KKT Conditions

Optimization Techniques (ENGG*6140)

School of Engineering,
University of Guelph, ON, Canada

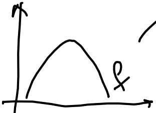
Course Instructor: Benyamin Ghojogh

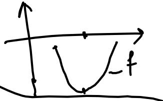
Winter 2023

The Lagrangian Function

Optimization Problem

x-1





$$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} &\quad f(\boldsymbol{x})\\ \text{subject to} &\quad y_i(\boldsymbol{x})\le 0,\quad i\in\{1,\dots,m_1\},\\ &\quad h_i(\boldsymbol{x})=0,\quad i\in\{1,\dots,m_2\}.\end{aligned} \qquad (1)$$

$h_1(x)=0$
$\vdots$
$h_{m_2}(x)=0$

$y_1(x)\le 0 \rightarrow x_1^2+3x_2\le 0$
$\rightarrow \tan(x)\le$
$y_2(x)\le 0$
$\vdots$
$y_{m_1}(x)\le 0$

Inequality constraints:

- $y_i(\boldsymbol{x})\le 0$
- $k_i(\boldsymbol{x})\ge 0 \overset{x-1}{\Longrightarrow} -k_i(\boldsymbol{x})\le 0 \Longrightarrow y_i(\boldsymbol{x})\le 0,$ where $y_i(\boldsymbol{x}) := -k_i(\boldsymbol{x})$

$y_i(x)=-k_i(x)$

$x_1^2+3x_2\le 0$
$x_1^3-2x_1\ge 0$
$-x_1^3+2x_2\le 0$

Maximization:

- $\underset{\boldsymbol{x}}{\text{maximize}}\ f(\boldsymbol{x}) \Longrightarrow \underset{\boldsymbol{x}}{\text{minimize}}\ -f(\boldsymbol{x}) \Longrightarrow \underset{\boldsymbol{x}}{\text{minimize}}\ g(\boldsymbol{x})$

$x_1^3+2x_2\le 1$
$x_1^3+2x_2-1\le 0$

Examples:

for $f_i(\boldsymbol{x})$: $\boldsymbol{x}^\top\boldsymbol{x}+\boldsymbol{a}^\top\boldsymbol{x}+b$, $x_1^3-4x_2-2x_3^2$, $\tan(x_1)-4\sin(x_3)-2$

for $y_i(\boldsymbol{x})\le 0$: $\boldsymbol{x}^\top\boldsymbol{x}+\boldsymbol{a}^\top\boldsymbol{x}+b\le\boldsymbol{0}$

for $y_i(\boldsymbol{x})\le 0$: $x_1^2+3x_2+2\le 0$, $x_2^3-4x_3-2\le 0$, $\tan(x_3)-4\sin(x_1)-2\le 0$, ...

for $h_i(\boldsymbol{x})=0$: $\boldsymbol{x}^\top\boldsymbol{x}+\boldsymbol{a}^\top\boldsymbol{x}+b=\boldsymbol{0}$

for $h_i(\boldsymbol{x})=0$: $x_1^2+3x_2+2=0$, $x_2^3-4x_3-2=0$, $\tan(x_3)-4\sin(x_1)-2=0$, ...

Lagrangian and Dual Variables

$x \in \mathbb{R}^d$

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boxed{y_i(\boldsymbol{x}) \le 0, \; i \in \{1, \dots, m_1\},} \\ & \boxed{h_i(\boldsymbol{x}) = 0, \; i \in \{1, \dots, m_2\}.} \end{aligned}$$

Definition (Lagrangian and dual variables)

The **Lagrangian function** for the optimization problem (1) is $\mathcal{L} : \underbrace{\mathbb{R}^{\overset{x}{d}} \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}} \to \mathbb{R}$ with domain $\mathcal{D} \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, defined as:

$$\star \; \mathcal{L}(\underbrace{\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}}) := f(\boldsymbol{x}) + \underbrace{\sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x})} + \underbrace{\sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x})} = f(\boldsymbol{x}) + \overbrace{\boldsymbol{\lambda}^\top \boldsymbol{y}(\boldsymbol{x})} + \underbrace{\boldsymbol{\nu}^\top \boldsymbol{h}(\boldsymbol{x})}, \tag{2}$$

where $\{\lambda_i\}_{i=1}^{m_1}$ and $\{\nu_i\}_{i=1}^{m_2}$ are the **Lagrange multipliers**, also called the **dual variables**, corresponding to inequality and equality constraints, respectively.

- $\boldsymbol{\lambda} := [\lambda_1, \dots, \lambda_{m_1}]^\top \in \mathbb{R}^{m_1}$, $\boldsymbol{\nu} := [\nu_1, \dots, \nu_{m_2}]^\top \in \mathbb{R}^{m_2}$, $\boldsymbol{y}(\boldsymbol{x}) := [y_1(\boldsymbol{x}), \dots, y_{m_1}(\boldsymbol{x})]^\top \in \mathbb{R}^{m_1}$, $\boldsymbol{h}(\boldsymbol{x}) := [h_1(\boldsymbol{x}), \dots, h_{m_2}(\boldsymbol{x})]^\top \in \mathbb{R}^{m_2}$.
- Eq. (2) is also called the **Lagrange relaxation** of the optimization problem (1).

Sign of Terms in Lagrangian

$$-y_i(x) \le 0 \rightarrow y_i(x) \ge 0$$

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \overbrace{y_i(\boldsymbol{x}) \le 0}^{\times}, \ i \in \{1, \dots, m_1\}, \\ & h_i(\boldsymbol{x}) = 0, \ i \in \{1, \dots, m_2\}. \end{aligned}$$

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) := \underbrace{f(\boldsymbol{x})}_{} \textcircled{+} \underbrace{\sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x})}_{} \textcircled{+} \underbrace{\sum_{i=1}^{m_2} \nu_i \overbrace{h_i(\boldsymbol{x})}^{}}_{} = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{y}(\boldsymbol{x}) + \boldsymbol{\nu}^\top \boldsymbol{h}(\boldsymbol{x}).$$

- sometimes, the plus sign behind $\sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x})$ is replaced with the negative sign. As $h_i(\boldsymbol{x})$ is for equality constraint, its sign is not important in the Lagrangian function.

$$\left\{ \begin{array}{l} \underbrace{h_i(\boldsymbol{x}) = 0}_{} \xrightarrow{\times -1} \underbrace{-\overbrace{h_i(\boldsymbol{x})}^{h_i(x)} = 0}_{}, \\ \rightarrow \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) := f(\boldsymbol{x}) + \sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x}) \textcircled{-} \underbrace{\sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x})}_{} = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{y}(\boldsymbol{x}) \textcircled{-} \boldsymbol{\nu}^\top \boldsymbol{h}(\boldsymbol{x}). \end{array} \right.$$

$$\underline{\cancel{+}}\, \nu_1 h_1(x) - \nu_2 h_2(x) + \dots$$

$$\rightarrow + \sum_{i=1}^{m_2} \underbrace{(-\nu_i)}_{z_i} h_i(x)$$

- <u>However</u>, the <u>sign</u> of the term $\underbrace{\sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x})}_{}$ is important because the sign of inequality constraint is <u>important</u>.

Interpretation of Lagrangian

constrained

cost

hard restriction

★

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} && f(\boldsymbol{x}) \\ &\text{subject to} && y_i(\boldsymbol{x}) \le 0,\ i \in \{1, \dots, m_1\}, \\ &&& h_i(\boldsymbol{x}) = 0,\ i \in \{1, \dots, m_2\}. \end{aligned}$$

Our goals:

- minimize the cost function: ★ $\boxed{\underset{\boldsymbol{x}}{\text{minimize}} \quad f(\boldsymbol{x})}$
- while satisfying the constraints by **penalizing** them:

penalty for not satisfying inequality $y_i(x) \le 0$

$$\boxed{y_i(\boldsymbol{x}) \le 0} \Longrightarrow \boxed{\underset{\boldsymbol{x},\lambda_i}{\text{minimize}}} \ \lambda_i y_i(\boldsymbol{x}), \text{ where } \boxed{\lambda_i \ge 0.}$$

$$h_i(\boldsymbol{x}) = 0 \Longrightarrow \underset{\boldsymbol{x},\nu_i}{\text{minimize}} \ \nu_i h_i(\boldsymbol{x}), \text{ where } \nu_i \ge 0.$$

penalty for not satisfying $h_i(x) = 0$

- let's combine these using **regularized** minimization:

$$\star \ \boxed{\min_{\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})} = f(\boldsymbol{x}) + \sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x}) + \sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x}) = \underbrace{f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{y}(\boldsymbol{x}) + \boldsymbol{\nu}^\top \boldsymbol{h}(\boldsymbol{x})}.$$

- So, Lagrangian is the relaxation of optimization problem to an unconstrained problem.

regularized cost

soft restriction

Interpretation of Lagrangian

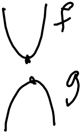
$$\min_{x} \underbrace{f(x)}_{\text{cost}} + \underbrace{\lambda}_{\text{reg parameter}} \underbrace{\ell(x)}_{\text{penalty (regularization term)}} \qquad \boxed{\lambda \geq 0}$$

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & y_i(\boldsymbol{x}) \leq 0, \; i \in \{1, \dots, m_1\}, \\ & h_i(\boldsymbol{x}) = 0, \; i \in \{1, \dots, m_2\}. \end{aligned}$$

- We want to minimize the objective function $f(\boldsymbol{x})$. We create a cost function consisting of the objective function.
- The optimization problem has constraints so its constraints should also be satisfied while minimizing the objective function. Therefore, we penalize the cost function if the constraints are not satisfied.
- For this, we can add the constraints to the objective function as the regularization (or penalty) terms and we minimize the regularized cost.
- The dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ can be seen as the regularization parameters which weight the penalties compared to the objective function $f(\boldsymbol{x})$.
- This regularized cost function is the Lagrangian function or the Lagrangian relaxation of the problem (1).
- Minimization of the regularized cost function minimizes the function $f(\boldsymbol{x})$ while trying to satisfy the constraints.

$$\min_{\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) := \underbrace{f(\boldsymbol{x})}_{} + \sum_{i=1}^{m_1} \boxed{\lambda_i} y_i(\boldsymbol{x}) + \sum_{i=1}^{m_2} \boxed{\nu_i} \underbrace{h_i(\boldsymbol{x})}_{} = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{y}(\boldsymbol{x}) + \boldsymbol{\nu}^\top \boldsymbol{h}(\boldsymbol{x}).$$

Lagrange Dual Function



f: primal function
g: dual function

Definition (Lagrange dual function)

The **Lagrange dual function** (also called the **dual function**) $\underbrace{g}$: $\underbrace{\mathbb{R}^{m_1}} \times \underbrace{\mathbb{R}^{m_2}} \to \mathbb{R}$ is defined as:

not a function of x

$$\boxed{g(\boldsymbol{\lambda}, \boldsymbol{\nu})} \overset{\text{define}}{:=} \inf_{\boldsymbol{x}\in\mathcal{D}} \underbrace{\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})} \xrightarrow{\text{min}} \mathcal{L}^*$$

$\boldsymbol{\lambda} \in \mathbb{R}^{m_1}$, $\boldsymbol{\nu} \in \mathbb{R}^{m_2}$

$$= \inf_{\boldsymbol{x}\in\mathcal{D}} \underbrace{\left(f(\boldsymbol{x}) + \sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x}) + \sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x}) \right)}. \quad (3)$$

- For **convex minimization**: the dual function g is a **concave** function. We will see later that we maximize this concave function in a so-called **dual problem**.
- For **convex maximization**: the dual function g is a **convex** function. We will see later that we minimize this convex function in a so-called **dual problem**.

Primal and Dual Feasibility

Primal Feasibility

$$
\begin{aligned}
&\underset{\boldsymbol{x}}{\text{minimize}} && f(\boldsymbol{x}) \\
&\text{subject to} && y_i(\boldsymbol{x}) \leq 0, \; i \in \{1, \dots, m_1\}, \\
& && h_i(\boldsymbol{x}) = 0, \; i \in \{1, \dots, m_2\}.
\end{aligned}
$$

Definition (The optimal point and the optimum)

The solution of this optimization problem is the **optimal point** denoted by $\boldsymbol{x}^*$. The minimum function from this solution, i.e., $f^* := f(\boldsymbol{x}^*)$, is called the **optimum function** of this problem.

The optimal point $\boldsymbol{x}^*$ is one of the **feasible points** which minimizes function $f(.)$ with constraints in problem (1). Hence, the optimal point is a feasible point:

$$
\begin{cases}
y_i(\boldsymbol{x}^*) \leq 0, \quad \forall i \in \{1, \dots, m_1\}, & (4) \\
h_i(\boldsymbol{x}^*) = 0, \quad \forall i \in \{1, \dots, m_2\}. & (5)
\end{cases}
$$

These are called the **primal feasibility**.

Feasibility in dual function

x^* solution of primal problem

$$\left\{\begin{array}{ll}\underset{\boldsymbol{x}}{\text{minimize}} & f(\boldsymbol{x}) \\ \text{subject to} & y_i(\boldsymbol{x}) \leq 0,\ i \in \{1, \dots, m_1\}, \\ & h_i(\boldsymbol{x}) = 0,\ i \in \{1, \dots, m_2\}.\end{array}\right.$$

$$\star\quad \underline{\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})} := f(\boldsymbol{x}) + \underbrace{\sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x}) + \sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x})} = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{y}(\boldsymbol{x}) + \boldsymbol{\nu}^\top \boldsymbol{h}(\boldsymbol{x}).$$

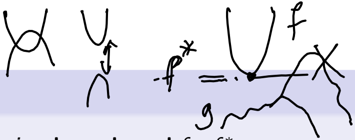
- The optimal point $\boldsymbol{x}^*$ minimizes the Lagrangian function because Lagrangian is the relaxation of optimization problem to an unconstrained problem.
- On the other hand, according to Eq. (3),

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

the dual function is the minimum of Lagrangian w.r.t. $\boldsymbol{x}$. Hence, we can write the dual function as:

$$\star\quad \underbrace{\boxed{g(\boldsymbol{\lambda}, \boldsymbol{\nu})} \overset{(3)}{=} \boxed{\inf_{\boldsymbol{x} \in \mathcal{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}}_{\text{def of dual function}} = \boxed{\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\nu}).}\quad x^* = \arg\inf_{x} \mathcal{L}(x, \lambda, \nu) \tag{6}$$

Dual Feasibility



Lemma (Dual function as a lower bound)

For **minimization** *problem, if* $\boldsymbol{\lambda} \succeq \mathbf{0}$, *then the dual function is a* **lower bound** *for* f^**:*

all elements of λ ↓ $\lambda_1, \dots, \lambda_{m_1} \geq 0$

$$\boxed{g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f^*.} \tag{7}$$

→ $g^* \leq f^*$

Proof.

Let $\lambda \succeq \mathbf{0}$ which means $\lambda_i \geq 0, \forall i$. Consider a feasible $\widetilde{\boldsymbol{x}}$ for problem (1). We have:

$$\rightarrow \mathcal{L}(\widetilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \overset{(2)}{=} f(\widetilde{\boldsymbol{x}}) + \underbrace{\sum_{i=1}^{m_1} \underbrace{\lambda_i}_{\geq 0} \underbrace{y_i(\widetilde{\boldsymbol{x}})}_{\leq 0}}_{\leq 0} + \overbrace{\cancel{\sum_{i=1}^{m_2} \nu_i \underbrace{h_i(\widetilde{\boldsymbol{x}})}_{=0}}}^{0} \leq f(\widetilde{\boldsymbol{x}}). \tag{8}$$

$\begin{cases} y_i(x) \leq 0 \\ h_i(x) = 0 \end{cases}$

Therefore, we have:

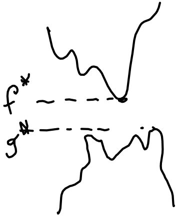
$$\rightarrow f(\widetilde{\boldsymbol{x}}) \overset{(8)}{\geq} \mathcal{L}(\widetilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\boldsymbol{x} \in \mathcal{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \overset{(3)}{=} g(\boldsymbol{\lambda}, \boldsymbol{\nu}).$$

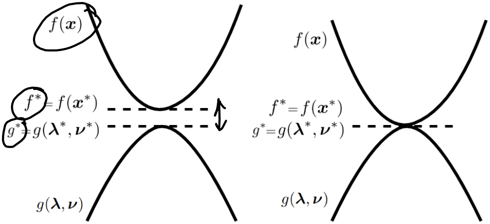
$\min_x f(x) \leq f(x) \quad \forall x \in D$

Hence, the dual function is a lower bound for the function of all feasible points. As the optimal point $\boldsymbol{x}^*$ is a feasible point, the dual function is a lower bound for f^*. ☐

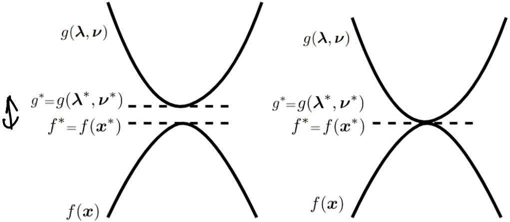
Dual function as a lower bound

For **minimization** problem: $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f^*$





For **maximization** problem: $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \geq f^*$



Nonnegativity of dual variables for inequality constraints

- From the above lemma, we conclude that for having the dual function as a lower bound for the optimum function, the dual variable $\{\lambda_i\}_{i=1}^{m_1}$ for inequality constraints (less than or equal to zero) should be non-negative, i.e.:

$$\star\ \boxed{\boldsymbol{\lambda} \succeq \mathbf{0}} \xrightarrow{\quad} \leqslant \quad \text{or} \quad \underbrace{\lambda_i \geq 0,\ \forall i \in \{1, \dots, m_1\}.} \tag{9}$$

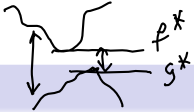
- We assume that the inequality constraints are less than or equal to zero. If some of the inequality constraints are greater than or equal to zero, we convert them to less than or equal to zero by multiplying them to -1. Or, if the inequality constraints are greater than or equal to zero, we should have $\lambda_i \leq 0, \forall i$ because $\underbrace{y_i(\boldsymbol{x}) \boxed{\geq} 0} \implies \underbrace{-y_i(\boldsymbol{x}) \leq 0}.$

$$\swarrow$$

$$\boxed{\lambda_i \leqslant 0}$$

The Dual Problem

Primal and dual problems



Definition (Primal and dual problems)

- We saw that the **dual function** is a **lower bound** for the optimum function: $\boxed{g(\boldsymbol{\lambda},\boldsymbol{\nu}) \leq f^*.}$
- We want to find the **best lower bound** so we **maximize** $g(\boldsymbol{\lambda},\boldsymbol{\nu})$ w.r.t. the dual variables $\boldsymbol{\lambda},\boldsymbol{\nu}$.
- Eq. (9) says that the dual variables for inequalities must be **nonnegative**: $\boldsymbol{\lambda} \succeq \mathbf{0}$
- Hence, we have the following optimization:

$$\left[\begin{array}{ll}\underset{\boldsymbol{\lambda},\boldsymbol{\nu}}{\text{maximize}} & \boxed{g(\boldsymbol{\lambda},\boldsymbol{\nu})}\\ \text{subject to} & \boxed{\boldsymbol{\lambda} \succeq \mathbf{0}.}\end{array}\right] \tag{10}$$

- The problem (10) is called the **Lagrange dual optimization problem** for problem (1).
- The problem (1) is also referred to as the **primal optimization problem**.

- The variable of problem (1), i.e. $\boldsymbol{x}$, is called the **primal variable** while the variables of problem (10), i.e. $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$, are called the **dual variables**.
- Let the solutions of the dual problem be denoted by $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$. We denote:

$f^* = f(x^*)$

$$\boxed{g^*} := \underline{g(\boldsymbol{\lambda}^*,\boldsymbol{\nu}^*)} = \boxed{\sup_{\boldsymbol{\lambda},\boldsymbol{\nu}} g.} \quad \text{s.t.} \quad \lambda \geq 0$$

Weak and strong duality

Definition (Weak and strong duality)

For all convex and nonconvex minimization problems, the optimum dual problem is a lower bound for the optimum function:

$g^* = f^*$ → strong duality

$g^* \leq f^*$ → weak duality

$$\underbrace{g^* \leq f^*}_{} \quad \text{i.e.,} \quad \underbrace{g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \leq f(\boldsymbol{x}^*).}_{} \tag{11}$$

This is called the **weak duality**. For some optimization problems, we have **strong duality** which is when the optimum dual problem is equal to the optimum function:

$$\boxed{g^* = f^*} \quad \text{i.e.,} \quad \boxed{g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = f(\boldsymbol{x}^*).} \;\bigstar \tag{12}$$

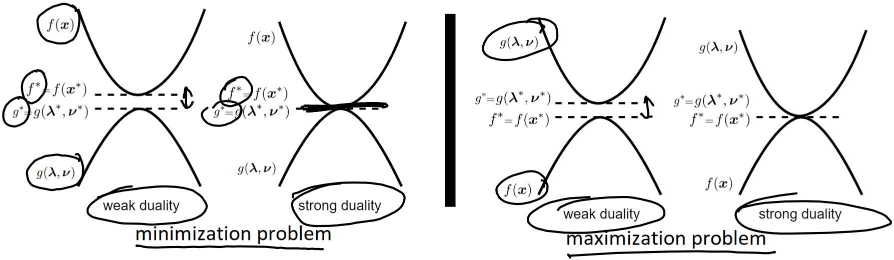
The strong duality usually holds for convex optimization problems.

Weak and strong duality

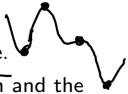
Corollary

Eqs. (11) and (12) show that, in a minimization problem, the optimum dual function always provides a lower-bound for the optimum primal function:

$$g^* \leq f^*. \tag{13}$$



Weak and strong duality in iterative optimization



- If optimization is **iterative**, the solution is updated iteratively until convergence.
- The series of primal optimal and dual optimal **converge** to the optimal solution and the dual optimal, respectively.
- in convex problem:

$$
\begin{aligned}
&\{\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots\} \to \boldsymbol{x}^*,\\
&\{\boldsymbol{\nu}^{(0)}, \boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)}, \dots\} \to \boldsymbol{\nu}^*,\\
&\{\boldsymbol{\lambda}^{(0)}, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)}, \dots\} \to \boldsymbol{\lambda}^*,\\
&\boxed{f(\boldsymbol{x}^{(0)}) \geq f(\boldsymbol{x}^{(1)}) \geq f(\boldsymbol{x}^{(2)}) \geq \dots \geq f(\boldsymbol{x}^*),}\\
&g(\boldsymbol{\lambda}^{(0)}, \boldsymbol{\nu}^{(0)}) \leq g(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\nu}^{(1)}) \leq \dots \leq g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*).\\
&g(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)}) \leq f(\boldsymbol{x}^{(k)}), \quad \forall k.
\end{aligned} \tag{14}
$$

$$\boxed{\min_x \quad f(x)} \leftarrow$$

$$\frac{\partial f(x)}{\partial x}$$

~~convex~~ set.

$$\tan(x) + \frac{\sin(x)}{x^2+1} = 0$$

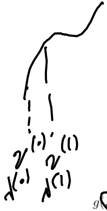
$\cancel{x_1}$

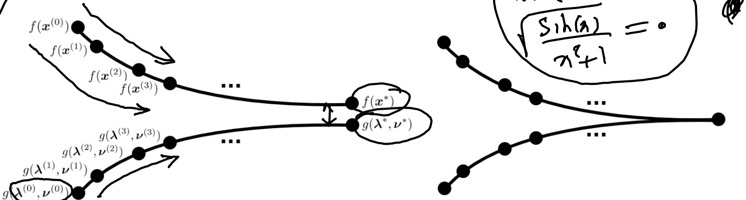
$\cancel{x_2} \to$ max

$x_3 \to$

x_4

$f(x_3) < f(x_4)$





(a) (b)

Slater's condition

$$\overbrace{\begin{bmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & & \\ a_{m_2 1} & \cdots & a_{m_2 d} \end{bmatrix}}^{\mathbb{R}^{m_2 \times d}} \overset{\mathbb{R}^{d}}{\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}} = \overset{\mathbb{R}^{m_2}}{\begin{bmatrix} b_1 \\ \vdots \\ b_{m_2} \end{bmatrix}}$$

Lemma (Slater's condition [1])

For a convex optimization problem in the form:

$$\begin{aligned} & \underset{\boldsymbol{x}}{\text{minimize}} && \boxed{f(\boldsymbol{x})} \\ & \text{subject to} && \boxed{y_i(\boldsymbol{x}) \leq 0,} \; i \in \{1, \dots, m_1\}, \\ & && \boxed{\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},} \end{aligned}$$

$$\underbrace{h_i(x) = 0 \quad i \in \{1, \dots, m_2\}}$$

we have **strong duality** *if it is strictly feasible, i.e.:*

$$\exists \boldsymbol{x} \in \underline{\textbf{int}(\mathcal{D})} : y_i(\boldsymbol{x}) < 0, \; \forall i \in \{1, \dots, m_1\}, \quad (15)$$
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

exists

f^*

g^*

In other words, for at least one point in the interior of domain (not on the boundary of domain), all the inequality constraints hold strictly. *This is called the* **Slater's condition**.

Example for Slater's condition

Example:

$$\left[\begin{aligned} &\underset{\boldsymbol{x}=[x_1,x_2]^\top}{\text{minimize}} && x_1^2 + x_2^2 + 1 \\ &\text{subject to} && x_1 + x_2 \geq 1, \\ & && x_1 - x_2 = 2. \end{aligned}\right.$$

The constraints:

$$\begin{aligned} x_1 + x_2 \geq 1 &\implies -x_1 - x_2 + 1 \leq 0, \\ x_1 - x_2 = 2 &\implies x_1 - x_2 - 2 = 0. \end{aligned}$$

So, the problem is:

$$\begin{aligned} &\underset{\boldsymbol{x}=[x_1,x_2]^\top}{\text{minimize}} && x_1^2 + x_2^2 + 1 \\ &\text{subject to} && -x_1 - x_2 + 1 \leq 0, \\ & && x_1 - x_2 - 2 = 0. \end{aligned}$$

Example for Slater's condition

$f^* = g^*$

g^*

$$
\begin{aligned}
&\underset{x=[x_1,x_2]^\top}{\text{minimize}} \quad && x_1^2 + x_2^2 + 1 \\
&\text{subject to} && -x_1 - x_2 + 1 \leq 0, \\
& && x_1 - x_2 - 2 = 0.
\end{aligned}
$$

Can we find at least one $[x_1, x_2]^\top$ to satisfy the following?

$$
\begin{cases}
-x_1 - x_2 + 1 < 0, \\
x_1 - x_2 - 2 = 0.
\end{cases}
$$

We have:

$$x_1 - x_2 - 2 = 0 \implies x_1 = x_2 + 2 \implies -(x_2 + 2) - x_2 + 1 < 0 \implies x_2 > -0.5$$

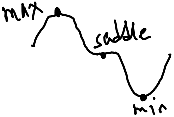
We can take $x_2 = 0$, then:

$$x_1 = x_2 + 2 = 0 + 2 = 2$$

So, we can find $x_1 = 2, x_2 = 0$ which satisfies the Slater's condition. Therefore, strong duality holds for this problem.

Stationarity Condition

Stationarity condition



$\min_{x} f(x)$
w.r.t. $y_i(x)$
$h_i(x)$

★★ $$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) := \underbrace{f(\boldsymbol{x})} + \underbrace{\sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x})} + \underbrace{\sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x})}.$$

- As was explained before, the Lagrangian function can be interpreted as a regularized cost function to be minimized. Hence:

★ $$\boxed{\underset{\boldsymbol{x}}{\text{minimize}}\ \mathcal{L}(\boldsymbol{x}, \underline{\boldsymbol{\lambda}, \boldsymbol{\nu}}).} \qquad \frac{\partial \mathcal{L}}{\partial x} \tag{16}$$

- We can find its minimum by setting its derivative w.r.t. $\boldsymbol{x}$, denoted by $\boxed{\nabla_{\boldsymbol{x}} \mathcal{L}}$, to zero:

$$\underbrace{\nabla_{x} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})} \overset{set}{=} 0 \implies \boxed{\underbrace{\nabla_{\boldsymbol{x}} f(\boldsymbol{x})} + \sum_{\underbrace{i=1}}^{m_1} \lambda_i \nabla_{\boldsymbol{x}} y_i(\boldsymbol{x}) + \underbrace{\sum_{i=1}^{m_2} \nu_i \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x})} \overset{set}{=} 0.} \ ★ \tag{17}$$

- This equation is called the **stationarity condition** because this shows that the gradient of Lagrangian w.r.t. $\boldsymbol{x}$ should vanish to zero (note that a stationary point of a function is a point where the derivative of function is zero).
- ★ This derivative holds for all dual variables and not just for the optimal dual variables.

λ_i^* ν_i^* → $\max g(\lambda, \nu)$

Complementary Slackness

Complementary slackness

- Assume that the problem has strong duality with $\underline{\boldsymbol{x}^*}$, $\underline{\boldsymbol{\lambda}^*}$, and $\underline{\boldsymbol{\nu}^*}$.
- According to Eq. (12), $\underline{g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = f(\boldsymbol{x}^*)}$, and Eq. (3), $\underline{g(\boldsymbol{\lambda}, \boldsymbol{\nu}) := \inf_{\boldsymbol{x} \in \mathcal{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}$, we have:

$$\bigstar\ f(\boldsymbol{x}^*) \overset{(12)}{=} g(\underline{\boldsymbol{\lambda}}^*, \underline{\boldsymbol{\nu}}^*) \overset{(3)}{=} \inf_{\boldsymbol{x} \in \mathcal{D}} \overbrace{\Big(f(\boldsymbol{x}) + \sum_{i=1}^{m_1} \lambda_i^* y_i(\boldsymbol{x}) + \sum_{i=1}^{m_2} \nu_i^* h_i(\boldsymbol{x}) \Big)}^{\mathcal{L}(x, \lambda^*, \nu^*)}$$

$$\overset{(a)}{=} f(\underline{\boldsymbol{x}^*}) + \sum_{i=1}^{m_1} \lambda_i^* y_i(\underline{\boldsymbol{x}^*}) + \overset{0}{\cancel{\sum_{i=1}^{m_2} \nu_i^* h_i(\boldsymbol{x}^*)}} \overset{(b)}{=} \underbrace{f(\boldsymbol{x}^*)} + \underbrace{\sum_{i=1}^{m_1} \overset{\geq 0}{\lambda_i^*} \overset{\leq 0}{y_i(\boldsymbol{x}^*)}}_{\leq 0} \overset{(c)}{\leq} \underbrace{f(\boldsymbol{x}^*)}, \quad (18)$$

where (a) is because $\boldsymbol{x}^*$ is the primal optimal solution for problem <u>(1) and it minimizes the Lagrangian</u>, (b) is because $\boldsymbol{x}^*$ is a feasible point and satisfies $\boxed{h_i(\boldsymbol{x}^*) = 0}$, and (c) is because $\underbrace{\lambda_i^* \geq 0}$ according to <u>Eq. (9)</u> and the feasible $\boldsymbol{x}^*$ satisfies $\underbrace{y_i(\boldsymbol{x}^*) \leq 0}$, so we have:

$$\bigstar\quad \underbrace{\lambda_i^* y_i(\boldsymbol{x}^*) \leq 0}, \quad \underline{\forall i \in \{1, \dots, m_1\}}. \quad (19)$$

$\lambda_i^* \geq 0 \qquad y_i(x^*) \leq 0$

Complementary slackness

- We obtained Eqs. (18) and (19):

$$* \quad f(\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + \sum_{i=1}^{m_1} \lambda_i^* y_i(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^*),$$

$$** \quad \underbrace{\lambda_i^* y_i(\boldsymbol{x}^*) \leq 0,} \quad \forall i \in \{1, \dots, m_1\}.$$

- Therefore:

$$\overbrace{f(\boldsymbol{x}^*)} = \overbrace{f(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m_1} \lambda_i^* y_i(\boldsymbol{x}^*)}} \leq \overbrace{f(\boldsymbol{x}^*)}$$

$$\Longrightarrow \underbrace{\sum_{i=1}^{m_1} \lambda_i^* y_i(\boldsymbol{x}^*) = 0} \overset{(19)}{\Longrightarrow} \underbrace{\lambda_i^* y_i(\boldsymbol{x}^*) = 0, \forall i.} \Leftarrow \{1, \dots, m_1\}$$

- Therefore, the multiplication of every optimal dual variable λ_i^* with $y_i(.)$ of optimal primal solution $\boldsymbol{x}^*$ must be zero. This is called the <u>**complementary slackness**</u>:

$$\boxed{\lambda_i^* \overbrace{y_i(\boldsymbol{x}^*)} = 0,} \quad \underline{\forall i \in \{1, \dots, m_1\}.} \tag{20}$$

Complementary slackness

$y_i(x^*) \leq 0$

$$\begin{cases} y_i(x^*) = 0 \\ y_i(x^*) < 0 \end{cases}$$

$\lambda_i^* \geq 0$

$$\begin{cases} \lambda_i^* > 0 \\ \lambda_i^* = 0 \end{cases}$$

- The **complementary slackness**:

$$\ast \quad \lambda_i^* \, y_i(\boldsymbol{x}^*) = 0, \quad \forall i \in \{1, \ldots, m_1\}.$$

- These conditions can be restated as:

$$\ast \begin{cases} \underline{\lambda_i^* > 0} \implies \underline{y_i(\boldsymbol{x}^*) = 0}, & (21) \\ \underline{y_i(\boldsymbol{x}^*) < 0} \implies \underline{\lambda_i^* = 0}, & (22) \end{cases}$$

which means that, for an inequality constraint,

- If the dual optimal is nonzero, its inequality function of the primal optimal must be zero.
- If the inequality function of the primal optimal is nonzero, its dual optimal must be zero.

KKT Conditions

Karush-Kuhn-Tucker (KKT) conditions

- In previous slides, we derived the **primal feasibility**, **dual feasibility**, **stationarity condition**, and **complementary slackness**. These four conditions are called the **Karush-Kuhn-Tucker (KKT) conditions** [2, 3].
- The primal optimal variable $\boldsymbol{x}^*$ and the dual optimal variables $\boldsymbol{\lambda}^* = [\lambda_1^*, \ldots, \lambda_{m_1}^*]^\top$, $\boldsymbol{\nu}^* = [\nu_1^*, \ldots, \nu_{m_2}^*]^\top$ must satisfy the KKT conditions.

Karush-Kuhn-Tucker (KKT) conditions

We summarize the KKT conditions in the following:

1. **Stationarity condition**:

$$\bigstar \quad \underline{\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) + \sum_{i=1}^{m_1} \lambda_i \nabla_{\boldsymbol{x}} y_i(\boldsymbol{x}) + \sum_{i=1}^{m_2} \nu_i \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}) = 0. \tag{23}$$

2. **Primal feasibility**:

$$\bigstar \quad y_i(\boldsymbol{x}^*) \leq 0, \quad \underline{\forall i} \in \{1, \dots, m_1\}, \tag{24}$$

$$\bigstar \quad h_i(\boldsymbol{x}^*) = 0, \quad \underline{\forall i} \in \{1, \dots, m_2\}. \tag{25}$$

3. **Dual feasibility**:

$$\underline{\boldsymbol{\lambda} \succeq \mathbf{0}} \quad \text{or} \quad \underline{\lambda_i \geq 0, \ \forall i \in \{1, \dots, m_1\}}. \quad \bigstar \tag{26}$$

4. **Complementary slackness**:

$$\boxed{\lambda_i^* \, y_i(\boldsymbol{x}^*) = 0,} \quad \underline{\forall i \in \{1, \dots, m_1\}}. \quad \bigstar \tag{27}$$

As listed above, KKT conditions impose constraints on the optimal dual variables of inequality constraints because the sign of inequalities is important.

Karush-Kuhn-Tucker (KKT) conditions

- Recall the dual problem (10):

$$\left[\begin{array}{ll}\underset{\boldsymbol{\lambda},\boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda},\boldsymbol{\nu}) \\ \text{subject to} & \boxed{\boldsymbol{\lambda} \succeq \mathbf{0}.}\end{array}\right.$$

The constraint in this problem is already satisfied by the dual feasibility in the KKT conditions. Hence, we can ignore the constraint of the dual problem:

$$\boxed{\underset{\boldsymbol{\lambda},\boldsymbol{\nu}}{\text{maximize}} \quad g(\boldsymbol{\lambda},\boldsymbol{\nu}),} \longrightarrow \lambda^*, \nu^* \tag{28}$$

which should give us $\boldsymbol{\lambda}^*$, $\boldsymbol{\nu}^*$, and $g^* = g(\boldsymbol{\lambda}^*,\boldsymbol{\nu}^*)$.

- This is an unconstrained optimization problem and for solving it, we should set the derivative of $g(\boldsymbol{\lambda},\boldsymbol{\nu})$ w.r.t. $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ to zero:

$$\nabla_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda},\boldsymbol{\nu}) \overset{\text{set}}{=} 0 \overset{(6)}{\Longrightarrow} \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda},\boldsymbol{\nu}) \overset{\text{set}}{=} 0. \rightarrow y_i(x) \tag{29}$$

$$\nabla_{\boldsymbol{\nu}} g(\boldsymbol{\lambda},\boldsymbol{\nu}) \overset{\text{set}}{=} 0 \overset{(6)}{\Longrightarrow} \nabla_{\boldsymbol{\nu}} \mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda},\boldsymbol{\nu}) \overset{\text{set}}{=} 0. \rightarrow h_i(x) \tag{30}$$

- Setting the derivatives of Lagrangian w.r.t. dual variables always gives back the corresponding constraints in the primal optimization problem.

Karush-Kuhn-Tucker (KKT) conditions

- Eqs. (23), (29), and (30) state that the primal and dual residuals must be zero:

$$\star\begin{cases} \underline{\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu})} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) + \sum_{i=1}^{m_1} \lambda_i \nabla_{\boldsymbol{x}} y_i(\boldsymbol{x}) + \sum_{i=1}^{m_2} \nu_i \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}) = 0, \\ \underline{\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda},\boldsymbol{\nu})} = 0, \\ \underline{\nabla_{\boldsymbol{\nu}}\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda},\boldsymbol{\nu})} = 0. \end{cases}$$

- Eqs. (3) and (28) can be summarized into the following max-min optimization problem:

$$\underline{\sup_{\boldsymbol{\lambda},\boldsymbol{\nu}} g(\boldsymbol{\lambda},\boldsymbol{\nu})} \overset{(3)}{=} \underline{\sup_{\boldsymbol{\lambda},\boldsymbol{\nu}} \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu})} = \underline{\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\nu}^*)}. \tag{31}$$

$$\begin{cases} \min(\max \mathcal{L}) \\ \max(\min \mathcal{L}) \end{cases}$$

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

inf - sup $\Leftarrow$ min - max

sup - inf $\Leftarrow$ max - min

Where the KKT name came from?

- The reason for the name KKT is as follows [4].
- In 1952, Kuhn and Tucker published an important paper proposing the conditions [3].
- However, later it was found out that there is a master's thesis by Karush, in 1939, at the University of Chicago, Illinois [2].
- That thesis had also proposed the conditions; however, researchers including Kuhn and Tucker were not aware of that thesis. Therefore, these conditions were named after all three of them.

Method of Lagrange Multipliers

Method of Lagrange multipliers

$\boxed{:=}$ define

$\boxed{:=}$ ← assign

We can solve the optimization problem (1):

$$\star\star\star\quad \begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & y_i(\boldsymbol{x}) \leq 0,\ i \in \{1, \dots, \underline{m_1}\}, \\ & h_i(\boldsymbol{x}) = 0,\ i \in \{1, \dots, \underline{m_2}\}, \end{aligned} \quad \longrightarrow x^*,\ f(x^*)$$

using <u>duality and KKT conditions</u>. This technique is also called the **method of Lagrange multipliers**. For this, we should do the following steps:

1. We write the Lagrangian as Eq. (2):

$$\star \quad \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) := \underbrace{f(\boldsymbol{x})}_{} + \underbrace{\sum_{i=1}^{m_1} \lambda_i y_i(\boldsymbol{x})}_{} + \underbrace{\sum_{i=1}^{m_2} \nu_i h_i(\boldsymbol{x})}_{} = \underbrace{f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{y}(\boldsymbol{x}) + \boldsymbol{\nu}^\top \boldsymbol{h}(\boldsymbol{x})}_{}.$$

2. We consider the <u>dual function</u> defined in Eq. (3) and we solve it:

$$\underbrace{\boldsymbol{x}^\dagger}_{} := \arg \underbrace{\min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}_{}. \tag{32}$$

It is an <u>unconstrained problem</u> and according to Eqs. (3) and (23), we solve this problem by taking the derivative of Lagrangian w.r.t. $\boldsymbol{x}$ and setting it to zero, i.e., $\underbrace{\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}_{} \overset{\text{set}}{=} 0$. This gives us the dual function: ← $g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)$

$$\boxed{g(\boldsymbol{\lambda}, \boldsymbol{\nu})} = \underbrace{\mathcal{L}(\boldsymbol{x}^\dagger, \boldsymbol{\lambda}, \boldsymbol{\nu})}_{}. \tag{33}$$

Method of Lagrange multipliers

3. We consider the dual problem, defined in Eq. (10) which is simplified to Eq. (28) because of Eq. (26). This gives us the optimal dual variables $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$:

$$\star \quad \underbrace{\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*} := \underbrace{\arg\max_{\boldsymbol{\lambda},\boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu})}. \tag{34}$$

It is an unconstrained problem and according to Eqs. (29) and (30), we solve this problem by taking the derivative of dual function w.r.t. $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ and setting them to zero, i.e., $\nabla_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \overset{\text{set}}{=} 0$ and $\nabla_{\boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \overset{\text{set}}{=} 0$. The optimum dual value is obtained as:

$$\star \quad \boxed{g^*} = \max_{\boldsymbol{\lambda},\boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \underbrace{g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)}. \tag{35}$$

4. We put the optimal dual variables $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ in Eq. (23) to find the optimal primal variable:

$$\boxed{\boldsymbol{x}^*} := \underbrace{\arg\min_{\boldsymbol{x}} \overbrace{\mathcal{L}(\boldsymbol{x}, \boxed{\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*})}}. \tag{36}$$

$\star$

It is an unconstrained problem and we solve this problem by taking the derivative of Lagrangian at optimal dual variables w.r.t. $\boldsymbol{x}$ and setting it to zero, i.e., $\underbrace{\boxed{\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)}}_{\downarrow\; x^*} \overset{\text{set}}{=} 0$. The optimum primal value is obtained as:

$$\boxed{f^*} = f(\boldsymbol{x}^*). \tag{37}$$

Example for Method of Lagrange multipliers

$f(\underline{x}) = f(x_1, x_2)$

Example:

$$\star\ \boxed{\begin{aligned} &\underset{\boldsymbol{x}=[\underline{x_1,x_2}]^\top}{\text{minimize}} && \boxed{x_1^2+x_2^2+1} \\ &\text{subject to} && x_1+x_2\geq 1, \\ &&& x_1-x_2=2. \end{aligned}}$$

$$\min_{x_1,x_2} \quad x_1^2+x_2^2+1$$

$$\text{s.t.} \quad -x_1-x_2+1\leq 0$$

$$x_1-x_2-2=0$$

The constraints:

$$\boxed{x_1+x_2\geq 1} \xRightarrow{\times -1} \underbrace{-x_1-x_2+1\leq \boxed{0},}$$

$$\underbrace{x_1-x_2=2} \implies x_1-x_2-2=\boxed{0}.$$

The Lagrangian is:

$$\star\ \underbrace{\mathcal{L}(\boldsymbol{x},\lambda,\nu)} = \underbrace{x_1^2+x_2^2+1} + \underbrace{\lambda(-x_1-x_2+1)} + \underbrace{\nu(x_1-x_2-2)}.$$

$$\star\ \nabla_{\boxed{\boldsymbol{x}}}\mathcal{L} = \underbrace{\begin{bmatrix}\nabla_{x_1}\mathcal{L} \\ \nabla_{x_2}\mathcal{L}\end{bmatrix}} = \begin{bmatrix}2x_1-\lambda+\nu \\ 2x_2-\lambda-\nu\end{bmatrix} \overset{\text{set}}{=} \begin{bmatrix}0 \\ 0\end{bmatrix}.$$

Therefore:

$$\begin{cases} 2x_1-\lambda+\nu=0 \\ 2x_2-\lambda-\nu=0 \end{cases} \implies x_1^\dagger = \frac{\lambda-\nu}{2}, \quad x_2^\dagger = \frac{\lambda+\nu}{2}.$$

Example for Method of Lagrange multipliers

$$
\begin{aligned}
\mathcal{L}(\overset{\lambda,\nu}{\boldsymbol{x}^\dagger}, \lambda, \nu) &= (x_1^\dagger)^2 + (x_2^\dagger)^2 + 1 + \lambda(-x_1^\dagger - x_2^\dagger + 1) + \nu(x_1^\dagger - x_2^\dagger - 2) \\
&= (\frac{\lambda - \nu}{2})^2 + (\frac{\lambda + \nu}{2})^2 + 1 + \lambda(-\frac{\lambda - \nu}{2} - \frac{\lambda + \nu}{2} + 1) + \nu(\frac{\lambda - \nu}{2} - \frac{\lambda + \nu}{2} - 2) \\
&= \underbrace{-\frac{1}{2}\lambda^2 - \frac{1}{2}\nu^2 + \lambda - 2\nu + 1}
\end{aligned}
$$

Hence:

$$
\bigstar \quad \underbrace{g(\lambda, \nu)} = \mathcal{L}(\boldsymbol{x}^\dagger, \lambda, \nu) = \overbrace{\underbrace{-\frac{1}{2}\lambda^2 - \frac{1}{2}\nu^2 + \lambda - 2\nu + 1}}.
$$

We have:

$$
\begin{cases}
\underline{\nabla_\lambda g(\lambda, \nu)} = -\lambda + 1 \overset{\text{set}}{=} 0 \\
\underbrace{\nabla_\nu g(\lambda, \nu)} = -\nu - 2 \underset{\text{set}}{=} 0
\end{cases}
\implies \underbrace{\lambda^* = 1, \quad \nu^* = -2.}
$$

Example for Method of Lagrange multipliers

We found $\lambda^* = 1, \quad \nu^* = -2$. Therefore:

$$\left.\begin{aligned}\mathcal{L}(\boldsymbol{x}, \lambda^*, \nu^*) &= x_1^2 + x_2^2 + 1 + \lambda^*(-x_1 - x_2 + 1) + \nu^*(x_1 - x_2 - 2) \\ &= x_1^2 + x_2^2 + 1 + (1)(-x_1 - x_2 + 1) + (-2)(x_1 - x_2 - 2)\end{aligned}\right\}$$

$$= \underbrace{x_1^2 + x_2^2 - 3x_1 + x_2 + 6}.$$

We have:

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda^*, \nu^*) = \begin{bmatrix} \nabla_{x_1} \mathcal{L}(\boldsymbol{x}, \lambda^*, \nu^*) \\ \nabla_{x_2} \mathcal{L}(\boldsymbol{x}, \lambda^*, \nu^*) \end{bmatrix} = \begin{bmatrix} 2x_1 - 3 \\ 2x_2 + 1 \end{bmatrix} \overset{set}{=} \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$$\begin{cases} 2x_1 - 3 = 0 \\ 2x_2 + 1 = 0 \end{cases} \implies \underbrace{x_1^* = \frac{3}{2}}, \quad \underbrace{x_2^* = -\frac{1}{2}}.$$

Example for Method of Lagrange multipliers

In summary, we have:

$$\underbrace{x_1^* = \frac{3}{2}, \quad x_2^* = -\frac{1}{2},} \quad \underbrace{\lambda^* = 1, \quad \nu^* = -2.}$$

$$\bigstar \quad f^* = f(\boldsymbol{x}^*) = \underbrace{3(x_1^*)^2 + 2(x_2^*)^2 + 1} = 3(\frac{3}{2})^2 + 2(-\frac{1}{2})^2 + 1 = \boxed{3.5} \rightarrow f^* = 3,5$$

$$\bigstar \quad \boxed{g^*} = \underbrace{g(\lambda^*, \nu^*)} = \mathcal{L}(\boldsymbol{x}^\dagger, \lambda^*, \nu^*) = \underbrace{-\frac{1}{2}(\lambda^*)^2 - \frac{1}{2}(\nu^*)^2 + \lambda^* - 2\nu^* + 1}$$

$$= -\frac{1}{2}(1)^2 - \frac{1}{2}(-2)^2 + 1 - 2(-2) + 1 = \boxed{3.5} \Longrightarrow g^* = 3.5$$

It is expected to have $f^* = g^*$, as we have strong duality (we saw in previous slides that this problem satisfies Slater's condition; see the example we provided for Slater's condition).

$$\boxed{g^* = f^*} \qquad\qquad \boxed{g^* \leq f^*}$$

References

[1] M. Slater, "Lagrange multipliers revisited," tech. rep., Cowles Commission Discussion Paper: Mathematics 403, Yale University, 1950.

[2] W. Karush, "Minima of functions of several variables with inequalities as side constraints," Master's thesis, Department of Mathematics, University of Chicago, Chicago, Illinois, 1939.

[3] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492, Berkeley: University of California Press, 1951.

[4] T. H. Kjeldsen, "A contextualized historical analysis of the Kuhn–Tucker theorem in nonlinear programming: the impact of world war II," *Historia mathematica*, vol. 27, no. 4, pp. 331–361, 2000.