Final exam: July 24 1:30-4:30Pm
Shaw F301

MAT3007 Optimization
Lecture 16 Convexity
Lagrangian Relaxation

Yuang Chen

School of Data Scienc
The Chinese University of Hong Kong, Shenzhen

July 10, 2025

Outline

Outline

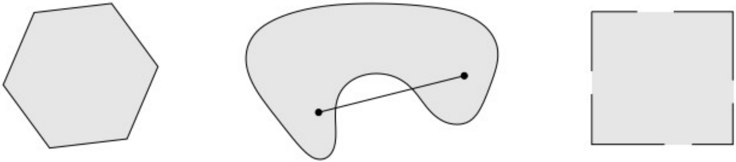
1. **Review Convex Set and Function**
2. Convex Optimization
3. Lagrangian Relaxation
4. Duality Theories
5. KKT Conditions

Convex Set

Definition

A set $\mathcal{X}$ is **convex** if $\forall x, y \in \mathcal{X}$ such that $\lambda x + (1-\lambda)y \in \mathcal{X}, \forall \lambda \in [0,1]$.

- Convex set contains line segment between any two points in the set
- Example:



Convex Function

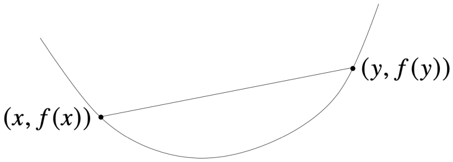
Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is a **convex** function iff the $\text{dom}(f)$ is a convex set and

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

for all $x, y \in \text{dom}(f)$ and $\lambda \in [0, 1]$.
f is **concave** if $-f$ is convex.
f is **strictly** convex iff the $\text{dom}(f)$ is a convex set and

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$



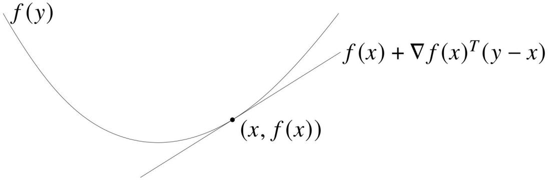
First-order Condition of Convex Functions

First-order condition of convex functions

A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^\mathsf{T}(y - x)$$

for all $x, y \in \mathrm{dom}(f)$.



Second-order Condition of Convex Functions

Second-order condition of convex functions

A twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ with convex domain is convex iff

$$\nabla^2 f(x) \succeq 0 \quad \text{(the Hessian is positive semi-definite)}$$

for all $x \in \text{dom}(f)$.

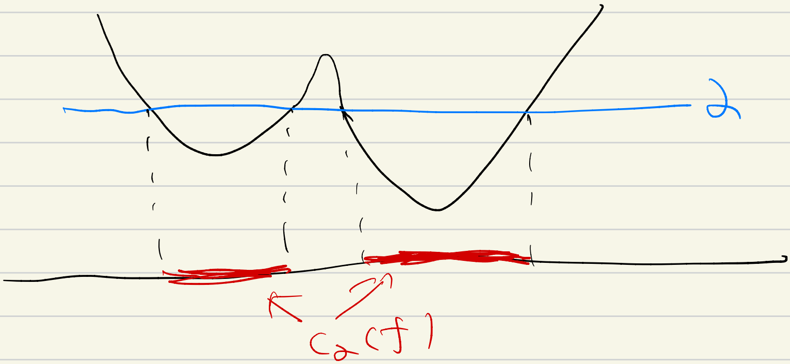
Positive Semi-definite

A $n \times n$ matrix A is positive semi-definite if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$. Equivalently, A is positive semi-definite if all its eigenvalues are nonnegative.

Remark: Concave function has a negative semi-definite Hessian.

α - Sublevel Set

$$C_\alpha(f) = \{x : f(x) \le \alpha\}$$



Prove: $C_\alpha(f)$ is convex set if f is convex function.

$\forall x, y \in C_\alpha(f), \quad \forall \lambda \in [0,1]$

$\hookrightarrow \begin{cases} f(x) \le \alpha \\ f(y) \le \alpha \end{cases}$

Since f is convex,

$$\begin{aligned} f(\lambda x + (1-\lambda) y) &\le \lambda f(x) + (1-\lambda) f(y) \\ &\le \lambda \alpha + (1-\lambda) \alpha \\ &= \alpha \end{aligned}$$

$\Rightarrow \lambda x + (1-\lambda) y \in C_\alpha(f)$

So, $C_\alpha(f)$ is convex

Remark: The converse is wrong.

Outline

Convex optimization

min. convex $\Leftrightarrow$ max. concave

$$\begin{aligned} \text{min.} \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

If f is a convex function and X is a convex set, then the problem is a convex optimization problem. (Minimize a convex function is equivalent to maximize a concave function.) Otherwise, we call it non-convex optimization problem

Examples:

- o $\min\{x^2 : -1 \le x \le 1\}$ is a convex opt. problem.
- o $\min\{-x^2 : -1 \le x \le 1\}$ is not a convex opt. problem.
- o $\min\{x^2 : -5 \le x \le 5, x \in \mathbb{Z}\}$ is not a convex opt. problem.

Recognizing Convex Optimization Problem

convex constraint

g is convex, $g(x) \le b$

g is concave, $g(x) \ge b$

① convex

$$\begin{aligned} \min. \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

② convex

③ linear

Suppose X is in the following format

$$X = \{g_i(x) \le b_i \ \ \forall i \in I, \quad h_j(x) = d_j \ \ \forall j \in J\},$$

if f is convex, g_i is convex for all $i \in I$ and h_j is affine for all $j \in J$, then the problem is a convex optimization problem.

Checking Convexity

- Check that all variables are continuous
- Check that the objective function is convex
- Check each equality constraint to see if it is linear
- Write each constraint as an inequality constraint in $\leq$ form with a constant on the right-hand-side, and check the convexity of the function on the left-hand-side

If it passes all the checks then you have a convex optimization problem. Otherwise, it may or may not be convex (the conditions are sufficient, not necessary).

Convex Optimization Example

$$\begin{aligned} \min. \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & \frac{x_1}{1 + x_2^2} \le 0 \\ & (x_1 + x_2)^2 = 0 \end{aligned}$$

Convex

not Convex in x_2

not linear

the problem is equivalent to $\Updownarrow$

$$\begin{aligned} \min. \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 \le 0 \\ & x_1 + x_2 = 0 \end{aligned}$$

Convex

linear

Example

min. $-xyz$

$$
\begin{aligned}
\max. \quad & xyz \\
\text{s.t.} \quad & z^2 - xy \le 0 \\
& x, y, z \ge 0
\end{aligned}
$$

Common 'trick' in ML

$$\max\ xyz \iff \max.\ \log xyz$$

$$\iff \max.\ \underbrace{\log x + \log y + \log z}_{\text{concave}}$$

$$\mathcal{X} = \{(x,y,z) : z^2 - xy \le 0, \quad x, y, z \ge 0\}$$

$$g(x,y,z) = z^2 - xy$$

Q: Is $g(x,y,z) \le 0$ a convex constraint?

$$\nabla g = \begin{bmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \\ \frac{\partial g}{\partial z} \end{bmatrix} = \begin{bmatrix} -y \\ -x \\ 2z \end{bmatrix}$$

$$\nabla^2 g = \begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} \quad x \quad & \quad y \quad & \quad z \quad \end{array} \\ \begin{bmatrix} \frac{\partial^2 g}{\partial x^2} & \frac{\partial^2 g}{\partial x \partial y} & \frac{\partial^2 g}{\partial x \partial z} \\ \frac{\partial^2 g}{\partial y \partial x} & \frac{\partial^2 g}{\partial y^2} & \frac{\partial^2 g}{\partial y \partial z} \\ \frac{\partial^2 g}{\partial z \partial x} & \frac{\partial^2 g}{\partial z \partial y} & \frac{\partial^2 g}{\partial z^2} \end{bmatrix} \end{array} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

eigenvalues of $\nabla^2 g$ are $-1, 1, 2$

$\nabla^2 g$ is not positive semi-definite

$g(x,y,z) = z^2 - xy$ is convex.

$g(x,y,z) \le 0$ is not a convex constraint.

$X = \{(x, y, z),\ z^2 - xy \leq 0,\ x, y, z \geq 0\}$

Prove X is a convex set.

$\forall\ (x_1, y_1, z_1)$ and $(x_2, y_2, z_2) \in X$

$\forall\ \lambda \in [0, 1]$

$(\lambda x_1 + (1-\lambda) x_2,\ \lambda y_1 + (1-\lambda) y_2,\ \lambda z_1 + (1-\lambda) z_2)$

$\geq 0 \qquad \geq 0 \qquad \geq 0$

$(\lambda z_1 + (1-\lambda) z_2)^2$

$= \lambda^2 z_1^2 + (1-\lambda)^2 z_2^2 + 2\lambda(1-\lambda) z_1 z_2$

$\leq \lambda^2 x_1 y_1 + (1-\lambda)^2 x_2 y_2 + \lambda(1-\lambda)(x_1 y_2 + x_2 y_1)$

$= \lambda^2 x_1 y_1 + (1-\lambda)^2 x_2 y_2 + \lambda(1-\lambda) x_1 y_2 + \lambda(1-\lambda) x_2 y_1$

$= (\lambda x_1 + (1-\lambda) x_2) \cdot (\lambda y_1 + (1-\lambda) y_2)$

$z_1^2 \leq x_1 y_1$

$z_2^2 \leq x_2 y_2$

$a + b \geq 2\sqrt{ab}$

$2 z_1 z_2$

$\leq 2\sqrt{x_1 y_1}\sqrt{x_2 y_2}$

$= 2\sqrt{(x_1 y_2)(x_2 y_1)}$

$\leq x_1 y_2 + x_2 y_1$

$\in X$

So X is a convex set.

Convex Optimization: Local is Global

$$(\text{P}): \quad \min.\{f(x): \ x \in X\}$$

Theorem

If (P) is a convex optimization problem then a solution $x^* \in X$ is a local optimal solution of (P) iff it is a global optimal solution.

$$\min f(x)$$
$$\text{s.t. } x \in \mathcal{X}$$

Proof:

Suppose x^* is a local optimal but not global.

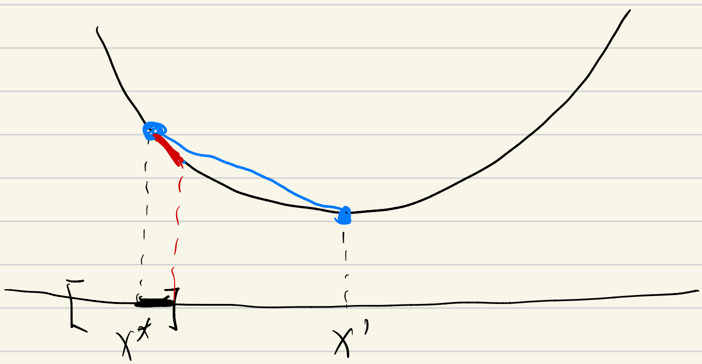
$\exists\, x' \in \mathcal{X}$, $f(x') < f(x^*)$

Since $f(x)$ is convex, $\forall \lambda \in [0,1]$

$$f(\lambda x' + (1-\lambda) x^*) \le \lambda f(x') + (1-\lambda) f(x^*)$$
$$< \lambda f(x^*) + (1-\lambda) f(x^*)$$
$$= f(x^*)$$

$\lambda \to 0$, $\lambda x' + (1-\lambda) x^*$ is in the neighbor-hood of x^*

This is a contradiction.



The 'Easy' and 'Difficult' Optimization Problems

- Linear v.s. nonlinear?
- Differentiable v.s. nondifferentiable?

Classify whether a problem is hard or easy: **Convex (easy)** v.s. **nonconvex (hard)**.

- **Convex optimization:** Benign global geometry and hence reasonable algorithms can almost always find the global minimum.
- **Nonconvex optimization:** Can be anything (e.g., the *saddle point*). In general, finding a stationary point is NP-hard.

Outline

Relaxation

$$\overset{V_P}{\overset{\shortparallel}{(P) : \min_x \{f(x) : x \in X\}}} \qquad \overset{V_Q}{\overset{\shortparallel}{(Q) : \min_x \{g(x) : x \in Y\}}}$$

Problem (Q) is a relaxation of (P) if:

- $X \subseteq Y$
- $\underline{f(x) \geq g(x) \quad \forall x \in X}$

$$\underline{V_P \geq V_Q}$$

Relaxation Properties

- The relaxation of an optimization problem should be easier to solve.
- The optimal value of the relaxation provides a lower bound on the original problem.
- If the relaxation is infeasible, then clearly the original problem is also infeasible.
- Suppose only the constraints are relaxed. If a solution to the relaxation is feasible for the original problem, then it must be an optimal solution to the original problem.

$$(P)\quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X \end{array} \qquad (Q)\quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in Y \end{array} \rightarrow x^* \in X \Rightarrow x^* \text{ is optimal for } (P)$$

x^* is optimal for (Q)

$$X \subseteq Y$$

$$
(P) \quad V_P = \begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) \le b_i, & \forall i \in I \quad \lambda \\ & h_j(x) = d_j, & \forall j \in J \quad \mu \end{array} \quad \longrightarrow \bar{X}
$$

The Lagrangian function is

$$
L(x, \lambda, \mu) = f(x) + \sum_{i \in I} \lambda_i \left[g_i(x) - b_i \right] + \sum_{j \in J} \mu_j \left[h_j(x) - d_j \right]
$$

where $\lambda_i \ge 0$ for all $i \in I$ and μ_j for $j \in J$ are Lagrangian multipliers. Then the following problem is the Lagrangian relaxation of (P):

$$
(Q) : \min_x L(x, \lambda, \mu)
$$

$$L(x;\lambda,\mu) = f(x) + \sum_i \underbrace{\underbrace{\lambda_i}_{\ge 0}\underbrace{[g_i(x)-b_i]}_{\le 0}}_{\le 0} + \sum_j \underbrace{\underbrace{\mu_j}_{\text{free}}\underbrace{[h_j(x)-d_j]}_{=0}}_{=0}$$

$$\le f(x), \quad \forall x \in X$$

Lagrangian Relaxation: (variable: x)

$$\mathcal{L}(\lambda,\mu) = \min_x \; L(x;\lambda,\mu)$$

$$V_p \ge \min_{x\in X} L(x;\lambda,\mu) \ge \min_x L(x;\lambda,\mu) := \mathcal{L}(\lambda,\mu)$$

V_p

better → ① $\lambda = 1M, \mu = 0$

② $\lambda = 100, \mu = -100$

③ $\lambda = 5, \mu = -6$

④ $\lambda = 1, \mu = -1$

$\mathcal{L}(\lambda,\mu)$

Lagrangian Relaxation

$$v_P = \min\{f(x) : g_i(x) \leq b_i, \forall i \in I,\ h_j(x) = d_j, \forall j \in J\}$$

For $\lambda_i \geq 0$, let:

$$\begin{aligned}\mathcal{L}(\lambda, \mu) &= \min_x L(x, \lambda, \mu) \\ &= \min_x \left\{ f(x) + \sum_{i \in I} \lambda_i [g_i(x) - b_i] + \sum_{j \in J} \mu_j [h_j(x) - d_j] \right\}.\end{aligned}$$

Then:

$$\mathcal{L}(\lambda, \mu) \leq v_P \quad \forall \lambda \geq 0.$$

Weak duality

Original Problem:

$$(P): \quad v_P = \min_x \{ f(x) : g_i(x) \le b_i, \, \forall i \in I, \, h_j(x) = d_j, \, \forall j \in J \}$$

(Lagrangian) Dual Problem:

$$(D): \quad v_D = \max_{\lambda,\mu} \{ \mathcal{L}(\lambda, \mu) : \lambda \ge 0 \}$$

where

$$\mathcal{L}(\lambda, \mu) = \min_x \left\{ f(x) + \sum_{i \in I} \lambda_i [g_i(x) - b_i] + \sum_{j \in J} \mu_j [h_j(x) - d_j] \right\}.$$

Weak duality: $v_D \le v_P$

$$
(P) \quad \begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le b_i, \quad \forall i \\ & h_j(x) = d_j, \quad \forall j \end{aligned}
$$

Lagrangian Relaxation. (λ, μ) are Parameters

$$
d(\lambda,\mu) = \min_{x} \ L(x; \lambda, \mu)
$$

Dual:

$$
(D) \quad \max_{\lambda,\mu} \ d(\lambda, \mu)
$$

$$
= \overbrace{\max_{\lambda,\mu} \ \underbrace{\min_{x} \ L(x; \lambda, \mu)}_{\text{Inner}}}^{\text{OUTER}}
$$

Example: two-way partitioning problem

$$\begin{aligned} \min \quad & x^{\mathsf{T}} W x \\ \text{s.t.} \quad & x_i^2 = 1, \forall i = 1, \ldots, n \qquad (\mu) \end{aligned}$$

- Feasible region $\{-1, 1\}^n$ contains 2^n discrete points
- Interpretation: partition in two sets, $x_i \in \{-1, 1\}$ is an assignment for item i.

Lagrangian function

$$L(x;\mu) = x^T W x + \sum_{i=1}^{n} \mu_i (x_i^2 - 1)$$

Lagrangian Relaxation:

$$\min_x L(x;\mu) = \min_x x^T W x + \sum_{i=1}^{n} \mu_i (x_i^2 - 1)$$

Dual:

$$\max_\mu \min_x x^T W x + \sum_{i=1}^{n} \mu_i (x_i^2 - 1)$$

$\cup$ $\cap$

$$\min\ ax^2 - b = \begin{cases} -b & \text{if } a \geq 0 \\ -\infty & \text{if } a < 0 \end{cases}$$

$$\min_x x^T W x + \sum_{i=1}^{n} \mu_i (x_i^2 - 1)$$

$$= \min_x x^T W x + x^T \operatorname{diag}(\mu) x - \sum_{i=1}^{n} \mu_i = \mathbf{1}^T \mu$$

$$= \min_x x^T (W + \operatorname{diag}(\mu)) x - \sum_{i=1}^{n} \mu_i$$

$$= \begin{cases} -\sum_{i=1}^{n} \mu_i & \text{if } W + \operatorname{diag}(\mu) \succeq 0 \\ -\infty & \text{o.w.} \end{cases}$$

$$\mu_1 x_1^2 + \cdots + \mu_n x_n^2$$

$$= [x_1 \cdots x_n] \underbrace{\begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{bmatrix}}_{\operatorname{diag}(\mu)} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Dual:

$$\max.\ -\sum_{i=1}^{n} \mu_i$$
$$\text{s.t. } W + \operatorname{diag}(\mu) \succeq 0$$

Outline

Lagrangian Dual Problem

Given original problem (P):

$$v_P = \min_x \{f(x) : g_i(x) \le b_i, \forall i \in I, \ h_j(x) = d_j, \forall j \in J\}$$

The dual function is defined as:

$$\begin{aligned}\mathcal{L}(\lambda, \mu) &= \min_x L(x, \lambda, \mu) \\ &= \min_x \left\{ f(x) + \sum_{i \in I} \lambda_i [g_i(x) - b_i] + \sum_{j \in J} \mu_j [h_j(x) - d_j] \right\}.\end{aligned}$$

Dual function provides a lower bound: $v_P \ge \mathcal{L}(\lambda, \mu)$ for all $\lambda \ge 0$ and μ.
The best lower bound is given by solving the dual problem:

$$(D): \quad v_D = \max_{\lambda, \mu} \{\mathcal{L}(\lambda, \mu) : \lambda \ge 0\}$$

Duality for General Optimization Problems

Now we have the dual problem

Nested

$$\max_{\lambda \geq 0, \mu} \quad \{\min_x L(x, \lambda, \mu)\}$$

no analytical result

Inner

Outer

- λ and μ are the dual variables.
- If the inner problem has an explicit form, then the dual problem is an explicit maximization problem in λ and μ (this is the case for linear optimization problems).
- However, in general, the inner problem may not have an explicit solution. In those cases, the dual problem doesn't have an explicit form (but it is still an optimization problem for λ and μ).

Dual is Convex Optimization

Remember the dual problem is: linear in λ and μ

$$\max_{\lambda \geq 0, \mu} \{\min_x L(x, \lambda, \mu)\}$$

where L is a linear function of λ and μ.

- Therefore, $\min_x L(x, \cancel{\mu}\,\lambda, \cancel{\nu}\,\mu)$ is always concave in $\cancel{\mu}\,\lambda$ and $\cancel{\nu}\,\mu$.
- Thus the dual problem is always a convex optimization problem — regardless of whether the primal problem is a convex optimization problem or not.
- The dual of dual problem does not necessarily equal to the primal problem.

Example

Consider the following problem:

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{x}} & \boldsymbol{x}^T\boldsymbol{x} \\ \text{subject to} & A\boldsymbol{x} \leq \boldsymbol{b} \quad (\lambda) \end{array}$$

It is a nonlinear optimization problem. Now let's construct its dual problem.

$$L(x;\lambda) = x^T x + \lambda^T (Ax - b)$$

$$\min_x \; x^T x + \lambda^T (Ax - b)$$

$$\nabla_x L(x, \lambda) = 2x + \underset{n \times m}{A^T} \underset{m \times 1}{\lambda} = 0$$

$$x = -\frac{1}{2} A^T \lambda$$

$$\min_x \; x^T x + \lambda^T (Ax - b)$$

$$= \left(-\frac{1}{2} A^T \lambda\right)^T \left(-\frac{1}{2} A^T \lambda\right) + \lambda^T A \left(-\frac{1}{2} A^T \lambda\right) - \lambda^T b$$

$$= \frac{1}{4} \lambda^T A A^T \lambda - \frac{1}{2} \lambda^T A A^T \lambda - \lambda^T b$$

$$= -\frac{1}{4} \lambda^T A A^T \lambda - \lambda^T b$$

Dual: $\max. \; -\frac{1}{4} \lambda^T A A^T \lambda - \lambda^T b$

s.t. $\lambda \geq 0$

Weak Duality

$$V_P \geq \min_{x \in \mathcal{X}} L(x;\lambda,\mu) \geq \boxed{\min_{x} L(x;\lambda,\mu)} := d(\lambda,\mu)$$

$$V_P \geq \max_{\lambda \geq 0,\mu} d(\lambda,\mu) := V_D$$

Theorem (Weak Duality)

If the primal and dual optimal value is v_P and v_D, then

$$v_D \leq v_P$$

Note that this always holds (even if the primal problem is nonconvex).

Weak Duality Always Holds!

Strong Duality

Duality GAP $= V_P - V_D \geq 0$

For LP, we have the strong duality theorem, saying that the optimal values of primal and dual problems are always the same.

Unfortunately, this is not necessarily true for nonlinear optimization problems. We call the difference between the primal optimal value and dual optimal value the duality gap.

- For LP, the duality gap is always 0.
- For general optimization problems, there could be a positive duality gap.
- Again, by the weak duality theorem, it must be that the optimal value of the minimization problem is larger than that of the maximization problem.

Strong Duality Theorem

Theorem (Strong Duality Theorem)

1 2

If the primal problem is a convex optimization problem and also Slater's condition holds, then the duality gap is 0, *i.e.,* $v_D = v_P$.

$$\begin{cases} g_i(x) \le b_i \\ h_j(x) = d_j \end{cases}$$

Definition (Slater's Condition)

If the original problem (P) is convex (i.e., f and g_i are convex for all $i \in I$, and h_j are affine for all $j \in J$), and there exists a strictly feasible $x \in \mathbb{R}^n$ such that:

$$g_i(x) < b_i \ \ \forall i \in I, \quad h_j(x) = d_j \ \ \forall j \in J,$$

then Slater's Condition holds and strong duality holds.

- Actually only need strict inequalities for non-affine h_i
- In other words, Slater's condition requires that there is a strict interior solution for nonlinear constraints.

Summary

- We use Lagrangian relaxation to derive dual problems for nonlinear optimization problems
- Dual problems for nonlinear optimization do not necessarily have an explicit form
- Weak duality always holds
- Strong duality holds when the problem is convex and the Slater's condition holds (otherwise, there could be a positive duality gap)

Outline

General Optimality Conditions

Now, We consider the following nonlinear program:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \quad \forall i = 1, \ldots, m, \\ & h_j(x) = 0, \quad \forall j = 1, \ldots, p. \end{aligned}$$

We will also work with the following index sets:

Definition: Active and Inactive Set

At a point $x \in X$, the set $\mathcal{A}(x) := \{i : g_i(x) = 0\}$ denotes the set of **active constraints**. The set of **inactive constraints** is then given by $\mathcal{I}(x) := \{i : g_i(x) < 0\}$.

Karush-Kuhn-Tucker Conditions

Karush-Kuhn-Tucker (KKT) Conditions:

- Main/Stationarity Condition

$$\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x) = 0$$

- Complementary Slackness

$$\lambda_i \cdot g_i(x) = 0 \quad \forall i = 1, \ldots, m$$

- Primal Feasibility

$$g_i(x) \leq 0, \quad h_j(x) = 0 \quad \forall i = 1, \ldots, m, \ \forall j = 1, \ldots, p$$

- Dual Feasibility

$$\lambda_i \geq 0 \quad \forall i = 1, \ldots, m$$

Theory with KKT Conditions

Theorem

1. *For any optimization problem with a constraint qualification (CQ) holds (e.g., strong duality), if* $x^\star$ *and* $\lambda^\star, \mu^\star$ *are primal and dual (local) optimal solutions, then* $x^\star$ *and* $\lambda^\star, \mu^\star$ *satisfy the KKT conditions.*
2. *For convex optimization problem, if* $x^\star$ *and* $\lambda^\star, \mu^\star$ *satisfy the KKT conditions, then* $x^\star$ *and* $\lambda^\star, \mu^\star$ *are primal and dual optimal solutions and strong duality holds.*

Constraint Qualifications

We require the collection of gradients

$$\{\nabla g_i(x) : i \in \mathcal{A}(x)\} \cup \{\nabla h_j(x) : j = 1, \ldots, p\} \qquad \text{(CQ)}$$

to be linearly independent or to have full rank.

- This condition is a **constraint qualification (CQ)** and is called **Linear Independence Constraint Qualification (LICQ)**.
- A feasible point x satisfying the LICQ is called **regular**.
- There are more CQs: ACQ, GCQ, MFCQ, PLICQ, Slater's condition (Optimization II (?)).

Failure of The KKT Conditions in the Absence of CQ

$$\begin{aligned} \min_{x_1,x_2} \quad & x_1 \\ \text{subject to} \quad & (x_1-1)^2+(x_2-1)^2 \leq 1, \\ & (x_1-1)^2+(x_2+1)^2 \leq 1. \end{aligned}$$

- Since there is only one feasible solution (i.e., $(x_1, x_2) = (1, 0)$), it is automatically optimal.
- However, this point does not satisfy the KKT conditions (exercise).
- This example says if a local min does not hold certain CQ, the KKT conditions may not be necessary at this point.
- This means that points satisfying KKT conditions may not contain all the local minimizers, since some of them may not be regular.

Remarks on KKT Conditions

Remarks:

- KKT conditions is first-order necessary conditions (FONC) for general constrained optimization problems.
- KKT conditions unify all formerly studied FONC.
- A (feasible) point (often the primal and dual variables (x, λ, μ) together) satisfying the KKT conditions is called a KKT point, regardless whether it satisfies CQ or not. We may also call x a KKT point for simplicity.
- KKT points are candidates for local optimal solutions – just like stationary points.

Example 1

Consider the problem:

$$\begin{array}{rc} \text{minimize} & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to} & x_1^2 + x_2^2 \leq 5 \\ & 3x_1 + x_2 \geq 3 \end{array}$$

Find the KKT conditions.

Example 2

Find the KKT conditions for

$$
\begin{aligned}
\text{minimize}_{\boldsymbol{x}} \quad & \boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{c}^T \boldsymbol{x} \\
\text{s.t.} \quad & A\boldsymbol{x} = \boldsymbol{b} \\
& C\boldsymbol{x} \geq \boldsymbol{d} \\
& \boldsymbol{x} \geq 0
\end{aligned}
$$

Example 3: Cylinder Volume

We want to build a cylinder with the maximum volume, with its surface area no larger than C.

- Decision variables: r (the radius of the base) and h (height).
- Then the optimization problem is:

$$\begin{aligned} \text{maximize}_{r,h} \quad & \pi r^2 h \\ \text{subject to} \quad & 2\pi r^2 + 2\pi rh \leq C \\ & r, h \geq 0 \end{aligned}$$

Example 4: Power Allocation

We have a collection of n communication channels and we need to decide how much power to allocate to each of them

- The capacity (communication rate) of channel i is $\log(\alpha_i + x_i)$ with a given $\alpha_i > 0$ and when x_i is allocated to it, and we have a budget constraint $\boldsymbol{e}^T\boldsymbol{x} = 1$, $\boldsymbol{x} \geq 0$

The optimization problem is:

$$\begin{array}{rl} \text{maximize}_{\boldsymbol{x}} & \sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{subject to} & \sum_{i=1}^n x_i = 1 \\ & \boldsymbol{x} \geq 0 \end{array}$$