

Optimization Theory and Algorithms

Instructor: Prof. LIAO, Guocheng (廖国成)

Email: liaogch6@mail.sysu.edu.cn

**School of Software Engineering
Sun Yat-sen University**

Outline

- Lagrange dual problem
- Weak duality and strong duality
- KKT conditions
- Saddle point

Motivation of duality theory

- Helps analyze and even solve the original difficult problem from an **easier** dual problem
- Obtain some **properties** of the original problem by analyzing dual problem
- Sensitivity analysis

Lagrangian

Standard form optimization problem

$$\begin{array}{ll}\min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p\end{array}$$

variable $x \in \mathbb{R}^n$; optimal value p^* ; not necessarily convex

Lagrangian: $L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0, i = 1, \dots, m$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0, i = 1, \dots, p$
- Lagrangian: objective function + weighted sum of constraint functions

Lagrangian dual function

Lagrange dual function (or just *dual function*): $g: \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$

$$\begin{aligned} g(\lambda, \nu) &= \inf_x L(x, \lambda, \nu) \\ &= \inf_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \end{aligned}$$

Dual function is the pointwise infimum of affine functions of (λ, ν) , so it is **concave**.

Lower bound property: for any $\lambda \geq 0$ and any ν , we have

$$g(\lambda, \nu) \leq p^*$$

Proof: let x' is feasible, i.e., $f_i(x') \leq 0$ and $h_i(x') = 0$:

$$\begin{aligned} g(\lambda, \nu) &= \inf_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \\ &\leq f_0(x') + \sum_{i=1}^m \lambda_i f_i(x') + \sum_{i=1}^p \nu_i h_i(x') \\ &\leq f_0(x') \end{aligned}$$

$g(\lambda, \nu) \leq f_0(x')$ holds for any feasible x' . Thus, $g(\lambda, \nu) \leq f_0(x^*) = p^*$.

Lagrange dual problem

Motivation: to make the lower bound $g(\lambda, \nu)$ of p^* as **large** as possible

Lagrange dual problem (or just *dual problem*):

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

- Dual problem is a convex problem (concave function maximization subject to convex constraint function)
- (λ, ν) is **dual feasible** if $\lambda \geq 0$ and $g(\lambda, \nu) > -\infty$
- (λ^*, ν^*) is dual optimal (or optimal Lagrange multipliers) if they are optimal for the dual problem

Primal problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

Dual problem

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

Examples

primal problem
(standard form LP)

$$\begin{array}{ll} \min & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \quad \longleftrightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{subject to} & b - Ax = 0 \\ & -x \leq 0 \end{array}$$

Lagrangian

$$\begin{aligned} L(x, \lambda, v) &= c^T x - \lambda^T x + v^T (b - Ax) \\ &= b^T v + (c - A^T v - \lambda)^T x \end{aligned}$$

- λ_i is associated with inequality constraint $f_i(x) = -x_i \leq 0, i = 1, \dots, n$
- v_i is associated with equality constraint $f_i(x) = b_i - a_i^T x, i = 1, \dots, m$

Dual function

$$g(\lambda, v) = \inf_x L(x, \lambda, v) = b^T v + \inf_x (c - A^T v - \lambda)^T x = \begin{cases} b^T v, & c - A^T v - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Dual problem

$$\begin{array}{ll} \max & g(\lambda, v) = \begin{cases} b^T v, & c - A^T v - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases} \\ \text{s.t.} & \lambda \geq 0 \end{array} \quad \longrightarrow \quad \begin{array}{ll} \max & b^T v \\ \text{s.t.} & \lambda \geq 0 \\ & c - A^T v - \lambda = 0 \end{array} \quad \longrightarrow \quad \begin{array}{ll} \max & b^T v \\ \text{s.t.} & A^T v \leq c \end{array}$$

Examples

primal problem
(inequality form LP)

$$\begin{array}{ll} \min & c^T x \\ \text{subject to} & Ax \preceq b \end{array} \quad \longleftrightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{subject to} & Ax - b \preceq 0 \end{array}$$

Lagrangian

$$\begin{aligned} L(x, \lambda) &= c^T x + \lambda^T (Ax - b) \\ &= -b^T \lambda + (c + A^T \lambda)^T x \end{aligned}$$

Dual function

$$g(\lambda) = \inf_x L(x, \lambda) = -b^T \lambda + \inf_x (c + A^T \lambda)^T x = \begin{cases} -b^T \lambda, & c + A^T \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Dual problem

$$\begin{array}{ll} \max & g(\lambda, \nu) = \begin{cases} -b^T \lambda, & c + A^T \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases} \\ \text{s.t.} & \lambda \geq 0 \end{array} \quad \longrightarrow \quad \begin{array}{ll} \max & -b^T \lambda \\ \text{s.t.} & c + A^T \lambda = 0 \\ & \lambda \geq 0 \end{array}$$

Examples

primal problem (quadratic programming)

$$\begin{array}{ll}\text{minimize} & x^T x \\ \text{subject to} & Ax = b\end{array}$$

Lagrangian

$$L(x, \nu) = x^T x + \nu^T (Ax - b)$$

Dual function

- Take the gradient with respect to x , and set the gradient equal to zero

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \implies \quad x = -(1/2)A^T \nu$$

- Plug in L to get g :

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

Dual problem

$$\max \quad -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

Examples

primal problem (non-convex)

$$\begin{array}{ll}\text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n\end{array}$$

Lagrangian

$$\begin{aligned}L(x, \nu) &= x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) \\ &= x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu.\end{aligned}$$

Dual function

$$\begin{aligned}g(\nu) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}\end{aligned}$$

The infimum of a quadratic form is either zero (positive semidefinite) or $-\infty$ (not positive semidefinite)

Dual problem

$$\begin{array}{ll}\text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0\end{array}$$

Take $\nu = -\lambda_{\min}(W)\mathbf{1}$, we get a lower bound $p^* \geq n\lambda_{\min}(W)$

Primal problem v.s. dual problem

Primal problem

$$\begin{array}{ll}\min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p\end{array}$$

Dual problem

$$\begin{array}{ll}\max & g(\lambda, v) \\ \text{s.t.} & \lambda \geq 0\end{array}$$

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p v_i h_i(x)$$

Dual problem:

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

Primal problem:

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



$$\begin{array}{ll}\min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p\end{array}$$

$$\max_{\lambda \geq 0, v} L(x, \lambda, v) = \begin{cases} f_0(x), & \text{if } f_i(x) \leq 0, h_i(x) = 0 \\ +\infty, & \text{otherwise} \end{cases}$$

Weak duality

Primal problem

$$\begin{aligned} p^* = \min & f_0(x) \\ \text{s.t. } & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

Dual problem

$$\begin{aligned} d^* = \max & g(\lambda, v) \\ \text{s.t. } & \lambda \geq 0 \end{aligned}$$

- p^* : optimal value of primal problem; d^* : optimal value of dual problem

Lower bound property: for any $\lambda \geq 0$ and any v , we have

$$g(\lambda, v) \leq p^*$$



Weak duality: $d^* \leq p^*$

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

Duality gap: $p^* - d^*$

Weak duality

Primal problem

$$\begin{aligned} p^* = \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

Dual problem

$$\begin{aligned} d^* = \max \quad & g(\lambda, v) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

- p^* : optimal value of primal problem; d^* : optimal value of dual problem

Weak duality: $d^* \leq p^*$

- $p^* = -\infty \implies d^* = -\infty$ (If the primal problem is **unbounded below**, dual problem is **infeasible**)
- $d^* = \infty \implies p^* = \infty$ (If the dual problem is **unbounded above**, primal problem is **infeasible**)

Strong duality

Primal problem

$$\begin{aligned} p^* = \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

Dual problem

$$\begin{aligned} d^* = \max \quad & g(\lambda, v) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

Strong duality: $d^* = p^*$

- The best bound obtained from dual function is tight.
- **Does not hold** in general
- Sufficient conditions for strong duality are called **constraint qualifications**
- Strong duality **usually** holds for convex optimization

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & a_i^T x = b_i, i = 1, \dots, p \end{aligned}$$

Slater's condition

One simple constraint qualification: convex optimization problem + *Slater's condition*

Convex optimization problem

$$\begin{array}{ll}\min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \dots, m \\ & Ax = b\end{array}$$

Slater's condition: there exists an $x' \in \text{int } \mathcal{D}$ such that $f_i(x') < 0, i = 1, \dots, m, Ax = b$.

 strictly feasible point

Slater's condition (weak form): if some inequality constraint functions are affine, i.e., f_1, \dots, f_k are affine: there exists an x' such that

$$f_i(x') \leq 0, i = 1, \dots, k, f_i(x') < 0, i = k + 1, \dots, m, Ax = b.$$

If the problem is a **convex optimization problem** and **Slater's condition** holds, then strong duality holds.

Complementary slackness

What can we learn from strong duality?

Suppose strong duality holds. Let x^* and (λ^*, ν^*) be primal and dual optimal, respectively.

$$g(\lambda^*, \nu^*) = f_0(x^*)$$

$$g(\lambda^*, \nu^*) = \inf_x f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x)$$

$$\begin{aligned} &\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$



Equality holds: $\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$

Complementary slackness: $\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$

$$\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0$$


$$f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$$

Lagrangian optimality

What can we learn from strong duality?

Suppose strong duality holds. Let x^* and (λ^*, v^*) be primal and dual optimal, respectively.

$$g(\lambda^*, v^*) = f_0(x^*)$$

$$\begin{aligned} g(\lambda^*, v^*) &= \inf_x f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p v_i^* h_i(x) \\ &\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p v_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$


Equality holds: $\inf_x L(x, \lambda^*, v^*) = L(x^*, \lambda^*, v^*)$

Lagrangian optimality: $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p v_i^* \nabla h_i(x^*) = 0$

Karush-Kuhn-Tucker (KKT) conditions

$$\begin{array}{ll}\min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p\end{array}$$

Suppose f_i and h_i are differentiable.

KKT conditions

- **Complementary slackness:** $\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$
- **Lagrangian optimality:** $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p v_i^* \nabla h_i(x^*) = 0$
- **Primal feasibility:** $f_i(x^*) \leq 0, i = 1, \dots, m, h_i(x^*) = 0, i = 1, \dots, p$
- **Dual feasibility:** $\lambda_i^* \geq 0, i = 1, \dots, m$

Strong duality holds. x^* and (λ^*, v^*) are primal and dual optimal, respectively.



x^* and (λ^*, v^*) satisfy KKT conditions.

KKT conditions are **necessary conditions** for strong duality and optimality

Karush-Kuhn-Tucker (KKT) conditions

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

Suppose f_i and h_i are differentiable, and the problem is a **convex optimization problem**.

KKT conditions

- **Complementary slackness:** $\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$
- **Lagrangian optimality:** $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p v_i^* \nabla h_i(x^*) = 0$
- **Primal feasibility:** $f_i(x^*) \leq 0, i = 1, \dots, m, h_i(x^*) = 0, i = 1, \dots, p$
- **Dual feasibility:** $\lambda_i^* \geq 0, i = 1, \dots, m$

x^* and (λ^*, v^*) satisfy KKT conditions.



Strong duality holds. x^* and (λ^*, v^*) are primal and dual optimal, respectively.

KKT conditions are **sufficient conditions** for strong duality and optimality of a **convex optimization problem**.

Karush-Kuhn-Tucker (KKT) conditions

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

Suppose f_i and h_i are differentiable, the problem is a **convex optimization problem** and satisfies **Slater's conditions**.

KKT conditions

- **Complementary slackness:** $\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$
- **Lagrangian optimality:** $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p v_i^* \nabla h_i(x^*) = 0$
- **Primal feasibility:** $f_i(x^*) \leq 0, i = 1, \dots, m, h_i(x^*) = 0, i = 1, \dots, p$
- **Dual feasibility:** $\lambda_i^* \geq 0, i = 1, \dots, m$

x^* and (λ^*, v^*) satisfy KKT conditions.



x^* and (λ^*, v^*) are primal and dual optimal, respectively.

For a **convex** optimization that satisfies **Slater's conditions**, KKT conditions are **sufficient and necessary conditions** for strong duality and optimality.

Example

Water-filling.

- To allocate power to a set of n communication channels to maximize total communication rate.
- $\log(\alpha_i + x_i)$ is the communication rates of channel i under power x_i and context-related parameter α_i


$$\begin{aligned} \min \quad & -\sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{s.t.} \quad & x \geq 0, \\ & \sum_{i=1}^n x_i = 1 \end{aligned}$$

KKT conditions

- **Complementary slackness:** $\lambda_i^* x_i^* = 0, i = 1, \dots, n$
- **Lagrangian optimality:** $-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0, i = 1, \dots, n \quad (1)$
- **Primal feasibility:** $x_i^* \geq 0, i = 1, \dots, n, \sum_{i=1}^n x_i^* = 1$
- **Dual feasibility:** $\lambda_i^* \geq 0, i = 1, \dots, m$

➤ If $x_i^* > 0$, then $\lambda_i^* = 0$. $\lambda_i^* = 0$ and (1) give $x_i^* = 1/\nu^* - \alpha_i$

➤ If $\lambda_i^* > 0$, then $x_i^* = 0$. $\lambda_i^* > 0$ and (1) give $\nu^* \geq 1/\alpha_i$



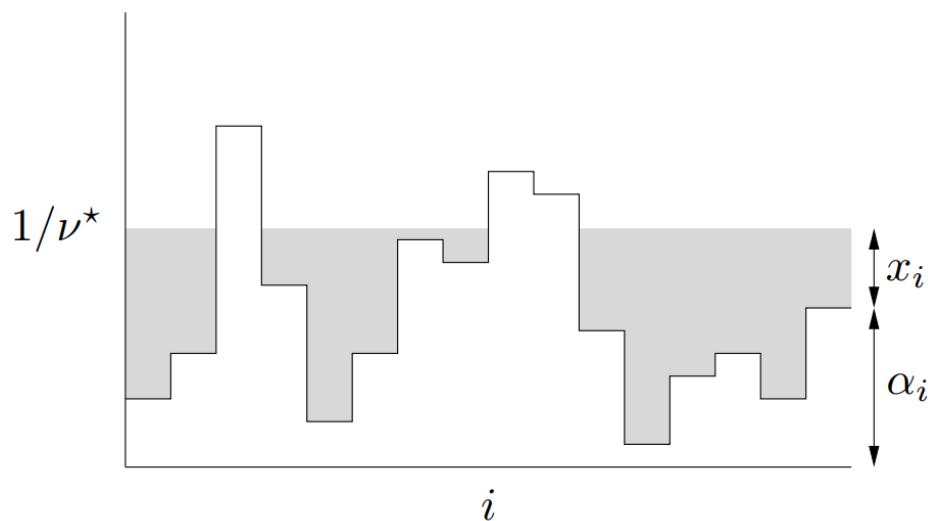
$$x_i^* = \begin{cases} \frac{1}{\nu^*} - \alpha_i, & \nu^* < \frac{1}{\alpha_i} \\ 0, & \nu^* \geq \frac{1}{\alpha_i} \end{cases}$$

Example

$$\begin{aligned} \min \quad & -\sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{s.t.} \quad & x \geq 0, \\ & \sum_{i=1}^n x_i = 1 \end{aligned}$$

Solution
$$x_i^* = \begin{cases} 1/\nu^* - \alpha_i, & \nu^* < 1/\alpha_i \\ 0, & \nu^* \geq 1/\alpha_i \end{cases} \quad \text{or} \quad x_i^* = \max\{0, 1/\nu^* - \alpha_i\},$$

where ν^* is such that $\sum_{i=1}^n \max\{0, 1/\nu^* - \alpha_i\} = 1$



Saddle point

Primal problem

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

Dual problem

$$\begin{aligned} \max & g(\lambda, v) \\ \text{s.t.} & \lambda \geq 0 \end{aligned}$$



$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p v_i h_i(x)$$



Primal problem:

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

Dual problem:

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

(x', λ', v') where $\lambda' \geq 0$ is a **saddle point** of the Lagrangian function L if

$$L(x', \lambda, v) \leq L(x', \lambda', v') \leq L(x, \lambda', v')$$

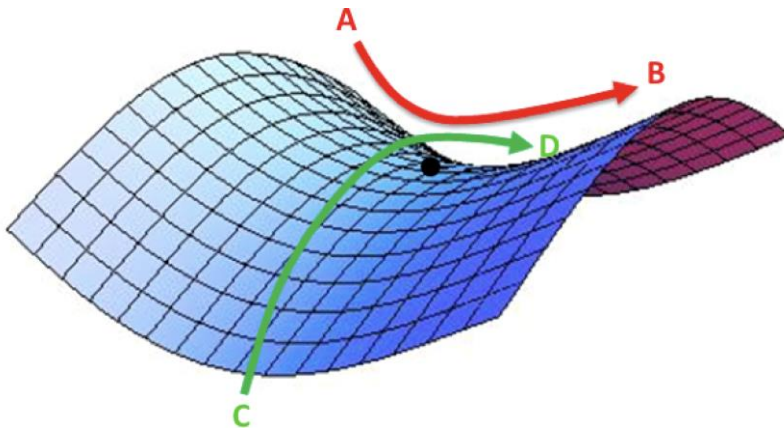
- x' minimizes L when (λ, v) is fixed at (λ', v') , i.e., $L(x', \lambda', v') = \min_x L(x, \lambda', v')$.
- (λ', v') maximize L when x is fixed at x' , i.e., $L(x', \lambda', v') = \max_{\lambda \geq 0, v} L(x', \lambda, v)$

Saddle point

(x', λ', ν') where $\lambda' \geq 0$ is a **saddle point** of the Lagrangian function L if

$$L(x', \lambda, \nu) \leq L(x', \lambda', \nu') \leq L(x, \lambda', \nu')$$

- x' minimizes L when (λ, ν) is fixed at (λ', ν') , i.e., $L(x', \lambda', \nu') = \min_x L(x, \lambda', \nu')$.
- (λ', ν') maximize L when x is fixed at x' , i.e., $L(x', \lambda', \nu') = \max_{\lambda \geq 0, \nu} L(x', \lambda, \nu)$



(x', λ', ν') is a saddle point of L **if and only if** x' and (λ', ν') are primal and dual optimal, respectively, and strong duality holds.

Shadow price Interpretation

$$\begin{array}{ll}
 \max & c^T x \\
 \text{subject to} & Ax \leq b \\
 & x \geq 0
 \end{array}
 \xrightarrow{\text{dual}}
 \begin{array}{ll}
 \min & b^T \lambda \\
 \text{s.t.} & A^T \lambda \geq c \\
 & \lambda \geq 0
 \end{array}$$

Production planning (primal problem).

- To determine the quantities of n products to maximize total profit s.t. resource constraints.
- c : profit; A : resource consumption; b : available resource.

Lagrange: $L(x, \lambda, \alpha) = -c^T x + \lambda^T (Ax - b) - \alpha^T x$

Shadow price for each resource:
Have unused resource: sell
Need more resource: buy

$$\begin{aligned}
 g(\lambda) &= \min_x L(x, \lambda, \alpha) = -b^T \lambda + \min_x x^T (A^T \lambda - c - \alpha) \\
 &= \begin{cases} -b^T \lambda, & A^T \lambda - c - \alpha = 0 \\ -\infty, & \text{otherwise} \end{cases}
 \end{aligned}$$

$$\begin{array}{ll}
 \max & g(\lambda, \nu) = \begin{cases} -b^T \lambda, & A^T \lambda - c - \alpha = 0 \\ -\infty, & \text{otherwise} \end{cases} \\
 \text{s.t.} & \lambda \geq 0, \alpha \geq 0
 \end{array}$$

Weak duality:
profit of product \leq value of resources

Resource purchase (dual problem).

- To determine the prices of m resources to minimize total cost.
- Constraint: for each product, payment of selling the resources \geq profit of selling the product