Lecture 10 - Duality in Convex Optimization

October 30, 2024

Happy Halloween!

A Convex (?) Set

Today's Agenda: Convex Duality

Primal Problem

$$
\begin{aligned}
(\mathscr{P}) \ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned} \tag{1}
$$

- Convex domain $X \subseteq \mathbb{R}^n$
- Every function $f_i : X \subseteq \mathbb{R}^n \to \mathbb{R}$ (real-valued), **convex**
- Equality constraints $Ax = b$ can be included in X

Today's Agenda: Convex Duality

Primal Problem

$$(\mathscr{P})\ \text{minimize}_x \quad f_0(x) \qquad f_i(x) \le 0, \quad i = 1, \dots, m \qquad x \in X. \tag{1}$$

- Convex domain $X \subseteq \mathbb{R}^n$
- Every function $f_i : X \subseteq \mathbb{R}^n \to \mathbb{R}$ (real-valued), **convex**
- Equality constraints $Ax = b$ can be included in X

- Many developments deal with the "interior" of X

Definition : Interior

The **interior** of a set X is the set of all points $x \in X$ so that:

$$\exists\, r > 0 \ :\ B(x, r) := \{y : \|y - x\| \le r\} \subseteq X$$

Must talk about the interior even if X is not full-dimensional ...

Relative Interior

- **Recall**: **Affine hull** of X is $\mathrm{aff}(X) := \left\{\theta_1 x_1 + \cdots + \theta_k x_k \;:\; x_i \in X,\, \sum_{i=1}^k \theta_i = 1\right\}$

Relative Interior

- **Recall**: **Affine hull** of X is $\mathrm{aff}(X) := \{\theta_1 x_1 + \cdots + \theta_k x_k \ : \ x_i \in X, \sum_{i=1}^k \theta_i = 1\}$

Definition Relative Interior

The **relative interior** of a set X is:

$$\text{rel int}(X) := \left\{x \in X \ : \ \exists\, r > 0 \text{ so that } B(x,r) \cap \mathrm{aff}(X) \subseteq C\right\}. \tag{2}$$

What is the relative interior of the following sets?

- $\{(x, y) \in \mathbb{R}^2 \mid (x, y) \in [0, 1]^2\}$
- $\{(x, y) \in \mathbb{R}^2 \mid x + y = 1, x \geq 0, y \geq 0\}$
- $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$

Convex Duality

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

- Convex domain $X \subseteq \mathbb{R}^n$
- Every function $f_i : X \subseteq \mathbb{R}^n \to \mathbb{R}$ (real-valued), **convex**
- Equality constraints $Ax = b$ can be included in X

- Assume $\mathrm{rel\ int}(X) \neq \emptyset$
- Assume that $(\mathscr{P})$ has an optimal solution x^*, optimal value $p^* = f_0(x^*)$

- **Core questions:**
  1. For x feasible for $(\mathscr{P})$, how to **quantify the optimality gap** $f_0(x) - p^*$?
  2. How to certify that x^* is **optimal** in $(\mathscr{P})$?

Convex Duality

Primal Problem

$$
(\mathscr{P}) \ \text{minimize}_x \quad \begin{aligned} & f_0(x) \\ & f_i(x) \leq 0, \qquad i = 1, \ldots, m \\ & x \in X. \end{aligned}
$$

- To construct **lower bounds** for $(\mathscr{P})$, define the **Lagrangian function**:

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

Convex Duality

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \le 0, \qquad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

- To construct **lower bounds** for $(\mathscr{P})$, define the **Lagrangian function**:

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

- By construction, $\mathcal{L}(x, \lambda) \le f_0(x)$ for any x feasible in $(\mathscr{P})$

Convex Duality

Primal Problem

$$(\mathscr{P})\ \text{minimize}_x \quad \begin{array}{ll} f_0(x) & \\ f_i(x) \leq 0, & i = 1, \ldots, m \\ x \in X. & \end{array}$$

- To construct **lower bounds** for $(\mathscr{P})$, define the **Lagrangian function**:

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

- By construction, $\mathcal{L}(x, \lambda) \leq f_0(x)$ for any x feasible in $(\mathscr{P})$
- For a lower bound on p^*, minimize $\mathcal{L}(x, \lambda)$ over $x \in X$

$$g(\lambda) := \inf_{x \in X} \mathcal{L}(x, \lambda).$$

Dual Problem

$$(\mathscr{D}) \quad \sup_{\lambda \geq 0} g(\lambda).$$

Q: Is the dual $(\mathscr{D})$ a convex optimization problem?

Convex Duality

Primal Problem

$$(\mathscr{P})\ \text{minimize}_x \quad \begin{array}{ll} f_0(x) & \\ f_i(x) \leq 0, & i = 1, \ldots, m \\ x \in X. & \end{array}$$

- To construct **lower bounds** for $(\mathscr{P})$, define the **Lagrangian function**:

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

- By construction, $\mathcal{L}(x, \lambda) \leq f_0(x)$ for any x feasible in $(\mathscr{P})$
- For a lower bound on p^*, minimize $\mathcal{L}(x, \lambda)$ over $x \in X$

$$g(\lambda) := \inf_{x \in X} \mathcal{L}(x, \lambda).$$

Dual Problem

$$(\mathscr{D}) \quad \sup_{\lambda \geq 0} g(\lambda).$$

Q: Is the dual $(\mathscr{D})$ a convex optimization problem?

Geometric Interpretation

Primal-Dual Pair

$$(\mathscr{P})\ p^* := \inf_{x\in X} f_0(x) \qquad\qquad (\mathscr{D})\quad d^* := \sup_{\lambda\geq 0} g(\lambda)$$
$$f_i(x) \leq 0,\ i = 1,\ldots,m$$

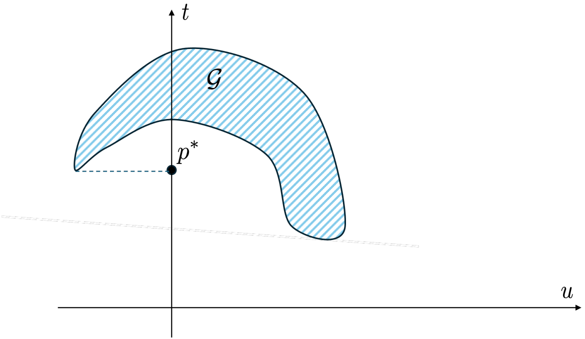
- Suppose $(\mathscr{P})$ has just one inequality constraint, i.e., $m = 1$
- Let $\mathcal{G} := \{(u,t) \in \mathbb{R}^2 : \exists x \in \mathbb{R}^n,\ t = f_0(x),\ u = f_1(x)\}$

Geometric Interpretation

Primal-Dual Pair

$$(\mathscr{P})\ p^* := \inf_{x \in X} f_0(x) \qquad\qquad (\mathscr{D}) \quad d^* := \sup_{\lambda \geq 0} g(\lambda)$$
$$f_i(x) \leq 0,\ i = 1, \ldots, m$$

- Suppose $(\mathscr{P})$ has just one inequality constraint, i.e., $m = 1$
- Let $\mathcal{G} := \{(u, t) \in \mathbb{R}^2 : \exists x \in \mathbb{R}^n,\ t = f_0(x),\ u = f_1(x)\}$



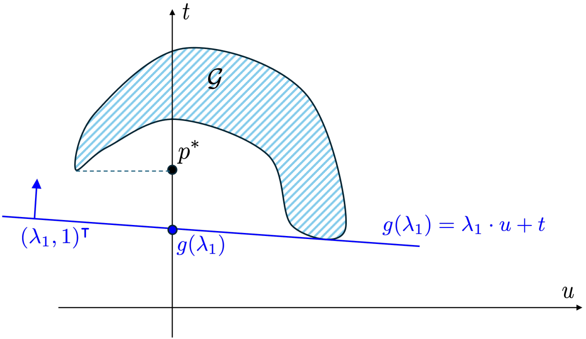
Geometric Interpretation

Primal-Dual Pair

$$(\mathscr{P}) \; p^* := \inf_{x\in X} \; f_0(x) \qquad\qquad (\mathscr{D}) \quad d^* := \sup_{\lambda\geq 0} g(\lambda)$$

$$f_i(x) \leq 0, \; i = 1, \ldots, m$$

- Suppose $(\mathscr{P})$ has just one inequality constraint, i.e., $m = 1$
- Let $\mathcal{G} := \{(u, t) \in \mathbb{R}^2 : \exists x \in \mathbb{R}^n, \; t = f_0(x), \; u = f_1(x)\}$



- Given $\lambda \geq 0$, to find $g(\lambda)$ we must minimize $t + \lambda \cdot u$ over $(u, t) \in \mathcal{G}$

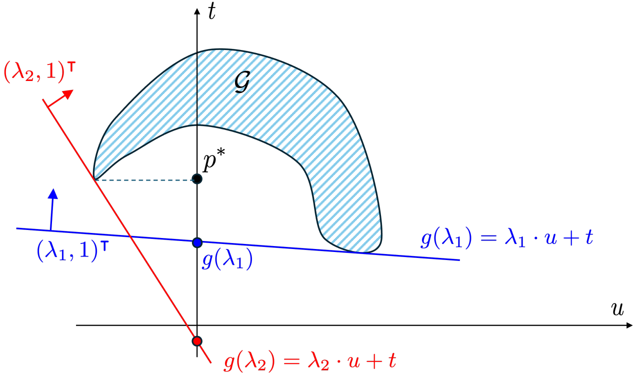
Geometric Interpretation

Primal-Dual Pair

$$(\mathscr{P})\ p^* := \inf_{x \in X} f_0(x) \qquad\qquad (\mathscr{D}) \quad d^* := \sup_{\lambda \geq 0} g(\lambda)$$

$$f_i(x) \leq 0,\ i = 1, \ldots, m$$

- Suppose $(\mathscr{P})$ has just one inequality constraint, i.e., $m = 1$
- Let $\mathcal{G} := \{(u, t) \in \mathbb{R}^2 : \exists x \in \mathbb{R}^n,\ t = f_0(x),\ u = f_1(x)\}$



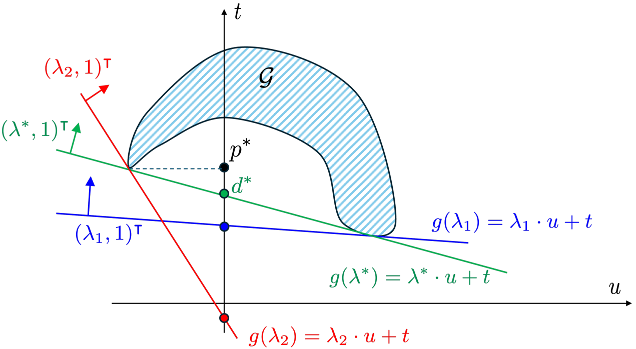
Geometric Interpretation

Primal-Dual Pair

$$(\mathscr{P})\ p^* := \inf_{x\in X} f_0(x) \qquad\qquad (\mathscr{D})\quad d^* := \sup_{\lambda\geq 0} g(\lambda)$$

$$f_i(x) \leq 0,\ i = 1,\ldots,m$$

- Suppose $(\mathscr{P})$ has just one inequality constraint, i.e., $m = 1$
- Let $\mathcal{G} := \{(u,t) \in \mathbb{R}^2 : \exists x \in \mathbb{R}^n,\ t = f_0(x),\ u = f_1(x)\}$



- Here, strong duality does not hold: $d^* < p^*$. But the set $\mathcal{G}$ is not convex!

Strong Duality in Convex Optimization?

Strong Duality in Convex Optimization?

Non-zero duality gap

Consider the example:

$$\underset{(x,y)\in X}{\text{minimize}}\ e^{-x}$$

$$x^2/y \leq 0$$

with domain $X = \{(x, y) \mid y \geq 1\}$.

Strong Duality in Convex Optimization?

Non-zero duality gap

Consider the example:

$$\underset{(x,y)\in X}{\text{minimize}}\; e^{-x}$$

$$x^2/y \leq 0$$

with domain $X = \{(x, y) \mid y \geq 1\}$.

- Convex optimization problem!
- *What are $p^*, \mathcal{L}$, g, d^*?*

Strong Duality in Convex Optimization?

Non-zero duality gap

Consider the example:

$$\underset{(x,y)\in X}{\text{minimize}}\ e^{-x}$$

$$x^2/y \leq 0$$

with domain $X = \{(x, y) \mid y \geq 1\}$.

- Convex optimization problem!
- *What are $p^*, \mathcal{L}$, g, d^*?*

Conditions Leading to Strong Duality

Primal Problem

$$
(\mathscr{P}) \ \text{minimize}_x \quad \begin{aligned} & f_0(x) \\ & f_i(x) \leq 0, \qquad i = 1, \ldots, m \\ & x \in X. \end{aligned}
$$

Conditions Leading to Strong Duality

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \le 0, \qquad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Slater Condition

The functions $f_1, \ldots, f_m : X \subseteq \mathbb{R}^n \to \mathbb{R}$ satisfy **the Slater condition on X** if there exists $x \in \mathrm{rel\ int}(X)$ such that

$$f_j(x) < 0, \quad j = 1, \ldots, m.$$

Conditions Leading to Strong Duality

Primal Problem

$$
\begin{aligned}
(\mathscr{P}) \ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Slater Condition

The functions $f_1, \ldots, f_m : X \subseteq \mathbb{R}^n \to \mathbb{R}$ satisfy **the Slater condition on X** if there exists $x \in \mathrm{rel\ int}(X)$ such that

$$f_j(x) < 0, \quad j = 1, \ldots, m.$$

- A point x that is **strictly feasible**
- Condition simpler if some f_i are affine: only require $f_i(x) < 0$ for the **non-linear** f_i

Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

Geometric intuition for proof:

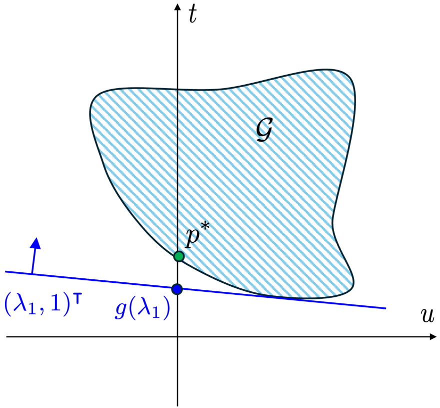
Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

Geometric intuition for proof:

- Recall case with $m = 1$ and $\mathcal{G} := \{(u, t) \in \mathbb{R}^2 : \exists x \in \mathbb{R}^n,\ t = f_0(x),\ u = f_1(x)\}$



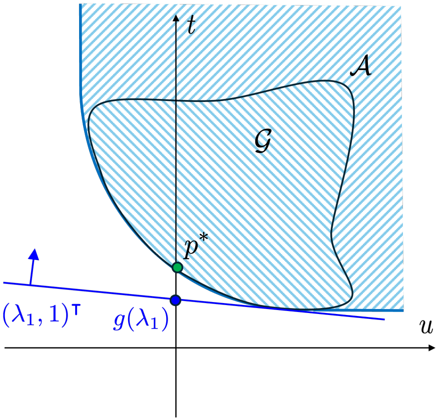
Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

Geometric intuition for proof:

- Nothing changes if we replace $\mathcal{G}$ with $\mathcal{A} = \mathcal{G} + \mathbb{R}^2_+$, which is a **convex set**



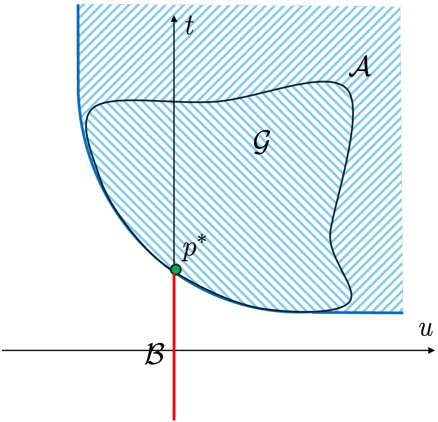
Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \dots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

Geometric intuition for proof:

- Define another convex set $\mathcal{B}$ with $\mathcal{A} \cap \mathcal{B} = \emptyset$



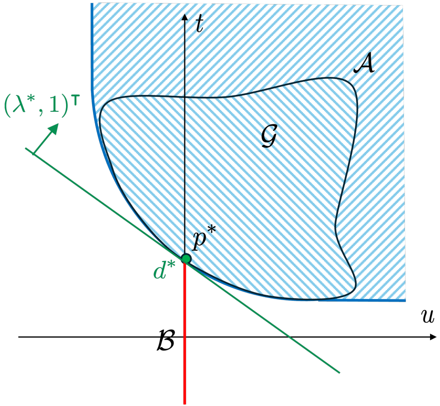
Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

Geometric intuition for proof:

- The Separating Hyperplane Theorem will give us the optimal λ^* and $p^* = d^*$



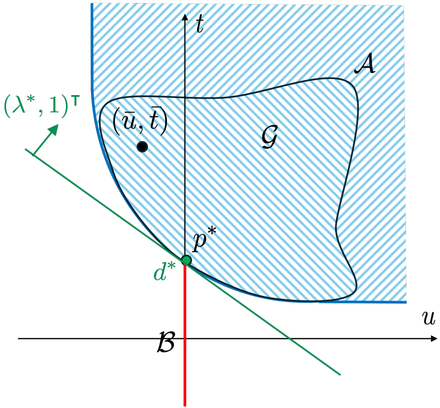
Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

Geometric intuition for proof:

- The Slater point will guarantee that the hyperplane is not vertical



Strong Duality in Convex Optimization

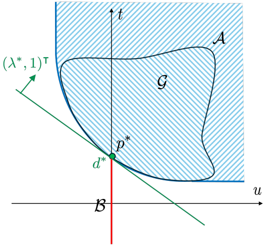
Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

- Define the set

$$\mathcal{A} = \big\{(u,t) \in \mathbb{R}^m \times \mathbb{R} : \exists x \in X,\\ t \geq f_0(x), u_i \geq f_i(x),\ i = 1, \ldots, m\big\}.$$

- $\mathcal{A}$ is convex. *Why?*



Strong Duality in Convex Optimization

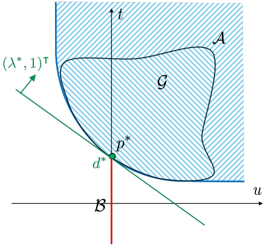
Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

- Define the set

$$\mathcal{A} = \{(u, t) \in \mathbb{R}^m \times \mathbb{R} : \exists x \in X, \\ t \geq f_0(x), u_i \geq f_i(x),\ i = 1, \ldots, m\}.$$

- $\mathcal{A}$ is convex. *Why?*
- Define the convex set $\mathcal{B} = \{(0, s) \in \mathbb{R}^m \times \mathbb{R} \mid s < p^\star\}$
- Claim: $\mathcal{A} \cap \mathcal{B} = \emptyset$. *Why?*



Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

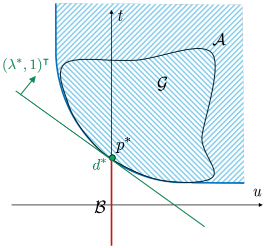
Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

- Define the set

$$\mathcal{A} = \big\{(u,t) \in \mathbb{R}^m \times \mathbb{R} : \exists\, x \in X, \\ t \geq f_0(x), u_i \geq f_i(x),\ i = 1, \ldots, m\big\}.$$

- $\mathcal{A}$ is convex. *Why?*
- Define the convex set $\mathcal{B} = \{(0,s) \in \mathbb{R}^m \times \mathbb{R} \mid s < p^\star\}$
- Claim: $\mathcal{A} \cap \mathcal{B} = \emptyset$. *Why?*
- Separating Hyperplane Theorem:

$$\exists\, (\lambda, \mu) \in \mathbb{R}^{m+1},\ \ b \in \mathbb{R} : \begin{cases} (\lambda, \mu) \neq 0, \\ \lambda^\intercal u + \mu t \geq b, \forall\, (u,t) \in A \\ \lambda^\intercal u + \mu t \leq b, \forall\, (u,t) \in B. \end{cases}$$



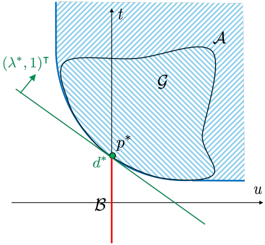
Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

- Separating Hyperplane Theorem:

$$\exists\, (\lambda, \mu) \in \mathbb{R}^{m+1}, \quad b \in \mathbb{R} : \begin{cases} (1) & (\lambda, \mu) \neq 0, \\ (2) & \lambda^\intercal u + \mu t \geq b, \forall\, (u, t) \in A \\ (3) & \lambda^\intercal u + \mu t \leq b, \forall\, (u, t) \in B. \end{cases}$$



- (2) implies $\lambda \geq 0$ and $\mu \geq 0$.
  Otherwise, $\inf_{(u,t)\in\mathcal{A}}(\lambda^\intercal u + \mu t) = -\infty$ so $\not\geq b$ (*Why?*)

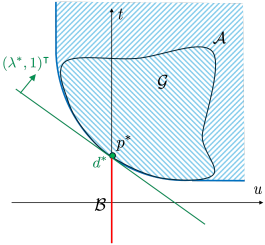
Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

- Separating Hyperplane Theorem:

$$\exists\, (\lambda, \mu) \in \mathbb{R}^{m+1}, \quad b \in \mathbb{R} : \begin{cases} (1) \quad (\lambda, \mu) \neq 0, \\ (2) \quad \lambda^\intercal u + \mu t \geq b, \forall\, (u,t) \in A \\ (3) \quad \lambda^\intercal u + \mu t \leq b, \forall\, (u,t) \in B. \end{cases}$$



- (2) implies $\lambda \geq 0$ and $\mu \geq 0$.
  Otherwise, $\inf_{(u,t)\in\mathcal{A}}(\lambda^\intercal u + \mu t) = -\infty$ so $\not\geq b$ (*Why?*)
- (3) simplifies to $\mu t \leq b$ for all $t < p^\star$, so $\mu p^\star \leq b$.
- We found $\lambda \geq 0, \mu \geq 0$:

$$(4)\ \mathcal{L}(x, \lambda) := \sum_{i=1}^{m} \lambda_i f_i(x) + \mu f_0(x) \geq b \geq \mu p^\star, \forall\, x \in X$$

Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

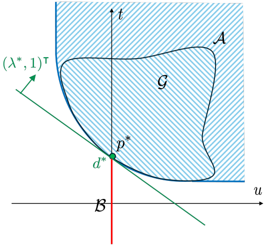
- Separating Hyperplane Theorem:

$$\exists\,(\lambda,\mu) \in \mathbb{R}^{m+1}, \quad b \in \mathbb{R} : \begin{cases} (1) \quad (\lambda,\mu) \neq 0, \\ (2) \quad \lambda^\intercal u + \mu t \geq b, \forall\,(u,t) \in A \\ (3) \quad \lambda^\intercal u + \mu t \leq b, \forall\,(u,t) \in B. \end{cases}$$

- We found $\lambda \geq 0, \mu \geq 0$:

$$(4)\ \mathcal{L}(x,\lambda) := \sum_{i=1}^{m} \lambda_i f_i(x) + \mu f_0(x) \geq b \geq \mu p^\star,\ \forall\, x \in X$$

- **Case 1.** $\mu > 0$ (non-vertical hyper-plane)



Strong Duality in Convex Optimization

Theorem (Strong Duality in Convex Optimization)

Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

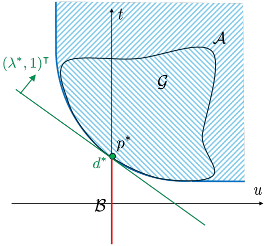
- Separating Hyperplane Theorem:

$$\exists\,(\lambda, \mu) \in \mathbb{R}^{m+1}, \quad b \in \mathbb{R} : \begin{cases} (1) \quad (\lambda, \mu) \neq 0, \\ (2) \quad \lambda^\intercal u + \mu t \geq b, \forall\,(u, t) \in A \\ (3) \quad \lambda^\intercal u + \mu t \leq b, \forall\,(u, t) \in B. \end{cases}$$

- We found $\lambda \geq 0, \mu \geq 0$:

$$(4)\ \mathcal{L}(x, \lambda) := \sum_{i=1}^{m} \lambda_i f_i(x) + \mu f_0(x) \geq b \geq \mu p^\star, \forall\, x \in X$$

- **Case 1.** $\mu > 0$ (non-vertical hyper-plane)
- Divide (4) by μ to get: $\mathcal{L}(x, \lambda/\mu) \geq p^\star, \forall\, x \in X$.
- This implies $g(\lambda/\mu) \geq p^\star$
- Weak duality: $g(\lambda/\mu) \leq p^\star$, so $g(\lambda/\mu) = p^\star$
- Strong duality holds and the dual optimum is attained



Strong Duality in Convex Optimization

Strong Duality in Convex Optimization

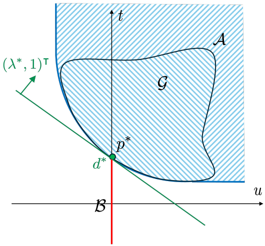
Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

- Separating Hyperplane Theorem:

$$\exists\, (\lambda, \mu) \in \mathbb{R}^{m+1}, \;\; b \in \mathbb{R} : \begin{cases} (1) & (\lambda, \mu) \neq 0, \\ (2) & \lambda^\intercal u + \mu t \geq b, \forall\, (u,t) \in A \\ (3) & \lambda^\intercal u + \mu t \leq b, \forall\, (u,t) \in B. \end{cases}$$

- We found $\lambda \geq 0, \mu \geq 0$:

$$(4)\ \mathcal{L}(x, \lambda) := \sum_{i=1}^{m} \lambda_i f_i(x) + \mu f_0(x) \geq b \geq \mu p^\star, \forall\, x \in X$$



- **Case 2.** $\mu = 0$ (vertical hyperplane)
- (4) implies $\sum_{i=1}^{m} \lambda_i f_i(x) \geq 0, \forall\, x \in X$

Strong Duality in Convex Optimization

Strong Duality in Convex Optimization

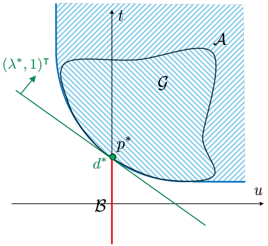
Let $X \subset \mathbb{R}^n$ be convex and $f_0, f_1, \ldots, f_m : X \to \mathbb{R}$ convex functions on X satisfying the Slater condition on X. Then, $p^* = d^*$ and the dual attains its optimal value.

- Separating Hyperplane Theorem:

$$\exists\, (\lambda, \mu) \in \mathbb{R}^{m+1}, \;\; b \in \mathbb{R} : \begin{cases} (1) & (\lambda, \mu) \neq 0, \\ (2) & \lambda^\intercal u + \mu t \geq b, \forall\, (u, t) \in A \\ (3) & \lambda^\intercal u + \mu t \leq b, \forall\, (u, t) \in B. \end{cases}$$

- We found $\lambda \geq 0, \mu \geq 0$:

$$(4)\; \mathcal{L}(x, \lambda) := \sum_{i=1}^{m} \lambda_i f_i(x) + \mu f_0(x) \geq b \geq \mu p^\star, \forall\, x \in X$$



- **Case 2.** $\mu = 0$ (vertical hyperplane)
- (4) implies $\sum_{i=1}^{m} \lambda_i f_i(x) \geq 0, \forall\, x \in X$
- $\bar{x}$ satisfies Slater condition $\Rightarrow f_i(\bar{x}) < 0$ for $i = 1, \ldots, m$
- This together with $\lambda \geq 0 \Rightarrow \lambda = 0$
- Contradicts (1) that $(\lambda, \mu) \neq 0$.

Explicit Equality Constraints

- In applications, useful to make the **equality constraints explicit**:

$$\begin{aligned} \text{minimize}_{x \in \mathcal{X}} \ & f_0(x) \\ \text{subject to } & f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & Ax = b. \end{aligned}$$

  where $f_i, i = 0, \dots, m$ are convex and $A \in \mathbb{R}^{p \times n}$ has rank p.

Explicit Equality Constraints

- In applications, useful to make the **equality constraints explicit**:

$$\begin{aligned} \text{minimize}_{x \in X} \ & f_0(x) \\ \text{subject to } & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ & Ax = b. \end{aligned}$$

  where $f_i, i = 0, \ldots, m$ are convex and $A \in \mathbb{R}^{p \times n}$ has rank p.

- With $\nu \in \mathbb{R}^p$ denoting Lagrange multipliers for $Ax = b$, Lagrangian is:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^\mathsf{T}(Ax - b),$$

Explicit Equality Constraints

- In applications, useful to make the **equality constraints explicit**:

$$
\begin{aligned}
\text{minimize}_{x \in X} \; & f_0(x) \\
\text{subject to } & f_i(x) \leq 0, \quad i = 1, \dots, m, \\
& Ax = b.
\end{aligned}
$$

  where $f_i, i = 0, \dots, m$ are convex and $A \in \mathbb{R}^{p \times n}$ has rank p.

- With $\nu \in \mathbb{R}^p$ denoting Lagrange multipliers for $Ax = b$, Lagrangian is:

$$
\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^\mathsf{T}(Ax - b),
$$

- With $g(\lambda, \nu) := \inf_{x \in X} \mathcal{L}(x, \lambda, \nu)$, the dual problem becomes:

$$
\begin{aligned}
\text{maximize } & g(\lambda, \nu) \\
\text{subject to } & \lambda \geq 0.
\end{aligned}
$$

  No sign constraints on ν!

Nonlinear Farkas Lemma

Proposition (Nonlinear Farkas Lemma)

Let $X \subset \mathbb{R}^n$ be convex, let $f_0, f_1, \ldots, f_m$ be real-valued convex functions on X, and assume $f_1, \ldots, f_m$ satisfy the Slater condition on X.

Then, the following system of inequalities has a solution

$$\exists\, x \;:\; f_0(x) < z, \quad f_j(x) \leq 0,\; j = 1, \ldots, m, \quad x \in X,$$

if and only if the following system has no solution:

$$\exists\, \lambda \;:\; \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j g_j(x) \right] \geq z, \;\; \lambda_j \geq 0,\; j = 1, \ldots, m.$$

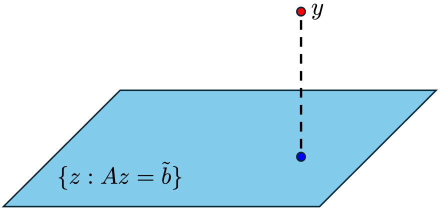
- Mirrors arguments used in strong duality proof

Minimum Euclidean Distance Problem

- Given $y \in \mathbb{R}^n$ and affine set $\{z : Az = \tilde{b}\}$
- $A \in \mathbb{R}^{p \times n}$, $\tilde{b} \in \mathbb{R}^p$ has rank p

$$\min_z \left\{ \|z - y\|_2^2 \ : \ Az = \tilde{b} \right\}$$

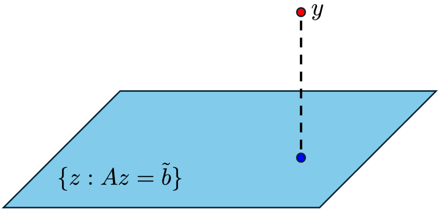
- *What is the optimal value p^*?*



Minimum Euclidean Distance Problem

- Given $y \in \mathbb{R}^n$ and affine set $\{z : Az = \tilde{b}\}$
- $A \in \mathbb{R}^{p \times n}$, $\tilde{b} \in \mathbb{R}^p$ has rank p

$$\min_z \left\{ \|z - y\|_2^2 \ : \ Az = \tilde{b} \right\}$$



Quadratic Programs - Preliminaries

Unconstrained Quadratic Program

For $Q = Q^\intercal$, consider the following unconstrained problem:

$$\min \; f(x) := \frac{1}{2} x^\intercal P x + q^\intercal x$$

- *What is the optimal value p^*?*

Quadratic Programs - Preliminaries

Unconstrained Quadratic Program

For $Q = Q^\intercal$, consider the following unconstrained problem:

$$\min \ f(x) := \frac{1}{2}x^\intercal P x + q^\intercal x$$

- *What is the optimal value p^*?*

$$\nabla_x f(x) = 0 \quad \Leftrightarrow \quad Px = -q$$

$$p^* = \begin{cases} -\frac{1}{2}q^\intercal P^\dagger q & \text{if } P \succeq 0 \text{ and } q \in \mathcal{R}(P) \\ -\infty & \text{otherwise.} \end{cases}$$

- $P^\dagger$ is the (Moore-Penrose) pseudo-inverse of P
- For A with singular value decomposition $A = U\Sigma V^\intercal$, $A^\dagger := V\Sigma^{-1}U^\intercal$
- Equals $(A^\intercal A)^{-1}A^\intercal$ if $\text{rank}(A) = n$ and $A^\intercal(AA^\intercal)^{-1}$ if $\text{rank}(A) = m$

QPs and QCQPs

Quadratic Programs

A **Quadratic Program (QP)** is an optimization problem of the form:

$$\min \frac{1}{2}x^\intercal P x + c^\intercal x$$
$$A_1 x = b_1$$
$$A_2 x \leq b_2$$

where $P = P^\intercal$.

QPs and QCQPs

Quadratic Programs

A **Quadratic Program (QP)** is an optimization problem of the form:

$$\begin{aligned} \min\ & \frac{1}{2}x^\intercal P x + c^\intercal x \\ & A_1 x = b_1 \\ & A_2 x \leq b_2 \end{aligned}$$

where $P = P^\intercal$.

Quadratically Constrained Quadratic Programs

A **Quadratically Constrainted Quadratic Program (QCQP)** is a problem:

$$\begin{aligned} \min\ & \frac{1}{2}x^\intercal P_0 x + c^\intercal x \\ & x^\intercal P_i x + q_i^\intercal x + b_i \leq 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where $Q_i, i = 0, \ldots, m$ are **symmetric** matrices.

Convex if $P \succeq 0, P_i \succeq 0$. Gurobi can now handle **non-convex** QCQPs!

Convex QP With Inequality Constraints

QP with Inequality Constraint

$$\underset{x}{\text{minimize}} \ \frac{1}{2}x^\intercal Q x + c^\intercal x$$
$$Ax \leq b$$

where $Q \succ 0$ is a **positive definite** matrix.

- *What is the Lagrangian? What is the dual? Does Slater Condition hold?*

Convex QP With Inequality Constraints

QP with Inequality Constraint

$$\underset{x}{\text{minimize}}\ \frac{1}{2}x^\intercal Qx + c^\intercal x$$
$$Ax \leq b$$

where $Q \succ 0$ is a **positive definite** matrix.

- *What is the Lagrangian? What is the dual? Does Slater Condition hold?*

Convex QCQP

QCQP

$$\begin{aligned} \text{minimize} \ & \frac{1}{2}x^\intercal P_0 x + q_0^\intercal x + r_0 \\ \text{subject to} \ & \frac{1}{2}x^\intercal P_i x + q_i^\intercal x + r_i \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $P_0 \succ 0$ and $P_i \succeq 0$

- *What is the Lagrangian? What is the dual? Does Slater Condition hold?*

A **Non-Convex** QCQP

A Special Non-Convex QCQP

For $A = A^\intercal$ and $A \not\succeq 0$, consider:

$$\begin{aligned} \text{minimize } & x^\intercal A x + 2b^\intercal x \\ & x^\intercal x \leq 1 \end{aligned}$$

- *What is the Lagrangian? What is the dual? Does Slater Condition hold?*

A **Non-Convex** QCQP

A Special Non-Convex QCQP

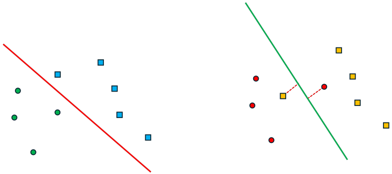
For $A = A^\intercal$ and $A \not\succeq 0$, consider:

$$\begin{aligned} \text{minimize } & x^\intercal A x + 2b^\intercal x \\ & x^\intercal x \leq 1 \end{aligned}$$

- Slater condition trivially satisfied!
- We actually have **zero duality gap**, $p^* = d^*$!
- A more general result: strong duality for any quadratic optimization problem with two constraints $\ell \leq x^\intercal P x \leq u$ if P and A are simultaneously diagonalizable

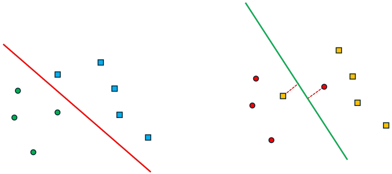
Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

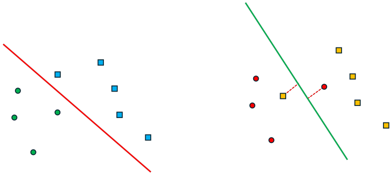


- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases}$$

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



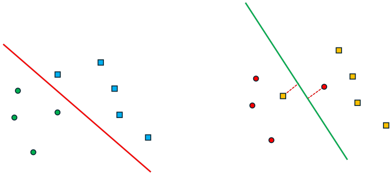
- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if} \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m.$$

- *How to solve this problem?*

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m.$$

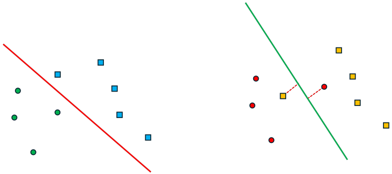
- *How to solve this problem?* This is **an LP**!

- In practice, non-separable. Find hyperplane minimizing total classification errors:

$$\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b)), \text{ where } \psi(t) = 1 \text{ if } t < 0 \text{ and } 0 \text{ otherwise.}$$

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m.$$

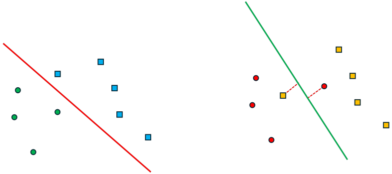
- *How to solve this problem?* This is **an LP**!

- In practice, non-separable. Find hyperplane minimizing total classification errors:

$$\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b)), \text{ where } \psi(t) = 1 \text{ if } t < 0 \text{ and } 0 \text{ otherwise.}$$

- **Hard (MIP) problem!**

Regularized Support Vector Machines (SVM)



- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

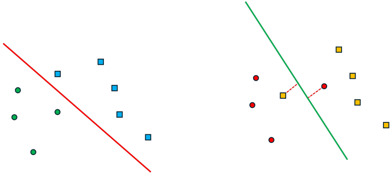
  - Separable if and only if $y_i(w^\intercal x_i + b) \geq 0,\ i = 1, \ldots, m$.

  - Minimize $\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b))$, where $\psi(t) = 1$ if $t < 0$ and 0 : **hard MIP!**

  - Replace $\psi(t)$ with upper bound $h(t) = (1 - t)_+ = \max(0, 1 - t)$ (**hinge function**)

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable if and only if $y_i(w^\intercal x_i + b) \geq 0, \ \ i = 1, \dots, m$.

- Minimize $\sum_{i=1}^m \psi(y_i(w^\intercal x_i + b))$, where $\psi(t) = 1$ if $t < 0$ and 0 : **hard MIP!**

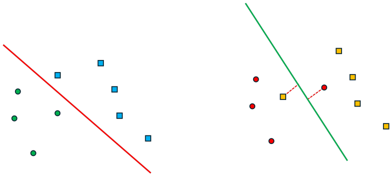
- Replace $\psi(t)$ with upper bound $h(t) = (1-t)_+ = \max(0, 1-t)$ (**hinge function**)

- Solve **regularized** version:

$$\min_{w,b} \ C \cdot \sum_{i=1}^m (1 - y_i(w^\intercal x_i + b))_+ + \frac{1}{2}\|w\|_2^2,$$

where parameter $C > 0$ controls trade-off between robustness and performance

Regularized Support Vector Machines (SVM)



- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

- Solve $\min_{w,b} \; C \cdot \sum_{i=1}^{m} (1 - y_i(w^\intercal x_i + b))_+ + \frac{1}{2}\|w\|_2^2$

- Can be written as a QP by introducing slack variables:

$$\min_{w,b,v} \; \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{m} v_i \quad : \quad v \geq 0, \; y_i(w^\intercal x_i + b) \geq 1 - v_i, \; i = 1, \ldots, m,$$

  or more compactly:

$$\min_{w,b,v} \; \frac{1}{2}\|w\|_2^2 + C \cdot 1^\intercal v \quad : \quad v \geq 0, \; v + Z^\intercal w + by \geq 1,$$

  where $Z^\intercal \in \mathbb{R}^{m \times n}$ is the matrix with rows given by $y_i \cdot x_i^\intercal$

- *What is the Lagrangian? What is the dual? Does Slater Condition hold?*

Regularized Support Vector Machines (SVM)

- Solve
$$\min_{w,b,v} \ \frac{1}{2}\|w\|_2^2 + C \cdot 1^\mathsf{T} v \quad : \quad v \geq 0,\ v + Z^\mathsf{T} w + by \geq 1,$$
where $Z^\mathsf{T} \in \mathbb{R}^{m \times n}$ is the matrix with rows given by $y_i \cdot x_i^\mathsf{T}$
- $\mathcal{L}(w, b, \lambda, \mu) = \frac{1}{2}\|w\|_2^2 + C \cdot v^\mathsf{T} 1 + \lambda^\mathsf{T}(1 - v - Z^\mathsf{T} w - by) - \mu^\mathsf{T} v$
- $g(\lambda, \mu) = \min_{w,b} \mathcal{L}(w, b, \lambda, \mu)$
- Taking gradients : $w(\lambda, \mu) = Z\lambda,\ C \cdot 1 = \lambda + \mu,\ \lambda^\mathsf{T} y = 0$
- We obtain
$$g(\lambda, \mu) = \begin{cases} \lambda^\mathsf{T} 1 - \frac{1}{2}\|Z\lambda\|_2^2 & \text{if } \lambda^\mathsf{T} y = 0,\ \lambda + \mu = C \cdot 1, \\ +\infty & \text{otherwise.} \end{cases}$$
- Dual problem
$$d^\star = \max_\lambda \left\{ \lambda^\mathsf{T} 1 - \frac{1}{2}\lambda^\mathsf{T} Z^\mathsf{T} Z \lambda \quad : \quad 0 \leq \lambda \leq C \cdot 1,\ \lambda^\mathsf{T} y = 0 \right\}.$$
- Strong duality holds, because the primal problem is a QP
- Dual objective depends only on the **kernel matrix** $K = Z^\mathsf{T} Z \in S_+^m$, and dual problem involves only m variables and $m + 1$ constraints
- Only dependence on the number of dimensions (features) n is through Z, requiring all products $x_i^\mathsf{T} x_j$, $1 \leq i \leq j \leq m$