

Learning objectives. By the end of class, you should be able to

1. use infeasibility certificates to prove that a feasible point is globally optimal;
2. derive weak duality from weak minimax, and explain why the dual function is always concave;
3. derive the dual problem, and use the primal–dual gap to certify global optimality or bound global suboptimality.

1 Recap: Weak alternative

Original problem; call it the **primal problem**

$$(P) \quad \text{find } x \in \mathbb{R}^n \text{ s.t. } f_i(x) \leq 0 \quad \text{for all } i \in \{1, \dots, m\}.$$

Certify infeasibility by solving the **dual problem**

$$(D) \quad \text{find } y \in \mathbb{R}_+^m \text{ s.t. } g(y) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, y) > 0.$$

Equivalently, maximize g over $y \in \mathbb{R}_+^m$ and check

$$\sup_{y \in \mathbb{R}_+^m} g(y) > 0.$$

The dual function g is *always* concave, even when $f_i(x)$ nonconvex.

Indeed, for $0 < \theta < 1$,

$$\begin{aligned} g(\theta y + (1 - \theta)z) &= \inf_x \{ \theta \mathcal{L}(x, y) + (1 - \theta) \mathcal{L}(x, z) \} \\ &\geq \theta \inf_x \{ \mathcal{L}(x, y) \} + (1 - \theta) \inf_x \{ \mathcal{L}(x, z) \} \\ &= \theta g(y) + (1 - \theta)g(z). \end{aligned}$$

Maximizing the dual function is *always* a convex optimization problem, although evaluating the function can be difficult

Question 1. Why is it called “weak” alternative?

Example 1. Certify that the following set is infeasible

$$\Omega = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1, x_1 + x_2 \geq 2\}.$$

2 Today: Optimality certification

Now we know how to certify that a problem has *no feasible solution*. Today, use it to certify when a feasible solution is globally optimal.

Example 2. Obviously, $x = 1$ is globally optimal below

$$\inf_{x \in \mathbb{R}} x^2 \quad \text{s.t. } x \geq 1.$$

Prove it by certifying that the *parameterized* feasibility set

$$\Omega_t = \{x \in \mathbb{R} : x^2 \leq t, 1 - x \leq 0\}$$

is infeasible for all $t < 1$.

Basic idea: No better solution exists, so $x = 1$ is globally optimal.

The ability to certify global optimality to a nonconvex problem is critically used in electric grid dispatch [Lavaei & Low, *IEEE TPWRS*, 2012], robot navigation [Carlone & Dellaert, *ICRA*, 2015], neural network verification [Chiu & Zhang, *ICML*, 2023], power system state estimation [Guzel & Zhang, *IEEE TCNS*, 2025].

Theorem 3. Global optimality certificate. Let $\hat{x} \in \mathbb{R}^n$ be *feasible* for the following

$$\inf_{x \in \mathbb{R}^n} f_0(x) \text{ s.t. } f_i(x) \leq 0 \text{ for all } i \in \{1, \dots, m\}.$$

If there exists corresponding $\hat{y} \in \mathbb{R}_+^m$ such that

$$f_0(\hat{x}) = g(\hat{y}),$$

where the dual function is written

$$g(y) = \inf_{x \in \mathbb{R}^n} \left\{ f_0(x) + \sum_{i=1}^m y_i f_i(x) \right\},$$

then $\hat{x}$ is globally optimal, and $\hat{y}$ is the associated certificate.

Basic idea: No better solution exists, so $\hat{x}$ is globally optimal.

Question 2. Checking global optimality is supposed to be NP-hard. What is the catch? How does Theorem 3 fail?

Remedy: construct a dual that can be evaluated in *closed-form*, and then prove *strong duality* (next lecture) to ensure that an arbitrarily good $\hat{y}$ exists.

Proof of Theorem 3 is by certifying that the *parameterized* feasibility set

$$\Omega_t = \{x \in \mathbb{R}^n : f_0(x) \leq t, f_1(x) \leq 0, \dots, f_m(x) \leq 0\}.$$

is infeasible for all $t < f_0(\hat{x})$.

3 Weak Lagrangian duality

In fact, everything we have learned so far are just special cases of the following foundational result. Note: it does not assume convexity, smoothness, coercivity, etc.

Theorem 4. Given $f_0, f_1, \dots, f_m$, we have $p^* \geq d^*$ where

$$p^* = \inf_{x \in \mathbb{R}^n} f_0(x) \text{ s.t. } f_i(x) \leq 0 \text{ for all } i \in \{1, \dots, m\}$$

$$d^* = \sup_{y \in \mathbb{R}_+^m} g(y) \text{ where } g(y) = \inf_{x \in \mathbb{R}^n} \underbrace{\left\{ f_0(x) + \sum_{i=1}^m y_i f_i(x) \right\}}_{L(x,y)}$$

In particular, any feasible $\hat{x} \in \Omega$ and any multiplier $\hat{y} \in \mathbb{R}_+^m$ satisfy

$$f_0(\hat{x}) \geq \inf_{x \in \Omega} f_0(x) \geq \sup_{y \in \mathbb{R}_+^m} g(y) \geq g(\hat{y}).$$

Certify the *amount* of global suboptimality for $\hat{x} \in \Omega$ by

$$f_0(\hat{x}) - \inf_{x \in \Omega} f_0(x) \leq \underbrace{f_0(\hat{x}) - g(\hat{y})}_{\text{duality gap}}.$$

Previous theorem is just special case when we find $\hat{y}$ with **zero duality gap**, i.e. $f_0(\hat{x}) = g(\hat{y})$.

In practice, may suffice if we can establish **small duality gap**, e.g. $|f_0(\hat{x}) - g(\hat{y})| = O(\sigma)$ for noise variance σ^2 .

Question 3. Why is it “weak” duality?

The proof is surprisingly simple.

Lemma 5. Weak minimax. For any function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}_+^m \rightarrow \mathbb{R}$, we have

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}_+^m} \mathcal{L}(x, y) \geq \sup_{y \in \mathbb{R}_+^m} \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, y)$$

Lemma 6. Primal form.

$$\begin{aligned} & \inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}_+^m} \left\{ f_0(x) + \sum_{i=1}^m y_i f_i(x) \right\} \\ &= \inf_{x \in \mathbb{R}^n} f_0(x) \text{ s.t. } f_i(x) \leq 0 \text{ for all } i \in \{1, \dots, m\} \end{aligned}$$

4 Application: Infeasibility

Question 4. Take the primal objective to be zero.

$$p^* = \inf_{x \in \mathbb{R}^n} 0 \text{ s.t. } f_i(x) \leq 0 \text{ for all } i \in \{1, \dots, m\}.$$

What is the dual objective d^* ?

Question 5. What happens when $d^* > 0$?

Question 6. What happens when $p^* = 0$?

We have in fact re-proved the duality form of the weak alternative.

Theorem 7. Weak alternative, duality form. At most one of the following is feasible

(P) find $x \in \mathbb{R}^n$ s.t. $f_i(x) \leq 0$ for all i

(D) find $y \in \mathbb{R}_+^m$ s.t. $g(y) = \inf_{x \in \mathbb{R}^n} \sum_{i=1}^m y_i f_i(x) > 0$.

5 Application: QCQPs

The following dual certificate for the *quadratically-constrained quadratic program* is used extensively in the literature.

Theorem 8. Given matrices $C, A_1, \dots, A_m$ and radius $R > 0$, define the feasible set

$$\Omega = \{x \in \mathbb{R}^n : \|x\| \leq R, x^T A_i x \leq b_i \text{ for all } i\}$$

Then, we have

$$\inf_{x \in \Omega} x^T C x \geq \sup_{y \in \mathbb{R}_+^m} -b^T y - \delta(y)$$

where

$$\delta(y) = R^2 \left[-\lambda_{\min}(C + \sum_{i=1}^m y_i A_i) \right]_+$$

The point is that any arbitrary choice of $y \in \mathbb{R}_+^m$ can yield a lower-bound on the global optimum. If it is sufficiently close, then it certifies approximate global optimality.

Note that interesting QCQPs are nonconvex.

Question 7. Suppose that the QCQP is convex, so $C \succeq 0$ and $A_i \succeq 0$. When is it feasible, and what is the optimal duality gap?

The proof is a standard application of weak duality.

But how to choose good $y \in \mathbb{R}_+^m$?

In practice, given feasible $\hat{x} \in \Omega$ with $\|\hat{x}\| < R$, a common strategy is to solve the KKT equations (next week) for a good choice of y :

$$\min_{y \in \mathbb{R}_+^m} \left\| \left(C + \sum_i y_i A_i \right) \hat{x} \right\|^2 + \sum y_i^2 (\hat{x}^T A_i \hat{x} - b_i)^2$$

which is a convex nonnegative linear least-squares problem. If $\hat{y} \in \mathbb{R}_+^m$ attains zero loss in the least-squares, so that

$$\left(C + \sum_i \hat{y}_i A_i \right) \hat{x} = 0, \quad \hat{y}_i (b_i - \hat{x}^T A_i \hat{x}) = 0,$$

then the duality gap further simplified

$$\begin{aligned} & \hat{x}^T C \hat{x} + b^T \hat{y} + \delta(\hat{y}) \\ &= \hat{x}^T \left(C + \sum_i \hat{y}_i A_i \right) \hat{x} + \sum_i \hat{y}_i (b_i - \hat{x}^T A_i \hat{x}) + \delta(\hat{y}) \\ &= \delta(\hat{y}). \end{aligned}$$

In special applications, it is possible to show that if $\hat{x}$ is globally optimal, then the least-squares $\hat{y}$ is unique and yields $\delta(\hat{y}) = 0$. For these special cases, global optimality can always be certified.

6 Summary

Every constrained optimization problem has a dual

$$p^* = \inf_{x \in \mathbb{R}^n} f_0(x) \text{ s.t. } f_i(x) \leq 0 \text{ for all } i \in \{1, \dots, m\}$$

$$d^* = \sup_{y \in \mathbb{R}_+^m} g(y) \text{ where } g(y) = \inf_{x \in \mathbb{R}^n} \underbrace{\left\{ f_0(x) + \sum_{i=1}^m y_i f_i(x) \right\}}_{L(x,y)}$$

Weak duality further says

$$g(\hat{y}) \leq d^* \leq p^* \leq f_0(\hat{x})$$

for every feasible $\hat{x} \in \Omega$ and $\hat{y} \in \mathbb{R}_+^m$.

Therefore, the primal–dual gap certifies global suboptimality:

$$0 \leq f_0(\hat{x}) - p^* \leq f_0(\hat{x}) - g(\hat{y}).$$

A zero primal–dual gap certifies global optimality.

Infeasibility and optimality are connected. Set $f_0(x) = 0$; then $g(y) > 0$ certifies infeasibility $\Omega = \emptyset$. Hence, weak alternatives is just a special case of weak duality.

For QCQPs, the dual lower bound can be explicitly computed using minimum eigenvalues:

$$-b^T y - R^2 \left[-\lambda_{\min} \left(C + \sum_i y_i A_i \right) \right]_+.$$

Next lecture:

- Strong duality. When does strong duality $p^* = d^*$ hold?
- Primal attainment. When do we have $x_* \in \Omega$ such that $p^* = f_0(x_*)$.
- Dual attainment. When do we have $y_* \in \mathbb{R}_+^m$ such that $d^* = g(y_*)$?

Classic answers: Linear objectives and constraints, or convex objectives and constraints + primal-dual Slater's conditions.