17. Duality

- Duality
- LPs and LCQPs

Duality

Constrained optimization

$x \in \{1, -1\}$
$\Leftrightarrow x^2 = 1$

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g_i(x) \leq 0,\ i = 1, ..., p \\ & h_j(x) = 0,\ j = 1, ..., q \end{array}$$

- Any optimization problem can be written in this way
- f, g_i's, h_j's may/may not be convex, smooth, continuous

Constrained optimization

$$\min_x \quad f(x) \qquad \text{st} \quad g_i(x) \le 0,\; h_j(x) = 0 \qquad (\star)$$

- We call $(\star)$ the primal problem
- The Lagrangian is an unconstrained (penalty) form

$$\mathcal{L}(x; u, v) = f(x) + u^T g(x) + v^T h(x)$$

  - $g(x) = [g_1(x), ..., g_p(x)]^T$, $h(x) = [h_1(x), ..., h_q(x)]^T$.
  - x is primal variable
  - u, v are Lagrange dual variables
- The dual problem is

$$\max_{u \ge 0, v} \quad \min_\theta \; \mathcal{L}(\theta; u, v)$$

Why it works

(P) $$\begin{array}{ll} \min\limits_{x} & f_P(x) := f(x) \\ \text{st} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

(D) $$\begin{array}{ll} \max\limits_{u,v} & \underbrace{\left(\min\limits_{x'} f(x') + u^T g(x') + v^T h(x') \right)}_{=: f_D(u,v)} \\ \text{st} & u \geq 0 \end{array}$$

Constraint enforcement in dual

D → +∞

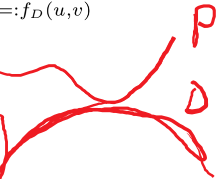
- Suppose $g_i(x') > 0$. Then dual is maximized by $u_i \to +\infty$
- Suppose $h_j(x) \neq 0$. Then dual is maximized by $v_j \to \mathbf{sign}(h_j(x)) \cdot +\infty$.

D → +∞ ⟶

Why it works

(P) $\quad \min_{x} \quad f_P(x) := f(x)$
$\quad\quad$ st $\quad g(x) \le 0$
$\quad\quad\quad\quad h(x) = 0$

(D) $\quad \max_{u,v} \quad \underbrace{\left(\min_{x'} f(x') + u^T g(x') + v^T h(x')\right)}_{=: f_D(u,v)}$
$\quad\quad$ st $\quad u \ge 0$



Weak duality:

- For x primal feasible, u dual feasible, $\boxed{f_P(x) \ge f_D(u,v)}$
- Proof: If $f_D(u,v) > f_P(x)$, then

$$f(x) + \underbrace{u^T g(x) + v^T h(x)}_{=0 \text{ since } x \text{ is primal-feasible}} < \min_{x'} f(x') + u^T g(x') + v^T h(x)$$

which is impossible.

Why it works

(P) $\min_x \quad f_P(x) := f(x)$
$\text{st} \quad g(x) \leq 0$
$\quad h(x) = 0$

(D) $\max_{u,v} \quad \underbrace{\left(\min_{x'} f(x') + u^T g(x') + v^T h(x')\right)}_{=: f_D(u,v)}$
$\text{st} \quad u \geq 0$

Strong duality:

- If f_P, g,h all convex, then under "mild conditions" *

$$f_P(x^*) = f_D(u^*, v^*)$$

where x^* optimizes the primal problem, u^* and v^* optimizes the dual problem.

- Proof is more complicated, but we will see some intuition.

* = Slater's conditions

Example

- Primal:

$$\min_x \quad \tfrac{1}{2}\|x-b\|_2^2,$$
$$\text{st} \quad Gx = f,\ x \geq 0$$

$+\infty$

$x_1 = -1$ $\quad$ $u_1 = +\infty$

- Lagrangian

$$\mathcal{L}(x;u,v) = \frac{1}{2}\|x-b\|_2^2 - u^T x + v^T(Gx-f)$$

- Given u, v, $\mathcal{L}$ is convex over θ, so find stationary point

$$0 = \nabla_x \mathcal{L}(x;u,v) \Rightarrow \boxed{x = b + u - G^T v}$$

Then dual is:

$$\max_{u,v} \min_x \mathcal{L}(x;u,v) = \max_{u,v} -\frac{1}{2}\|b+u-G^T v\|_2^2 + \frac{1}{2}b^T b - v^T f$$

st $u \geq 0$

Observations about "usefulness"

(P) $$\begin{array}{ll} \min\limits_{x} & f_P(x) := f(x) \\ \text{st} & g(x) \le 0 \\ & h(x) = 0 \end{array}$$

(D) $$\begin{array}{ll} \max\limits_{u,v} & \underbrace{\left(\min\limits_{x'} f(x') + u^T g(x') + v^T h(x')\right)}_{=:f_D(u,v)} \\ \text{st} & u \ge 0 \end{array}$$

- With constraints removed, Lagrangian minimization over x is often easier than solving the primal problem
- However, the resulting dual problem is <u>not necessarily</u> easier to solve than the primal problem
- Key word is <u>duality</u>, not <u>superiority</u>

Two-player game interpretation

Primal $\min_{x} \max_{u \geq 0, v} (\mathcal{L})$

$$\min_{x} \; \{f(x) : g(x) \leq 0, \;\; h(x) = 0\}$$

Lagrangian

$$\mathcal{L}(x; u, v) = f(x) + u^T g(x) + v^T h(x)$$

$u \geq 0$

Dual

$$\max_{u \geq 0, v} \; \min_{x} \mathcal{L}(x; u, v)$$

Player P

- Has control over x
- Goal: minimize $\mathcal{L}(x; u, v)$

Player D

- Has control over $u \geq 0$, v
- Goal: maximize $\mathcal{L}(x; u, v)$

Game play:

- P goes first, picks x
- D goes next, picks u, v

$$\min_{x} \max_{u \geq 0} \mathcal{L} \geq \max_{u \geq 0, v} \min_{x} \mathcal{L}$$

If primal went first, dual player enforces primal constraints

Primal

$$\min_x \{f(x) : g(x) \le 0,\ h(x) = 0\}$$

Lagrangian $, u \geq 0$

$$\mathcal{L}(x; u, v) = f(x) + u^T g(x) + v^T h(x)$$

Dual

$$\max_{u \ge 0, v} \boxed{\min_x \mathcal{L}(x; u, v)}$$

If P plays $g_k(x) > 0$ ...

$\vdots$

- D plays $u_k \to +\infty$...

If P plays $h_k(x) \neq 0$, $s = \mathbf{sign}(h_k(x))$

- D plays $sv_k \to +\infty$...

If x is not primal feasible, then

$$\max_{u,v} \mathcal{L}(x; u, v) = +\infty$$

P loses!

$$\max_{u \geq 0, v} \mathcal{L}(x'(u,v); u, v) = +\infty \text{ if } x' \text{ vio}$$

$\min_x \mathcal{L}(x,v) = \begin{cases} -\infty & \text{if } v < -1 \\ -v & \text{if } v \geq -1 \end{cases}$

Nonconvex case: order matters!

$2x + 2vx = 0$ if $v \neq -1$ $\to x = 0$ if $v > 1$

Consider

$$\min_x \quad x^2, \qquad \boxed{\mathcal{L}(x;v) = x^2 + v(x^2-1)}$$
$$\text{st} \quad x^2 = 1,$$

$x \in \{1, -1\}$

- If P goes first
  - either $x^2 = 1$ or $\min\limits_x \max\limits_v \mathcal{L}(x;v) \to +\infty$
  - $\mathcal{L}(x^*;v) = 1 + 0$, v can be anything

if $x^2 = 1$, D can choose any v

$V = \infty$: s.t. F

- If D goes first,
  - Given v, $\mathcal{L}$ convex in x so P picks stationary point $x = 0$ (primal infeasible)
  - $\mathcal{L}(x^*;v) = -v$
  - ~~If D were smart, it would have picked $v = 0$ to minimize worst case error~~

$V = -\infty$

When is order interchangable?

- Suppose that $L(x; u)$ is
  - a convex function in x
  - a concave function in u
- Then*

$$\min_x \max_u L(x; u) = \max_u \min_x L(x; u)$$

* = most of the time, unless Slater's conditions. For all the examples we consider in this class, you can assume this assumption is satisfied.

Convex* case: order interchangable

$$I_{\{\cdot \geq 0\}}(v) = \begin{cases} +\infty & v > 0 \\ 0 & v = 0 \end{cases}, \qquad I_{\{\cdot \geq 0\}}(v) \text{ is a convex function in } v$$

- Therefore

$$\mathcal{L}'(x; u, v) := f(x) + u^T g(x) + v^T h(x) - I_{\{\cdot \geq 0\}}(v)$$

concave in u, v
convex in x

is convex in x, concave in u and v

- If answer is not $-\infty$ then v is feasible
- Therefore

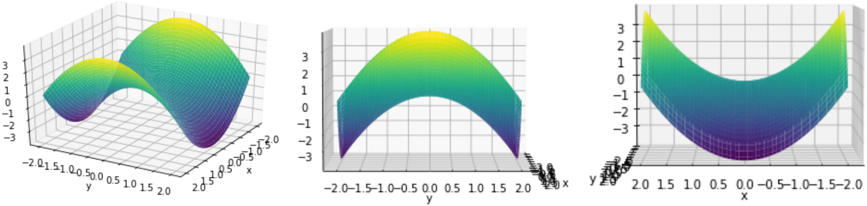
$$\begin{aligned} \text{Optimal dual objective} &= \max_{u,v} \min_{x} \mathcal{L}'(x; u, v) \\ &= \min_{x} \max_{u,v} \mathcal{L}'(x; u, v) \\ &= \text{Optimal dual objective} \end{aligned}$$

* = Slater's conditions

Saddle point optimization

$$\min_{x} \max_{u \geq 0, v} \mathcal{L}(x; u, v) = \max_{u \geq 0, v} \min_{x} \mathcal{L}(x; u, v) \qquad (\star\star)$$

- If $\mathcal{L}$ is convex in x / concave in u, v, $(\star\star)$ = saddle point optimization
- When strong duality holds, solving primal, dual, saddle are equivalent



Complementary slackness

(P) $\min_x \quad f_P(x) := f(x)$

st $\quad g(x) \leq 0$

$\quad h(x) = 0$

(D) $\max_{u,v} \quad \underbrace{\left(\min_{x'} f(x') + u^T g(x') + v^T h(x')\right)}_{=: f_D(u,v)}$

st $\quad u \geq 0$

- If x is primal feasible and u, v dual feasible, then
$$u_i g_i(x) \geq 0, \ \forall i, \qquad v_j h_j(x) = 0, \ \forall j$$
- This is a simple proof for weak duality
$$f_P(x) - f_D(u,v) \geq \underbrace{u^T g(x)}_{\geq 0} + \underbrace{v^T h(x)}_{=0} \geq 0$$
- **Complementary slackness**: at optimality, if strong duality holds,
$$g(x^*)_j u_j^* = 0 \ \forall j$$

either $g_j(x^*) = 0$

or $u_j^* = 0$

or both

(This is stronger than just $u^T g(x) = 0$)

Four equivalences when strong duality holds

$$\text{(P)} \quad \min_x \quad f_P(x) := f(x) \qquad \text{(D)} \quad \max_{u,v} \quad \underbrace{\left(\min_{x'} f(x') + u^T g(x') + v^T h(x')\right)}_{=: f_D(u,v)}$$

$$\text{st} \quad g(x) \leq 0 \qquad \qquad \text{st} \quad u \geq 0$$

$$h(x) = 0$$

- x solves the primal problem
- u, v solves the dual problem and I can get minimizer x' from u, v
- x, u, v solve the saddle optimization problem ← primal-dual solvers
- x is primal feasible, u, v are dual feasible, complementary slackness holds

Sequential minimization optimization (SMO

2 important examples where strong duality holds

(P)

Linear programs (LP)

(D)

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \qquad \begin{aligned} \underset{s\in\mathbb{R}^n,\, y\in\mathbb{R}^m}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & A^T y - s = c, \\ & s \geq 0, \end{aligned}$$

$A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$

$A^T v \geq c$

- Examples: Resource allocation, economics, famous "diet problem"
- Linear convexification of many nonconvex problems
  - Adding constraint $x_i =$ integer $\rightarrow$ mixed-integer linear programs (MILP)
- Primal or dual problem both often used

Field is too rich to summarize in one slide (Dantzig: Linear programming '63)

"How" $\cancel{0 \le c^T x + v^T b = c^T x + v^T A x}$

Duality and complementary slackness

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b,\ x \ge 0, \end{array}$$

Lagrangian

$$\mathcal{L}(x; u, v) = \boxed{c^T x} + v^T(Ax - b) - \boxed{u^T x}$$

$$= x^T(c + A^T v - u) - v^T b$$

$$\min_x \mathcal{L}(x; u, v) = \begin{cases} -v^T b & \text{if } c + A^T v - u = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem

$$\max_{u \ge 0} \; -v^T b$$

$$\text{s.t.} \quad c + A^T v - u = 0$$

u^*, v^*

$u_i^* x_i^* = 0$

Linearly constrained quadratic programs (LCQP)

$$
\begin{array}{ll}
\underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2} x^T Q x + p^T x \\
\text{subject to} & Ax \le b, \; Cx = d.
\end{array}
$$

$Q \in \mathbb{R}^{n \times n}$ p.s.d., $A \in \mathbb{R}^{m_1 \times n}$, $b \in \mathbb{R}^{m_1}$, $C \in \mathbb{R}^{m_2 \times n}$, $d \in \mathbb{R}^{m_2}$

- The example we really care about: SVMs
- Other examples:
  - constrained linear regression
  - 2nd-order approximations of more complicated problems
  - Trust region methods

$\nabla_x x^T A^T b = A^T b$ $\quad \nabla_x \frac{1}{2} x^T Q x = Qx, \quad \nabla_x c^T x = c$

Duality and complementary slackness

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \frac{1}{2}x^TQx + p^Tx \\ \text{subject to} & Ax \le b, \ Cx = d, \end{array} \qquad Q \text{ is invertible}$$

$u \qquad v$

Lagrangian

$$\mathcal{L}(x; u, v) = \frac{1}{2}x^TQx + p^Tx - u^T(b - Ax) + v^T(Cx - d)$$

$$\boxed{Qx + \underbrace{p + A^Tu + C^Tv}_{z} = 0}$$

Dual problem

$$= \frac{1}{2}x^TQx + x^Tz - u^Tb - v^Td$$

$$\boxed{x = -Q^{-1}z} \rightarrow -\frac{1}{2}z^TQ^{-1}z - u^Tb - v^Td$$

$$\max_{u, v, z} \quad u \ge 0, \quad z = p + A^Tu + C^Tv$$

Primal and dual LCQPs: Q is positive definite (invertible)

- Primal

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \frac{1}{2}x^TQx + p^Tx \\ \text{subject to} & Ax \leq b,\ Cx = d, \end{array}$$

- Dual

$$\begin{array}{ll} \underset{u\in\mathbb{R}^{m_1},v\in\mathbb{R}^{m_2},z}{\text{maximize}} & -\frac{1}{2}z^TQ^{-1}z - u^Tb - v^Td \\ \text{subject to} & z = p + A^Tu + C^Tv \in \mathbb{R}^n \\ & u \geq 0 \end{array}$$

- Complementary slackness

$$u_i^*(b_i - a_i^Tx^*) = 0, \quad i = 1, ..., m_1$$

- Recovery: $x = -Q^{-1}z$

Duality and complementary slackness

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2}x^T Qx + p^T x \\ \text{subject to} & Ax \leq b,\ Cx = d, \end{array}$$

Q is not necessarily invertible

- Lagrangian

$$\mathcal{L}(x; u, v) := \underbrace{\frac{1}{2}x^T Qx + z^T x}_{\text{depends on } x} + \mathcal{L}_0, \quad z = p + A^T u + C^T v,\ \mathcal{L}_0 = -u^T b - v^T d$$

- Linear decomposition theorem

$$z = z_N + z_R, \quad Qz_N = 0, \quad z_R = Q^T w$$

- Suppose $z_N \neq 0$. Then pick $x = -\alpha z_N$. Then

$$\min_x \mathcal{L}(x; u, v) \leq \underbrace{\frac{1}{2}x^T Qx}_{=0} - \alpha \|z_N\|_2^2 + \mathcal{L}_0 \overset{\alpha \to +\infty}{\to} -\infty$$

- Player D insists that $z = z_R \in \mathbf{range}(Q)$

Duality and complementary slackness

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \frac{1}{2}x^TQx + p^Tx \\ \text{subject to} & Ax \le b,\ Cx = d, \end{array} \qquad Q \text{ is not necessarily invertible}$$

- Lagrangian

$$\mathcal{L}(x; u, v) := \frac{1}{2}x^TQx + z^Tx + \mathcal{L}_0, \quad z = p + A^Tu + C^Tv$$

- Stationary point

$$Qx = z \in \mathbf{range}(Q) \iff x = Q^\dagger z$$

- **Proof.**

Duality and complementary slackness

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2} x^T Q x + p^T x \\ \text{subject to} & Ax \le b, \; Cx = d, \end{array}$$

Q is not necessarily invertible

- Lagrangian

$$\mathcal{L}(x; u, v) := \frac{1}{2} x^T Q x + z^T x + \mathcal{L}_0, \quad z = p + A^T u + C^T v$$

- Stationary point

$$Qx = z \in \mathbf{range}(Q) \iff x = Q^\dagger z$$

- **Proof.** Take skinny eigenvalue decomposition of Q

$$Qx = U \Lambda U^T x = z \iff \underbrace{U \Lambda^{-1} U^T}_{= Q^\dagger} z = x$$

Primal and dual LCQPs: Q is positive semidefinite

- Primal

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2} x^T Q x + p^T x \\ \text{subject to} & Ax \leq b,\ Cx = d, \end{array}$$

- Dual

$$\begin{array}{ll} \underset{u \in \mathbb{R}^{m_1}, v \in \mathbb{R}^{m_2}, z}{\text{maximize}} & -\frac{1}{2} z^T Q^\dagger z - u^T b - v^T d \\ \text{subject to} & z = p + A^T u + C^T v \in \mathbf{range}(Q) \\ & u \geq 0 \end{array}$$

- Complementary slackness

$$u_i^* (b_i - a_i^T x^*) = 0, \quad i = 1, ..., m_1$$

- Recovery: $x = -Q^\dagger z$

Support vector machines

$$\nabla_x (x^T y) = \nabla_x y^T x = y \qquad u^T Z \theta = \theta^T Z^T u$$

Hard Margin SVM duality

$$Z = \begin{bmatrix} y_1 x_1^T \\ \vdots \\ y_m x_m^T \end{bmatrix}$$

$$\begin{array}{ll} \underset{\theta}{\text{minimize}} & \frac{1}{2}\|\theta\|_2^2 \\ \text{subject to} & y_i x_i^T \theta \geq 1,\ i = 1, \ldots, m \end{array}$$

Lagrangian

$$\mathcal{L}(\theta; u) = \frac{1}{2}\|\theta\|_2^2 - u^T(Z\theta - \mathbb{1}), \quad \text{s.t. } u \geq 0$$

$$\min_{\theta} \mathcal{L}(\theta; u), \qquad \boxed{\theta = Z^T u}$$

Dual problem

$$\mathcal{L} = \frac{1}{2}\|Z^T u\|_2^2 - \|Z^T u\|_2^2 + \mathbb{1}^T u$$

$$\boxed{\max_u \ -\frac{1}{2}\|Z^T u\|_2^2 + \mathbb{1}^T u \quad \text{s.t. } \boxed{u \geq 0}}$$

Hard Margin SVM duality

(P) $\min_{\theta} \quad \frac{1}{2}\|\theta\|_2^2$
$\quad\;\;$ st $\quad y_i x_i^T \theta \geq 1, \; i = 1, ..., m$

(D) $\max_{u} \quad -\frac{1}{2}\sum_{i=1}^m (y_i x_i^T u)^2 + u^T \mathbf{1}$
$\quad\;\;$ st $\quad u \geq 0$

Primal recovery

$\theta = Z^T u$

Complementary slackness

$\text{dgap} = u^T(Z\theta - 1)$

$u_i(z_i^T\theta - 1) = 0 \quad \forall i$

$u_i > 0 \Rightarrow y_i x_i^T \theta = 1$ ← support vector

Hard Margin SVM duality summary

$$\text{(P)} \quad \begin{array}{ll} \min\limits_{\theta} & \frac{1}{2}\|\theta\|_2^2 \\ \text{st} & Z\theta \geq \mathbf{1} \end{array} \qquad \text{(D)} \quad \begin{array}{ll} \max\limits_{u} & -\frac{1}{2}\|Z^T u\|_2^2 + u^T \mathbf{1} \\ \text{st} & u \geq 0 \end{array}$$

- $Z = \begin{bmatrix} x_1 y_1 & \cdots & x_m y_m \end{bmatrix}^T$
- Primal recovery from dual optimal: $\boxed{\theta^* = Z^T u^*}$
- Complementary slackness

$$u_i > 0 \text{ or } y_i z_i > 1 \text{ but never both.}$$

- Dual recovery of support vectors:

$$\{i : u_i > 0\} \supseteq \{i : y_i x_i = 1\} = \text{ support vectors}$$

Hard margin SVM computation

$$\text{(P)} \quad \boxed{\begin{array}{ll} \min_{\theta} & \frac{1}{2}\|\theta\|_2^2 \\ \text{st} & Z\theta \geq \mathbf{1} \end{array}} \qquad \text{(D)} \quad \boxed{\begin{array}{ll} \max_{u} & -\frac{1}{2}\|Z^T u\|_2^2 + u^T\mathbf{1} \\ \text{st} & u \geq 0 \end{array}}$$

min

- $Z = \begin{bmatrix} x_1 y_1 & \cdots & x_m y_m \end{bmatrix}^T$
- Gradient of primal vs gradient of dual
  - Slightly more computation, requires one pass through data
- Projection on feasibility sets

  ZZ^T
  - Very difficult for primal, very simple for dual
- In general, solving dual SVM is preferred over solving primal SVM
- Suggested exercise: repeat all of this for soft margin SVM!