Optimization Review

Pattern Recognition: Fall 2019

University of Tehran
School of Electrical and Computer Engineering

Optimization Problems Introduction

Optimization Problem

$$\begin{array}{rll} \text{Minimize} & \mathrm{f}_0(x) & \\ \text{Subject to} & \mathrm{f}_i(x) \leq 0 & i = 1, ..., n \\ & \mathrm{h}_k(x) = 0 & k = 1, ..., p \end{array}$$

- $\mathrm{f}_0(x)$: Objective (or goal) function
- $\mathrm{f}_i(x)$: inequality constraint
- $\mathrm{h}_k(x)$: equality constraint

Optimization Problem

$$\begin{aligned} \text{Minimize} \quad & f_0(x) \\ \text{Subject to} \quad & f_i(x) \le 0 \qquad i = 1, ..., n \\ & h_k(x) = 0 \qquad k = 1, ..., p \end{aligned}$$

- Domain of opt. problem: $\mathcal{D} = \cap \mathcal{D}_{f_0} \ \cap_{i=1}^{n} \mathcal{D}_{f_i} \ \cap_{k=1}^{p} \mathcal{D}_{h_k}$
- Feasible set: $\{x \mid x \in \mathcal{D}, \ f_i(x) \le 0, \ h_k(x) = 0\}$
- Feasible point: a point of a feasible set
- If feasible set is empty, problem is called infeasible.

Types of optimization Problem

Optimization Problems can be categorized as follows:

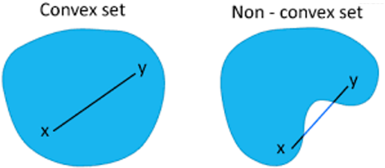
- Convex $\rightarrow$ $\mathrm{f}_0(x)$ and $\mathrm{f}_i(x)$ should be convex and $\mathrm{h}_k(x)$ should be affine functions.
- Non-convex $\rightarrow$ e.g. problems with binary or integer variables

Convex and Affine Functions

Convex set

$\mathcal{C}$ is a convex set if:

$$\forall x_1 \text{ and } x_2 \in \mathcal{C} \implies \theta x_1 + (1-\theta)x_2 \in \mathcal{C} \qquad \forall\, 0 \le \theta \le 1$$



Convex and Affine Functions

Convex set

$\mathcal{C}$ is a convex set if:

$$\forall x_1 \text{ and } x_2 \in \mathcal{C} \implies \theta x_1 + (1-\theta)x_2 \in \mathcal{C} \qquad \forall\, 0 \le \theta \le 1$$

Convex function

Function f: dom(f)$\in \mathbb{R}^n \to \mathbb{R}$ is a convex function if:

- dom(f) is a convex set
- $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y) \qquad \forall x, y \in \text{dom(f)} \qquad \forall\, 0 \le \theta \le 1$

Or if $\nabla^2 f(x) \succeq 0$
If $-f(x)$ is convex, $f(x)$ is concave.

Convex and Affine Functions

Convex set

$\mathcal{C}$ is a convex set if:

$$\forall x_1 \text{ and } x_2 \in \mathcal{C} \implies \theta x_1 + (1-\theta)x_2 \in \mathcal{C} \qquad \forall\, 0 \leq \theta \leq 1$$

Convex function

Function f: dom(f)$\in \mathbb{R}^n \to \mathbb{R}$ is a convex function if:

- dom(f) is a convex set
- $\mathrm{f}(\theta x + (1-\theta)y) \leq \theta \mathrm{f}(x) + (1-\theta)\mathrm{f}(y) \qquad \forall x, y \in \mathrm{dom(f)} \qquad \forall\, 0 \leq \theta \leq 1$

Or if $\bigtriangledown^2 \mathrm{f}(x) \succeq 0$
If $-\mathrm{f}(x)$ is convex, $\mathrm{f}(x)$ is concave.

Example

$$f(x) = \sum_{i=1}^{n} x_i \ln(x_i)$$

Convex and Affine Functions

Convex set

$\mathcal{C}$ is a convex set if:

$$\forall x_1 \text{ and } x_2 \in \mathcal{C} \implies \theta x_1 + (1-\theta)x_2 \in \mathcal{C} \qquad \forall\, 0 \leq \theta \leq 1$$

Convex function

Function f: dom(f)$\in \mathbb{R}^n \to \mathbb{R}$ is a convex function if:

- dom(f) is a convex set
- $\mathrm{f}(\theta x + (1-\theta)y) \leq \theta \mathrm{f}(x) + (1-\theta)\mathrm{f}(y) \qquad \forall x, y \in \mathrm{dom(f)} \qquad \forall\, 0 \leq \theta \leq 1$

Or if $\bigtriangledown^2 \mathrm{f}(x) \succeq 0$
If $-\mathrm{f}(x)$ is convex, $\mathrm{f}(x)$ is concave.

Affine set

$\mathcal{C}$ is a affine set if:

$$\forall x_1 \text{ and } x_2 \in \mathcal{C} \implies \theta x_1 + (1-\theta)x_2 \in \mathcal{C} \qquad \forall\, \theta$$

Why convexity of the problem is important?

- Easy known solution
- Global and local optimal values are identical.

First Order Optimality

For a convex problem, x^* is optimal iff: $\bigtriangledown \mathrm{f}^T(x^*)(y - x^*) \geq 0 \quad \forall y \in$ feasible set

Example

$$\begin{aligned} \text{Minimize} \quad & x_1^2 + (x_2 - 1)^2 \\ \text{Subject to} \quad & -1 \leq x_1 \leq 1 \\ & x_2 \geq 0 \end{aligned}$$

Duality and necessary optimality conditions

Lagrangian Function

$$
\begin{array}{rll}
\text{Minimize} & f_0(x) & \\
\text{Subject to} & f_i(x) \leq 0 & i = 1, \dots, n \\
 & h_k(x) = 0 & k = 1, \dots, p
\end{array}
$$

Lagrangian Function

$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{n} \lambda_i \, f_i(x) + \sum_{k=1}^{p} \nu_k \, h_k(x)$

ν and λ: Dual variables or Lagrangian multipliers

Lagrangian Function

$$\begin{array}{rlr} \text{Minimize} & f_0(x) & \\ \text{Subject to} & f_i(x) \leq 0 & i = 1, ..., n \\ & h_k(x) = 0 & k = 1, ..., p \end{array}$$

Lagrangian Function

$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{n} \lambda_i f_i(x) + \sum_{k=1}^{p} \nu_k h_k(x)$
ν and λ: Dual variables or Lagrangian multipliers

Lagrangian Dual Function

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

Dual function provides lower bound for p^* (optimal value of primal) for any $\lambda_i \geq 0$

proof

$$
\begin{aligned}
\mathrm{g}(\lambda, \nu) &= \inf_{x \in \mathcal{D}} \mathrm{L}(x, \lambda, \nu) \\
&\leq \mathrm{L}(x^*, \lambda, \nu) \\
&= \mathrm{f}_0(x^*) + \sum_{i=1}^{n} \lambda_i \, \mathrm{f}_i(x^*) + \sum_{k=1}^{p} \nu_k \, \mathrm{h}_k(x^*) \\
&\leq p^* \\
\implies \mathrm{g}(\lambda, \nu) &\leq p^*
\end{aligned}
$$

Dual Problem

$$\max \mathrm{g}(\lambda, \nu) = d^*$$
$$\lambda \geq 0$$

If primal problem is convex and there is at least one strong feasible point, then the duality gap $(d^* - p^*)$ is zero.

Strong feasible set: $\{x \mid x \in \mathcal{D}, \mathrm{f}_i < 0, \mathrm{h}_k = 0\}$

Dual Problem

$$\max \mathrm{g}(\lambda, \nu) = d^*$$
$$\lambda \geq 0$$

If primal problem is convex and there is at least one strong feasible point, then the duality gap $(d^* - p^*)$ is zero.

Strong feasible set: $\{x \mid x \in \mathcal{D}, \mathrm{f}_i < 0, \mathrm{h}_k = 0\}$

Example

$$\min x^2$$
$$s.t.\ ax + b \geq 0$$

Dual Problem

$$\max \mathrm{g}(\lambda, \nu) = d^*$$
$$\lambda \geq 0$$

If primal problem is convex and there is at least one strong feasible point, then the duality gap $(d^* - p^*)$ is zero.

Strong feasible set: $\{x \mid x \in \mathcal{D}, \mathrm{f}_i < 0, \mathrm{h}_k = 0\}$

Example

$$\min c^T x$$
$$s.t.\ Ax + b = 0, \quad x \geq 0$$

KKT Conditions

- Primal feasibility: $\mathrm{f}_i(x^*) \leq 0, \quad \mathrm{h}_k(x^*) = 0$
- Dual feasibility: $\lambda_i^* \geq 0$
- Complementary slackness: $\lambda_i^* \, \mathrm{f}_i(x^*) = 0$
- Lagrangian stability: $\bigtriangledown_x \mathrm{L}(x^*, \lambda^*, \nu^*) = 0$

KKT Conditions

- Primal feasibility: $\mathrm{f}_i(x^*) \leq 0, \quad \mathrm{h}_k(x^*) = 0$
- Dual feasibility: $\lambda_i^* \geq 0$
- Complementary slackness: $\lambda_i^* \, \mathrm{f}_i(x^*) = 0$
- Lagrangian stability: $\bigtriangledown_x \mathrm{L}(x^*, \lambda^*, \nu^*) = 0$

Example

$$
\begin{aligned}
\min\ & x^2 + 2y^2 \\
s.t.\ & x + y \geq 3 \\
& y - x^2 \geq 1
\end{aligned}
$$

KKT Conditions

- Primal feasibility: $\mathrm{f}_i(x^*) \leq 0, \quad \mathrm{h}_k(x^*) = 0$
- Dual feasibility: $\lambda_i^* \geq 0$
- Complementary slackness: $\lambda_i^* \, \mathrm{f}_i(x^*) = 0$
- Lagrangian stability: $\bigtriangledown_x \mathrm{L}(x^*, \lambda^*, \nu^*) = 0$

Example

$$\min \frac{1}{2} x^T Q x + q^T x + r$$
$$s.t. \; Ax + b = 0$$

Nonlinear Programming

Nonlinear Programming

Which problems can be solved by nonlinear optimization algorithms?

- Problems with nonlinear derivatives $\rightarrow$ Analytical solution is hard to find.
- non-convex problems.
- problems with more than one local optimums.
- initial point dependent problems.

Simplex

$$
\begin{aligned}
&\max z \\
&s.t.\ Ax = b \\
&\quad\ x \geq 0 \\
&\quad\ x \in \mathbb{R}^n, \quad b_{m\times n}, \quad A_{m\times n}
\end{aligned}
$$

$$
\begin{aligned}
z - c_1x_1 - ... - c_nx_n &= 0 \\
a_{11}x_1 + ... + a_{1n}x_n &= b_1 \\
&\vdots \\
a_{m1}x_1 + ... + a_{mn}x_n &= b_m
\end{aligned}
$$

Simplex

$$\begin{aligned} &\max z \\ &s.t.\ Ax = b \\ &\quad x \geq 0 \\ &\quad x \in \mathbb{R}^n, \quad b_{m\times n}, \quad A_{m\times n} \end{aligned}$$

$$\begin{aligned} z - c_1x_1 - \dots - c_nx_n &= 0 \\ a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n &= b_m \end{aligned}$$

Algorithm 1 Simplex Method

```
unbounded=False, optimal=False
while unbounded=False & optimal=False do
    if c_i ≤ 0 ∀i then optimal=True
    else
        choose variable j = arg max{c_i | c_i > 0}
        choose row k = arg min{b_i / a_ij | a_ij > 0}
        if a_ij ≤ 0 ∀i then unbounded=True
        obtain x_j from kth row and rewrite equations.
        omit row k.
```

Simplex

Algorithm 2 Simplex Method

unbounded=False, optimal=False
while unbounded=False & optimal=False **do**
  if $c_i \leq 0 \; \forall i$ **then** optimal=True
  else
    choose variable $j = \arg\max\{c_i \mid c_i > 0\}$
    choose row $k = \arg\min\{\frac{b_i}{a_{ij}} \mid a_{ij} > 0\}$
    if $a_{ij} \leq 0 \; \forall i$ **then** unbounded=True
    obtain x_j from kth row and rewrite equations.
    omit row k.

Example

$$\begin{aligned} &\max\, 5x_1 + 4x_2 + 3x_3 \\ &s.t.\; 2x_1 + 3x_2 + x_3 \leq 5 \\ &\quad\;\; 4x_1 + x_2 + 2x_3 \leq 11 \\ &\quad\;\; 3x_1 + 4x_2 + 2x_3 \leq 8 \qquad x_1, x_2, x_3 \geq 0 \end{aligned}$$

Example 1 (slide 5)

$$f(x) = \sum_{i=1}^{n} x_i \ln(x_i) \longrightarrow \frac{\partial f}{\partial x_i} = \ln(x_i + 1)$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = 0 \quad i \neq j \qquad \frac{\partial^2 f}{\partial x_i^2} = \frac{1}{x_i}$$

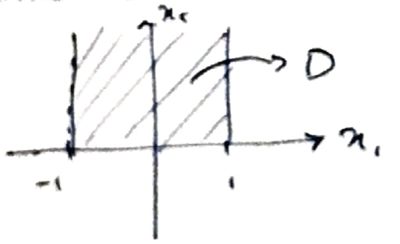
$$\Rightarrow \nabla^2 f(x) = \begin{bmatrix} \frac{1}{x_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{x_n} \end{bmatrix} \longrightarrow \left. \begin{array}{l} \det(\nabla^2 f(x)) = \prod_{i=1}^{n} \frac{1}{x_i} \\ D_f = \{x \mid x \in \mathbb{R}^+\} \end{array} \right\} \rightarrow \det(\nabla^2 f) > 0$$

$\Rightarrow f(x)$ is a convex function

Example 2 (slide 7)

$$f(x) = x_1^2 + (x_2 - 1)^2 \longrightarrow D_f = \mathbb{R}^2$$

the problem domain $= D = \mathbb{R}^2 \cap -1 \leq x_1 \leq 1 \cap x_2 \geq 0$



D is a convex set ①

$$\frac{\partial f}{\partial x_1} = 2x_1 \quad , \quad \frac{\partial f}{\partial x_2} = 2(x_2 - 1) \quad , \quad \frac{\partial f}{\partial x_1 \partial x_2} = 0 \quad , \quad \frac{\partial^2 f}{\partial x_1^2} = 2$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2$$

$$\rightarrow \nabla^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \longrightarrow f \text{ is a convex function } ②$$

① & ② $\longrightarrow$ the problem is convex

$$\begin{bmatrix} 2x_1 & 2(x_2 - 1) \end{bmatrix} \begin{bmatrix} y_1 - x_1 \\ y_2 - x_2 \end{bmatrix} \geq 0 \longrightarrow \begin{cases} x_1 (y_1 - x_1) \geq 0 \xrightarrow{\forall y_1} x_1 = 0 \\ (x_2 - 1)(y_2 - x_2) \geq 0 \xrightarrow{\forall y_2} x_2 = 1 \end{cases}$$

$$\Rightarrow x^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Example 3 (slide 11-1)

$x^2 \rightarrow$ convex function

$ax+b \rightarrow$ affine function $\Rightarrow$ convex function

} the problem is convex $\Rightarrow$ duality gap $=0$

$$L(x,\lambda) = x^2 - \lambda(ax+b)$$

$$\frac{\partial L}{\partial x} = 2x - a\lambda = 0 \longrightarrow x = \frac{a\lambda}{2}$$

$$\Rightarrow g(\lambda) = \inf_x L(x,\lambda) = \frac{a^2\lambda^2}{4} - \lambda\left(\frac{a^2\lambda}{2} + b\right) = -\frac{a^2\lambda^2}{4} - \lambda b$$

dual problem: $\max \; -\frac{a^2\lambda^2}{4} - \lambda b$

s.t. $\lambda \geq 0$

$$\frac{dg}{d\lambda} = -\frac{a^2}{2}\lambda - b = 0 \longrightarrow \lambda = \frac{-2b}{a^2}$$

$b \leq 0 \nearrow \quad p^* = d^* = -\frac{a^2}{4} \times \frac{4b^2}{a^4} + \frac{2b^2}{a^2} = \frac{b^2}{a^2}$

$b > 0 \searrow \quad \lambda = 0 \rightarrow d^* = p^* = 0$

Example 4 (slide 11-2)

$$L(x,v,\lambda) = c^T x + v(Ax+b) - \lambda^T x = (c^T + vA - \lambda^T)x + vb$$

$$g(\lambda, v) = \begin{cases} vb & c^T + vA - \lambda^T = 0 \\ -\infty & \text{o.w.} \end{cases}$$

$\Rightarrow$ dual problem: $\max \; vb$

s.t. $c^T + vA - \lambda^T = 0$

$\lambda \geq 0$

Example 5 (slide 12)

1) Primal feasibility: $x+y \geq 3$ $\quad y-x^2 \geq 1$

2) Dual feasibility: $\lambda_1 \geq 0$ $\quad \lambda_2 \geq 0$

3) Complementary Slackness: $\lambda_1(3-x-y)=0$ $\quad \lambda_2(1+x^2-y)=0$

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

$\nabla_x L = 2x-\lambda_1+2\lambda_2 = 0 \quad \nabla_y L = 4y-\lambda_1+\lambda_2 = 0$

① $\lambda_1=\lambda_2=0 \xrightarrow{4)} x=y=0 \xrightarrow{1)} x+y=0 \not\geq 3$ ✗

② $\lambda_1=0,\ \lambda_2>0 \xrightarrow{4)} x=\lambda_2,\ y=\frac{\lambda_2}{4} \xrightarrow{3)} 1+x^2-y=0 = 1+\lambda_2^2-\frac{\lambda_2}{4}=0$ ✗

③ $\lambda_1>0,\ \lambda_2=0 \xrightarrow{4)} x=\frac{\lambda_1}{2},\ y=\frac{\lambda_1}{4} \xrightarrow{3)} 3-\frac{\lambda_1}{2}-\frac{\lambda_1}{4}=0 \to \lambda_1=4$

$\to x=2,\ y=1 \xrightarrow{1)} y-x^2=1-4 \not\geq 1 \to$ ✗

4) $\lambda_1>0,\ \lambda_2>0 \xrightarrow{3)} \begin{array}{l} 3-x-y=0 \\ 1+x^2-y=0 \end{array} \to \begin{array}{l} x=-2,\ y=5 \\ x=1,\ y=2 \end{array}$

$x=-2,\ y=5 \xrightarrow{4)} \begin{array}{l} -4-\lambda_1+2\lambda_2=0 \\ 20-\lambda_1+\lambda_2=0 \end{array} \to \lambda_2=24,\ \lambda_1=44$ ✓

$x=1,\ y=2 \xrightarrow{4)} \begin{array}{l} 2-\lambda_1+2\lambda_2=0 \\ 8-\lambda_1+\lambda_2=0 \end{array} \to \lambda_1=14,\ \lambda_2=6$ ✓

for minimizing x^2+2y^2 we should choose: $x^*=1$, $y^*=2$

Example 6 (Slide 12-2)

KKT Conditions : 1) $Ax + b = 0$

2) $L(x, v) = \frac{1}{2} x^T Q x + q^T x + r + v(Ax+b)$

$$\rightarrow \frac{\partial L}{\partial x} = Qx + q + Av = 0$$

$$\rightarrow \begin{bmatrix} Q & A \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} -q \\ -b \end{bmatrix}$$

Example 7 (Slide 15)

$$Z - 5x_1 - 4x_2 - 3x_3 = 0$$

1) $2x_1 + 3x_2 + x_3 + x_4 = 5$

2) $4x_1 + x_2 + 2x_3 + x_5 = 11$

3) $3x_1 + 4x_2 + 2x_3 + x_6 = 8$

x_4, x_5, x_6 : Slack Variables

$x_i \geq 0 \quad i = 1, 2, \dots, 6$

$\max\{5, 4, 3\} = 5 \rightarrow x_1$ is chosen

$\min\left\{\frac{5}{2}, \frac{11}{4}, \frac{8}{3}\right\} = \frac{5}{2} \rightarrow$ row 1 is chosen

$$\rightarrow x_1 = -\frac{3}{2}x_2 - \frac{x_3}{2} - \frac{x_4}{2} + \frac{5}{2}$$

$$\Rightarrow Z + \frac{7}{2}x_2 - \frac{1}{2}x_3 + \frac{5}{2}x_4 - \frac{25}{2} = 0$$

$$x_1 + \frac{3}{2}x_2 + \frac{x_3}{2} + \frac{x_4}{2} = \frac{5}{2}$$

$$-5x_2 - x_3 - 2x_4 + x_5 = 1$$

$$-\frac{1}{2}x_2 + \frac{1}{2}x_3 - \frac{3}{2}x_4 + x_6 = \frac{1}{2}$$

$\max\left\{\frac{1}{2}\right\} = \frac{1}{2} \rightarrow x_3$ is chosen

$\min\left\{\frac{5/2}{1/2}, \frac{1/2}{1/2}\right\} = \frac{1/2}{1/2} \rightarrow$ row 3 is chosen

$x_3 = 1 + x_2 + 3x_4 - 2x_6$

$\Rightarrow \quad Z + x_4 + 3x_2 + x_6 = 13$ → all coefficients are negative

$x_1 + 2x_2 + 2x_4 - x_6 = 2$ ⇒ optimal point:

$-5x_2 - 2x_4 + x_5 = 1$ $\quad x_4^* = x_2^* = x_6^* = 0$

$x_3 - x_2 - 3x_4 + 2x_6 = 1$ $\quad x_1^* = 2, \; x_5^* = 1, \; x_3^* = 1$

$Z^* = 13^*$