$\Rightarrow$ $\tilde{x}$ must be primal optimal.

$(\tilde{\lambda}, \tilde{v})$ must be dual optimal.

Why? Since $\tilde{\lambda}_i \ge 0$ for all i,

$$L(x, \tilde{\lambda}, \tilde{v}) = f_0(x) + \sum_i \tilde{\lambda}_i f_i(x) + \sum_i \tilde{v}_i h_i(x)$$

$L(x, \tilde{\lambda}, \tilde{v})$ is convex in x.

Stationarity guarantees that $\nabla_x L(x, \tilde{\lambda}, \tilde{v}) = 0$ vanishes at $\tilde{x}$

$\Rightarrow$ $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{v})$. Hence,

$$g(\tilde{\lambda}, \tilde{v}) = \inf_{x \in D} L(x, \tilde{\lambda}, \tilde{v})$$

$\tilde{x}$ minimises the Lagrangian

$$= L(\tilde{x}, \tilde{\lambda}, \tilde{v})$$

$$= f_0(\tilde{x}) + \underbrace{\cancel{\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x})}}_{0 \text{ due to complementary slackness}} + \underbrace{\cancel{\sum_{i=1}^{p} \tilde{v}_i h_i(\tilde{x})}}_{0 \text{ since } h_i(\tilde{x}) = 0}$$

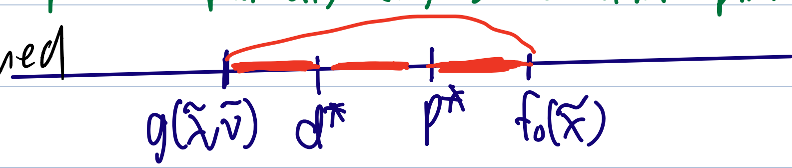
$\hat{x}$, $\hat{\lambda}$ and $\hat{v}$ all satisfy the KKT condition

$= f_0(\tilde{x})$

$\tilde{x}, (\tilde{\lambda}, \tilde{\nu})$ have zero duality gap

$\tilde{x}$ is primal optimal; $(\tilde{\lambda}, \tilde{\nu})$ is dual optimal

all sandwiched to be 0



$x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$. Problem convex

TFAE:

i) x & (λ, ν) together satisfy the KKT conditions.

ii) x & (λ, ν) are primal-dual optimal & strong duality holds.

$x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$. Problem convex and Slater satisfied

TFAE:

i) x & (λ, ν) together satisfy the KKT conditions.

ii) x & (λ, ν) are primal-dual optimal

Example: If Slater/SD does not hold, it possible that x, (λ, ν) are primal-dual optimal but do not satisfy KKT.

$$\min_{x \in \mathbb{R}, y > 0} \underbrace{e^{-x}}_{\text{convex}} \quad \text{s.t.} \quad \underbrace{x^2/y \leq 0}_{\text{convex}} \qquad D = \{(x,y) : y > 0\}$$

Primal optimal: $x^*=0,\ y^*>0$ } PF ✓
Dual optimal: $\lambda^* \geq 0$ } DF ✓

does not violate feasibility conditions of KKT

$$L(x,y,\lambda) = e^{-x} + \lambda\, x^2/y$$

$$g(\lambda) = \inf_{x,y>0} \{ e^{-x} + \lambda\, x^2/y \} = 0$$

Lagrange multiplier × function defining inequality constraint

$$d^* = \max_{\lambda \geq 0} g(\lambda) = \underline{\max_{\lambda \geq 0} 0} \qquad \lambda^* \geq 0$$

C.S. : $\lambda^* \cdot \frac{(x^*)^2}{y^*} = 0$ ✓ $\qquad \lambda_i f_i(x) = 0$

$f_0(x) + \lambda f_1(x) \qquad \nabla f_0(x) + \sum_i \lambda_i \nabla f_i(x) = 0$

Stationarity: $L(x,y,\lambda) = e^{-x} + \lambda \frac{x^2}{y}$.

2 condition

$$\frac{\partial}{\partial x} L(x,y,\lambda) = -e^{-x} + \lambda \frac{2x}{y} = -\underbrace{e^{-0}}_{=1} + \lambda^* \frac{0}{y^*} \overset{?}{=} 0 \quad \text{no!}$$

$$\frac{\partial}{\partial y} L(x,y,\lambda) = -\lambda \frac{x^2}{y^2} = 0 \quad \text{when } x = x^* = 0,\ y = y^* > 0.$$

$\Rightarrow$ KKT is not satisfied! (because slater strong duality does not hold)

Equality Constrained Convex Quadratic Minimization

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T P x + q^T x + r \qquad P \in S_+^n$$
$$\text{s.t. } Ax = b \qquad A \in \mathbb{R}^{p \times n},\ b \in \mathbb{R}^p.$$

convex opt prob
$\Downarrow$
KKT both neccesary and sufficient

KKT conditions: (x^*, ν^*) is primal-dual optimal

No inequality constraints
↳ no need to state dual feasibility and complementary slackness conditions

PF: $\quad Ax^* = b.$ $\qquad (A^T v^*)^T x$

Stationarity: $\quad L(x, v^*) = \frac{1}{2}x^T P x + q^T x + r + (v^*)^T (Ax - b)$

$\nabla_x L(x, v^*) = 0$

$$Px^* + q + A^T v^* = 0. \qquad x \in \mathbb{R}^n, \quad v \in \mathbb{R}^p$$

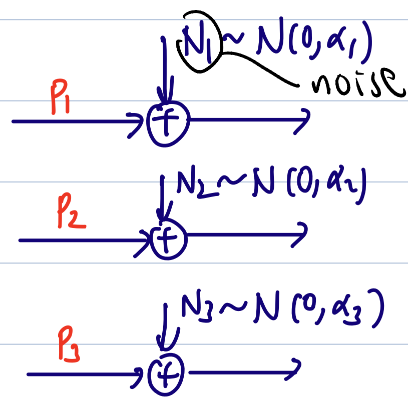
$$\Rightarrow \begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

Solving this linear system of eqns in $n+p$ variables (x^*, v^*) gives us the optimal primal & dual variables.

Water-filling

$\sum_{i=1}^{n} P_i = P$

← sum capacity



3 communication channels

$$\min_{x \in \mathbb{R}^n} \; -\sum_{i=1}^{n} \log(\alpha_i + x_i)$$

$-\log$ is convex

$$\text{s.t.} \quad x \geq 0, \quad \underbrace{\sum_{i=1}^{n} x_i = 1}_{1^T x = 1}$$

$-x \leq 0$

good if small α

Arises in information theory

$$L(x, \lambda, v) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) + \lambda^T(-x) + v(1^T x - 1)$$

$\lambda \in \mathbb{R}^n$, $v \in \mathbb{R}$

ln

KKT conditions: (x^*, λ^*, v^*) is primal-dual optimal if differentiate wrt x_i

Stationarity: $-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0 \quad \forall i = 1, \dots, n$

PF: $x^* \geq 0, \quad \mathbf{1}^T x^* = 1.$

DF: $\lambda^* \geq 0$

CS: $\lambda_i^* x_i^* = 0 \quad \forall i = 1, \dots, n.$

Stat: $\lambda_i^* = -\frac{1}{\alpha_i + x_i^*} + \nu^* \quad \forall i$

CS: $x_i^* \left(-\frac{1}{\alpha_i + x_i^*} + \nu^*\right) = 0 \quad \forall i.$

DF + Stat: $\nu^* \geq \frac{1}{\alpha_i + x_i^*}. \quad \forall i$ since $-\frac{1}{\alpha_i + x_i^*} + \nu^* \geq 0$

If $\nu^* < \frac{1}{\alpha_i}$, the last condition can only hold if $x_i^* > 0$. positive power

$\overset{\text{C.S.}}{\Rightarrow} \nu^* = \frac{1}{\alpha_i + x_i^*} \Rightarrow x_i^* = 1/\nu^* - \alpha_i.$

(must be lower bound)

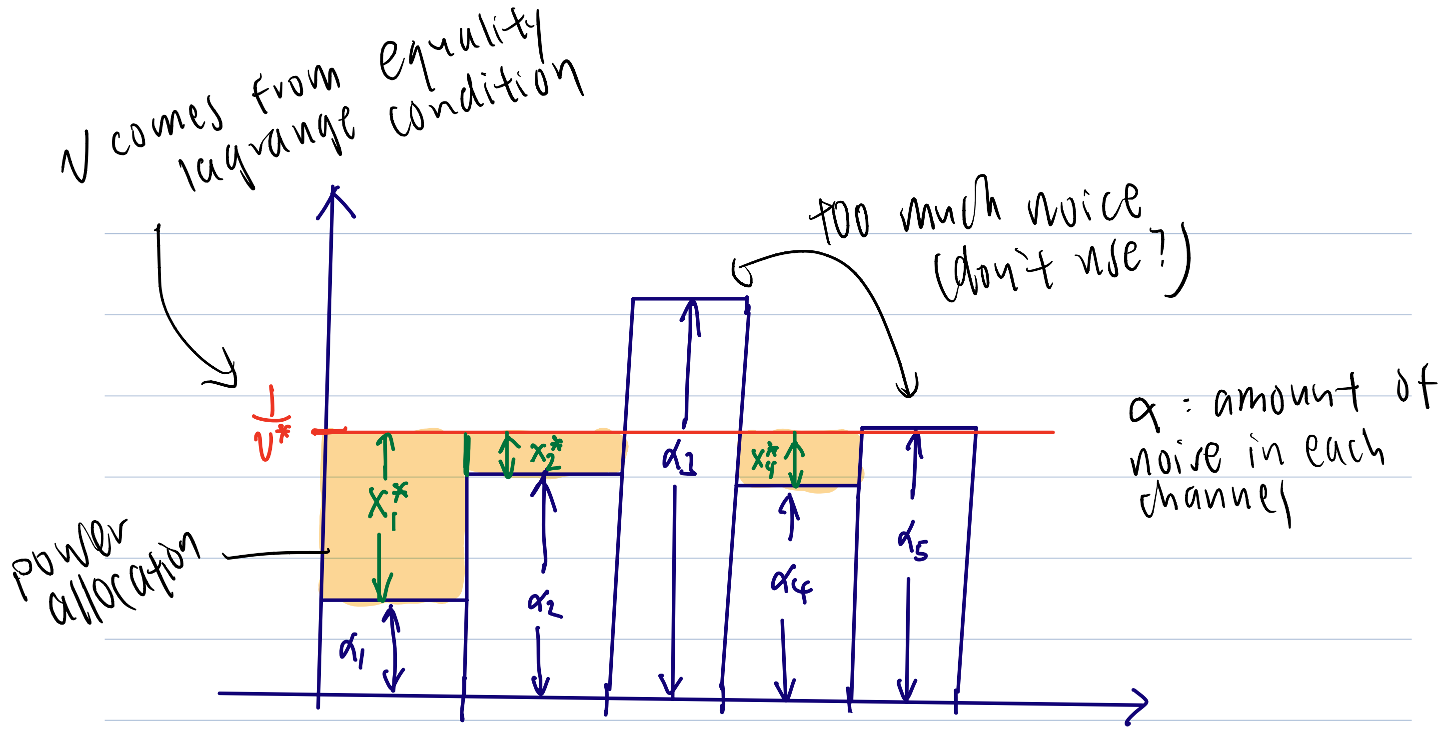
If $\nu^* \geq \frac{1}{\alpha_i}$, then $x_i^* > 0$ is impossible because this would imply that $\nu^* \geq 1/\alpha_i > \frac{1}{\alpha_i + x_i^*}$, which violates C.S.

$\Rightarrow x_i^* = 0$ if $\nu^* \geq 1/\alpha_i.$

$$x_i^* = \begin{cases} 1/\nu^* - \alpha_i, & \nu^* < 1/\alpha_i \\ 0 & \nu^* \geq 1/\alpha_i \end{cases}$$

$$x_i^* = \max\left\{0, \frac{1}{\nu^*} - \alpha_i\right\}$$

$$1 = \sum_{i=1}^{n} x_i^* = \sum_{i=1}^{n} \max\left\{0, \frac{1}{\nu^*} - \alpha_i\right\}$$



Ex: $\min_{x,y} \quad (x-2)^2 + 2(y-1)^2$ quadratic ⇒ convex opt problem

s.t. $x+4y \le 3, \quad -x+y \le 0.$

Lagrangian: $L(x,y,\lambda_1,\lambda_2) = (x-2)^2 + 2(y-1)^2 + \lambda_1(x+4y-3) + \lambda_2(-x+y)$

Stat: $\frac{\partial}{\partial x} L(x,y,\lambda_1,\lambda_2) = 2(x-2) + \lambda_1 - \lambda_2 = 0$

$\frac{\partial}{\partial y} L(x,y,\lambda_1,\lambda_2) = 4(y-1) + 4\lambda_1 + \lambda_2 = 0$

CS: $\lambda_1(x+4y-3) = 0, \quad \lambda_2(-x+y) = 0$

PF: $x+4y \le 3, \quad \underline{-x+y \le 0}.$

DF: $\lambda_1 \ge 0, \quad \lambda_2 \ge 0.$

Check 4 conditions in C.S.

always look at complementary slackness first.

$\lambda_2(-x+y) \Rightarrow \lambda_2$ not necessarily 0 but $-x+y=0$

look at stationary

① $\lambda_1=\lambda_2=0 \Rightarrow x=2, y=1.$ not PF ✗

② $\lambda_1=0, x=y \Rightarrow 2(x-2) = -4(y-1) = -4(x-1)$

$y=x=4/3, \quad \lambda_2 = 2(x-2) = -4/3.$ not DF ✗

③ $\lambda_2=0, x+4y-3=0 \Rightarrow 2(x-2) = y-1$ ✓

$2(3-4y-2) = y-1 \Rightarrow y = 1/3$

$x = 5/3, \quad \lambda_1 = -2(x-2) = 2/3$ ✓

④ $x=y, x+4y-3=0 \Rightarrow x=y=5/3$

want both inequality constraints to be active

$\lambda_1 = 22/25, \quad \lambda_2 = -48/25.$ not DF ✗

$(x = 5/3, y = 1/3)$ $\quad x+4y \leq 3$ ✓ $\quad -x+y \leq 0$ ✓

$(x^* = 5/3, y^* = 1/3)$ must be primal optimal

$\Rightarrow f_0(x^*, y^*) = (x^*-2)^2 + 2(y^*-1)^2 = 1.$

Duality & Problem Reformulations.

i) $\min_x f_0(Ax+b)$ $\quad$ Dualisation $\Rightarrow$ when have constraints

Introduce new variables & equality constraints.

$$\min_{x,y} f_0(y) \quad \text{s.t.} \quad y = Ax+b.$$

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

negative conjugate

linear function in x

$$= \inf_{x,y} \{ \boxed{f_0(y) - v^T y} + \cancel{(A^T v)^T} x + b^T v \}$$

$$= \begin{cases} -f_0^*(v) + b^T v & A^T v = 0 \\ -\infty & \text{else} \end{cases}$$

Dual Problem: $\max_v \; b^T v - f_0^*(v)$ s.t. $A^T v = 0$.

Implicit Constraints.

LP constraints

explicit constraints

$$\min_x \; c^T x \quad \text{s.t.} \quad \boxed{\begin{matrix} Ax = b \\ -\mathbf{1} \le x \le \mathbf{1} \end{matrix}}$$

x inside a box

Dual Problem: $g(v, \lambda_1, \lambda_2) = \inf_{x \in \mathbb{R}^n} \{ c^T x + v^T(Ax - b) + \lambda_1^T(x - \mathbf{1}) + \lambda_2^T(-\mathbf{1} - x) \}$

also an LP.

$$\max_{\lambda_1, \lambda_2} \; -b^T v - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2$$

$$c + A^T v + \lambda_1 - \lambda_2 = 0$$

$$\lambda_1, \lambda_2 \ge 0$$

Reformulate the box constraints so that they are implicit.

$$f_0(x) = \begin{cases} c^T x & -\mathbf{1} \le x \le \mathbf{1} \\ \infty & \text{else} \end{cases}$$

defined in terms of the box constraints

$$\min_x \; f_0(x) \quad \text{s.t.} \quad Ax = b.$$

no inequality constraints
inf over xs that respect the box constraint (domain restricted but fewer constraints)

$$g(\nu) = \inf_{x: -1 \le x \le 1} \left\{ c^T x + \nu^T (Ax - b) \right\}$$

$$= \inf_{x: -1 \le x \le 1} \left\{ c^T x + (A^T \nu)^T x - b^T \nu \right\}$$

$$= -b^T \nu + \inf_{x: -1 \le x \le 1} (c + A^T \nu)^T x$$

$$\inf_{-1 \le x \le 1} d^T x = \inf_{x: \|x\|_\infty \le 1} d^T x = -\|x\|_1$$

one norm is dual to infinity norm

largest abs value of $x \le 1$

$$g(\nu) = -b^T \nu - \| c + A^T \nu \|_1$$

Dual optimization problem: $\max_{\nu} \; -b^T \nu - \| c + A^T \nu \|_1$

but d^* values are the same