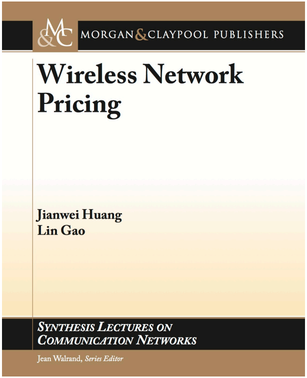
Wireless Network Pricing
Chapter 4: Social Optimal Pricing

Jianwei Huang & Lin Gao

Network Communications and Economics Lab (NCEL)
Information Engineering Department
The Chinese University of Hong Kong

The Book



- E-Book freely downloadable from NCEL website: `http://ncel.ie.cuhk.edu.hk/content/wireless-network-pricing`
- Physical book available for purchase from Morgan & Claypool (`http://goo.gl/JFGlai`) and Amazon (`http://goo.gl/JQKaEq`)

Chapter 4: Social Optimal Pricing

Outline

- Focus: social optimal pricing, where one service provider chooses prices to maximize the social welfare.
- Theory: Convex Optimization

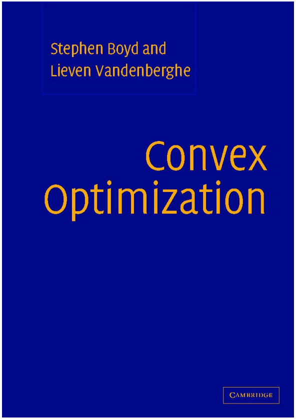
Convex Optimization

Definition (Convex Optimization)

Convex optimization studies the problem of minimizing convex functions (or equivalently, maximizing concave functions) over convex sets.

Convex Optimization

- Mainly follow the book "Convex Optimization" by Boyd and Vandenberghe.



Section 4.1
Theory: Dual-based Optimization

Prelims

- Notations
  - $\mathbb{R}^n$: the set of all real n-vectors
    - Each vector in $\mathbb{R}^n$ is called a *point* of $\mathbb{R}^n$.
    - $\mathbb{R}^2$ is the set of all 2-D vectors
    - $\mathbb{R}^1$ or $\mathbb{R}$ denotes the set of all real 1-vectors (all real numbers).
  - $\mathbb{R}^{m \times n}$: the set of all $m \times n$ real matrices
  - $f : \mathbb{R}^n \to \mathbb{R}^m$: a function that maps some real n-vectors (called the *domain* of function f) into real m-vectors
    - $\mathcal{D}(f)$: the domain of function f
- Key Concepts
  - Convex Set
  - Convex Function
  - Convex Optimization
  - Dual Algorithm

Convex Set

Convex Set

Definition (Convex Set)

A nonempty set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex, if for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ and any $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$, we have:

$$\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in \mathcal{X}$$

Convex Set

- Geometrically, a set is convex if every point in the set can be reached by every other point, along an inner straight path between them.
- Examples of convex and non-convex sets:

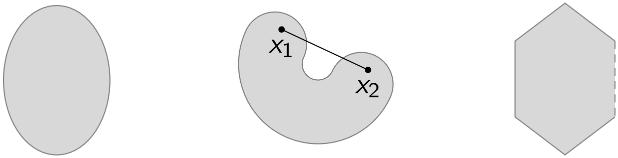


Figure: (i) Convex (ii) Non-convex (iii) Non-convex

Convex Combination

Definition (Convex Combination)

A convex combination of points $\boldsymbol{x}_1, ..., \boldsymbol{x}_k$ can be expressed as

$$\boldsymbol{y} = \theta_1 \boldsymbol{x}_1 + ... + \theta_k \boldsymbol{x}_k,$$

with $\theta_1 + ... + \theta_k = 1$ and $\theta_i \geq 0, i = 1, ..., k$.

Lemma (4.2)

A nonempty set $\mathcal{X}$ is convex, if and only if the convex combination of any points in $\mathcal{X}$ also lies in $\mathcal{X}$.

Convex Hull

- The convex hull of a set $\mathcal{X}$, denoted $\mathcal{H}(\mathcal{X})$, is the smallest convex set that contains $\mathcal{X}$.

Definition (Convex Hull)

The convex hull $\mathcal{H}(\mathcal{X})$ of a set $\mathcal{X}$ consists of the convex combinations of all points in $\mathcal{X}$, i.e.,

$$\{\theta_1 \boldsymbol{x}_1 + \ldots + \theta_k \boldsymbol{x}_k \mid \theta_1 + \ldots + \theta_k = 1, \theta_i \geq 0, \boldsymbol{x}_i \in \mathcal{X}, i = 1, \ldots, k\}.$$

- Properties
  - $\mathcal{H}(\mathcal{X})$ is always convex (even if $\mathcal{X}$ is not convex);
  - $\mathcal{X} \subseteq \mathcal{H}(\mathcal{X})$;
  - If $\mathcal{X}$ is a convex set, then $\mathcal{X} = \mathcal{H}(\mathcal{X})$;
  - If $\mathcal{Y}$ is a convex set that contains $\mathcal{X}$, then $\mathcal{H}(\mathcal{X}) \subseteq \mathcal{Y}$.

Examples of Convex Hull

- Source sets $\mathcal{X}$:

- Convex hulls $\mathcal{H}(\mathcal{X})$:

Operations Preserving the Convexity of Sets

- Intersection: Suppose $\mathcal{X}_1, ..., \mathcal{X}_k$ are convex sets. Then, the intersection of $\mathcal{X}_1, ..., \mathcal{X}_k$

$$\mathcal{X} \triangleq \mathcal{X}_1 \cap ... \cap \mathcal{X}_k$$

  is also a convex set.

- Affine Mapping: Suppose $\mathcal{X}$ is a convex set in $\mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, and $\boldsymbol{b} \in \mathbb{R}^m$. Then, the affine mapping of $\mathcal{X}$

$$\mathcal{Y} \triangleq \{\mathbf{A}\boldsymbol{x} + \boldsymbol{b} \mid \boldsymbol{x} \in \mathcal{X}\}$$

  is also a convex set.

Convex Function

Convex (and Concave) Function

Definition (Convex Function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, if

1. $\mathcal{D}(f)$ is a convex set, and
2. for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}(f)$ and $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$, we have:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}).$$

Convex (and Concave) Function

Definition (Convex Function)

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, if

1. $\mathcal{D}(f)$ is a convex set, and
2. for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}(f)$ and $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$, we have:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}).$$

Definition (Concave Function)

A function $f(\cdot)$ is concave if and only if $-f(\cdot)$ is convex.

Convex (and Concave) Function

Definition (Convex Function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, if

1. $\mathcal{D}(f)$ is a convex set, and
2. for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}(f)$ and $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$, we have:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}).$$

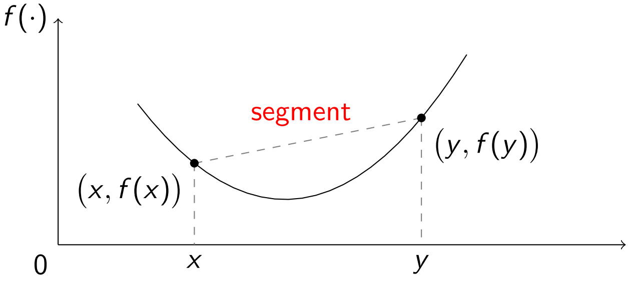
Definition (Concave Function)

A function $f(\cdot)$ is concave if and only if $-f(\cdot)$ is convex.

- A function $f(\cdot)$ can be neither convex nor concave, e.g., $f(x) = x^3$ over $\mathcal{D}(f) = \mathbb{R}$.

Convex Function

- Assume that $\mathcal{D}(f)$ is convex.
- Geometrically, a function $f(\cdot)$ is convex if the segment between any two points $(\boldsymbol{x}, f(\boldsymbol{x}))$ and $(\boldsymbol{y}, f(\boldsymbol{y}))$ lies above $f(\cdot)$.
- Illustration of Convex Function $f(\cdot)$:



Strictly Convex (and Concave) Function

Definition (Strictly Convex Function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex, if

1. $\mathcal{D}(f)$ is a convex set, and
2. for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}(f)$ and $\theta \in \mathbb{R}$ with $0 < \theta < 1$, we have:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) < \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

Strictly Convex (and Concave) Function

Definition (Strictly Convex Function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex, if

1. $\mathcal{D}(f)$ is a convex set, and
2. for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}(f)$ and $\theta \in \mathbb{R}$ with $0 < \theta < 1$, we have:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) < \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

Definition (Strictly Concave Function)

A function $f(\cdot)$ is strictly concave if and only if $-f(\cdot)$ is strictly convex.

Generalized Definition of Convex Function

Definition (Convex Function)

A function $f(\cdot)$ is convex, if and only if

1. $\mathcal{D}(f)$ is convex, and
2. For any $\boldsymbol{x}_1, ..., \boldsymbol{x}_k \in \mathcal{D}(f)$,

$$f(\theta_1 \boldsymbol{x}_1 + ... + \theta_k \boldsymbol{x}_k) \leq \theta_1 f(\boldsymbol{x}_1) + ... + \theta_k f(\boldsymbol{x}_k),$$

when $\theta_1 + ... + \theta_k = 1$ and $\theta_i \geq 0, i = 1, ..., k$.

Examples of Convex Functions

- Examples of convex functions over $\mathcal{D}(f) = (0, \infty)$
  - 2^x, 3^x, e^x, etc.
  - x^2, x^4, x^6, etc.
  - $-log_2(x)$, $-ln(x)$, etc.
- Question: What about
  - $f(x) = x^3$ over $\mathcal{D}(f) = [0, \infty)$?
  - $f(x) = x^3$ over $\mathcal{D}(f) = (-\infty, 0]$?
  - $f(x) = 2x$ over $\mathcal{D}(f) = \mathbb{R}$?
  - x^2 over $\mathcal{D}(f) = \mathbb{R}$?
  - x^2 over $\mathcal{D}(f) = (-\infty, 0) \cup [1, \infty)$?
  - $-log_2(x)$ over $\mathcal{D}(f) = \mathbb{R}$?

Gradient

- Consider a scalar-valued function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$

Definition (First-Order Derivative (Gradient))

The first-order derivative of a scalar-valued function $f(\cdot)$ at a point $\boldsymbol{x} \in \mathcal{D}(f)$, denoted by $\nabla f(\boldsymbol{x})$, is an n-dimensional vector with the i-th component given by

$$\nabla f(\boldsymbol{x})_i = \frac{\partial f(\boldsymbol{x})}{\partial x_i}, \ i = 1, ..., n,$$

- x_i: the i-th coordinate of the vector $\boldsymbol{x}$;
- $\partial f(\boldsymbol{x})/\partial x_i$: the partial derivative of $f(\boldsymbol{x})$ with respect to x_i.

First-Order Condition of a Convex Function

Lemma (First-Order Condition of a Convex Function)

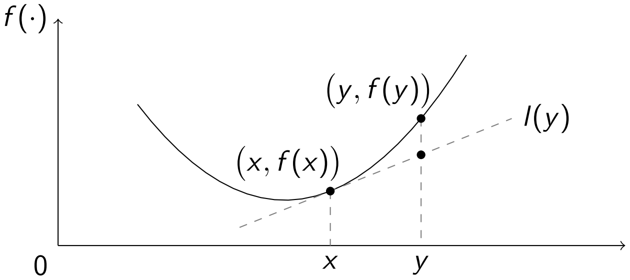
A differentiable function $f(\cdot)$ is convex, if and only if $\mathcal{D}(f)$ is convex and

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}(f).$$

First-Order Condition of a Convex Function

- Geometrically, the line passing through any point $(\boldsymbol{x}, f(\boldsymbol{x}))$ along the gradient direction $\nabla f(\boldsymbol{x})$ lies under the graph of $f(\cdot)$.

- Illustration of First-order Condition:



- Example: $f(x) = (x-3)^2$, we have $f(8) \geq f(5) + f'(5)(8-5)$

Hessian Matrix

Definition (Second-Order Derivative (Hessian Matrix))

The second-order derivative of a scalar-valued function $f(\cdot)$ at a point $\boldsymbol{x} \in \mathcal{D}(f)$, denoted by $\nabla^2 f(\boldsymbol{x})$, is an $n \times n$ matrix, given by

$$\nabla^2 f(\boldsymbol{x})_{ij} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}, \ i = 1, ..., n, j = 1, ..., n.$$

- $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$: the second partial derivative of $f(\boldsymbol{x})$ with respect to x_i and x_j.

- Example: Calculate the Hessian Matrix of $f(\boldsymbol{x}) = x_1^2 + 3x_2^2 + 5x_1x_2$

Second-Order Condition of a Convex Function

Lemma (Second-Order Condition)

A twice differentiable function $f(\cdot)$ is convex, if and only if $\mathcal{D}(f)$ is convex and its Hessian matrix is positive semidefinite, i.e.,

$$\nabla^2 f(\boldsymbol{x}) \succeq 0, \quad \forall \boldsymbol{x} \in \mathcal{D}(f).$$

- Positive semidefinite matrix: `https://en.wikipedia.org/wiki/Positive-definite_matrix`

Second-Order Condition of a Convex Function

Lemma (Second-Order Condition)

A twice differentiable function $f(\cdot)$ is convex, if and only if $\mathcal{D}(f)$ is convex and its Hessian matrix is positive semidefinite, i.e.,

$$\nabla^2 f(\boldsymbol{x}) \succeq 0, \quad \forall \boldsymbol{x} \in \mathcal{D}(f).$$

- Positive semidefinite matrix: `https://en.wikipedia.org/wiki/Positive-definite_matrix`
- Question: what about the special case of $f(\boldsymbol{x}) : \mathbb{R} \to \mathbb{R}$?

Operations Preserving Convexity of Functions

- Nonnegative weighted sums: Suppose $f_1(\cdot), ..., f_k(\cdot)$ are convex, and $\theta_1, ..., \theta_k \geq 0$. Then the following function is convex:

$$f(\boldsymbol{x}) \triangleq \theta_1 f_1(\boldsymbol{x}) + ... + \theta_k f_k(\boldsymbol{x})$$

- Point-wise maximum: Suppose $f_1(\cdot), ..., f_k(\cdot)$ are convex. Then the following function is convex:

$$f(\boldsymbol{x}) \triangleq \max\{f_1(\boldsymbol{x}), ..., f_k(\boldsymbol{x})\}$$

  - An illustration

- Composition with an affine mapping: Suppose $g(\cdot)$ is a convex function on $\mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n\times m}$, and $\boldsymbol{b} \in \mathbb{R}^n$. Then the following function is convex:

$$f(\boldsymbol{x}) \triangleq g(\mathbf{A}\boldsymbol{x} + \boldsymbol{b})$$

Operations Preserving Convexity of Functions

- Nonnegative weighted sums: Suppose $f_1(\cdot), ..., f_k(\cdot)$ are convex, and $\theta_1, ..., \theta_k \geq 0$. Then the following function is convex:

$$f(\boldsymbol{x}) \triangleq \theta_1 f_1(\boldsymbol{x}) + ... + \theta_k f_k(\boldsymbol{x})$$

- Point-wise maximum: Suppose $f_1(\cdot), ..., f_k(\cdot)$ are convex. Then the following function is convex:

$$f(\boldsymbol{x}) \triangleq \max\{f_1(\boldsymbol{x}), ..., f_k(\boldsymbol{x})\}$$

  - An illustration

- Composition with an affine mapping: Suppose $g(\cdot)$ is a convex function on $\mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n \times m}$, and $\boldsymbol{b} \in \mathbb{R}^n$. Then the following function is convex:

$$f(\boldsymbol{x}) \triangleq g(\mathbf{A}\boldsymbol{x} + \boldsymbol{b})$$

- Question: What is the relationship between $\mathcal{D}(f)$ and the domain(s) of the old function(s)?

Convex Optimization

Convex Optimization

- Optimization Problem: the problem of finding a point $\boldsymbol{x}$ over a feasible set that minimizes an objective function:

Optimization Problem

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \ \ i = 1, ..., m. \end{aligned}$$

  - Objective function $f(\cdot)$: the objective to be minimized;
  - Constraint functions $f_i(\cdot)$: the constraints to be satisfied;
  - Feasible set $\mathcal{C}$: the set of all feasible points that satisfy all constraints,

$$\mathcal{C} \triangleq \{\boldsymbol{x} \in \mathcal{D} \mid f_i(\boldsymbol{x}) \leq 0, i = 1, ..., m\}.$$

    - ★ Here $\mathcal{D} = \mathcal{D}(f) \cap \mathcal{D}(f_1) \cap \cdots \cap \mathcal{D}(f_m)$.

- Convex Optimization Problem: an optimization problem with a convex objective function and a convex feasible set.

Unconstrained Convex Optimization

- Unconstrained Convex Optimization: a convex optimization problem without any constraint:

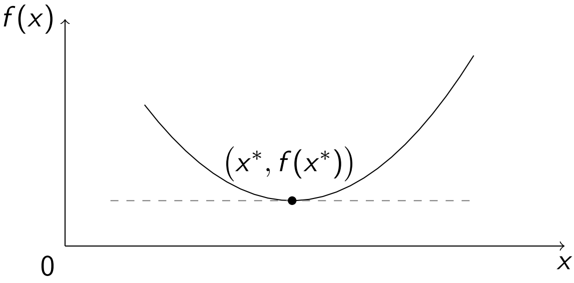
$$\text{minimize} \quad f(\boldsymbol{x})$$

Lemma (4.5)

Suppose $f(\cdot)$ is convex and differentiable. A feasible point $\boldsymbol{x}^* \in \mathcal{C}$ is a global minimizer of $f(\cdot)$ if and only if

$$\nabla f(\boldsymbol{x}^*)_i = \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} = 0, \quad \forall i = 1, ..., n.$$

Unconstrained Convex Optimization



Solving Unconstrained Convex Optimization

- Computational Methods: find an algorithm that computes a sequence of feasible points $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, ...\boldsymbol{x}^{(k)}$, with

$$f(\boldsymbol{x}^{(k)}) \to f(\boldsymbol{x}^{*}) \text{ as } k \to \infty$$

- Gradient-based Algorithms:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \gamma^{(k)} \boldsymbol{d}^{(k)}$$

  - $\boldsymbol{d}^{(k)}$: a gradient-based n-vector (called search direction) at iteration k;
    - ★ Example: $\boldsymbol{d}^{(k)} \triangleq -\nabla f(\boldsymbol{x}^{(k)})$ (Gradient Descent Method)
  - $\gamma^{(k)}$: a positive scalar (called step size) at iteration k;
    - ★ Question: What happen is the step size is too large or too small?

Constrained Convex Optimization

- Constrained Convex Optimization: a convex optimization problem with convex constraints (i.e., $f_i(\cdot)$ function is convex for each i):

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \; i = 1, ..., m, \end{aligned}$$

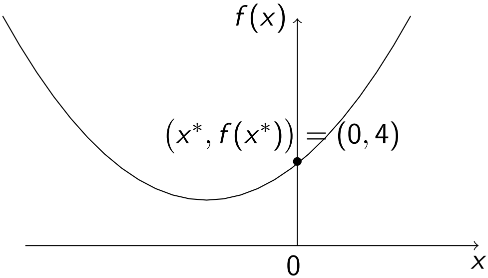
Lemma (4.6)

Suppose $f(\cdot)$ is convex and differentiable. A feasible point $\boldsymbol{x}^* \in \mathcal{C}$ is a global minimizer of $f(\cdot)$ if and only if

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \forall \boldsymbol{x} \in \mathcal{C}.$$

A Numerical Example

$$\begin{aligned} \text{minimize} \quad & (x+2)^2 \\ \text{subject to} \quad & x \geq 0. \end{aligned}$$



- $\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = 2(x^* + 2)(x - x^*) = 2(0+2) \cdot (x - 0) \geq 0$

Constrained Convex Optimization

- Geometrically, at a minimizer $\boldsymbol{x}^*$, the gradient $\nabla f(\boldsymbol{x}^*)$ makes an angle less than or equal to 90 degrees with all feasible variations $\boldsymbol{x} - \boldsymbol{x}^*$.
- Illustration of optimal $\boldsymbol{x}^*$:

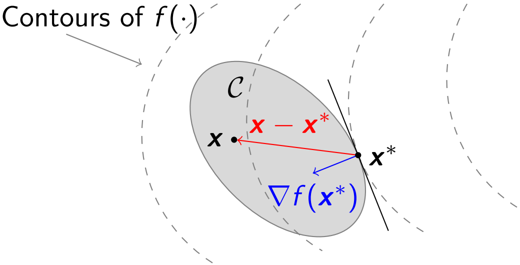


Figure: The gradient $\nabla f(\boldsymbol{x}^*)$ (blue arrow) makes an angle less than or equal to 90 degrees with all feasible variations $\boldsymbol{x} - \boldsymbol{x}^*$ (red arrow).

Dual Algorithm

Duality Principle

- An important theoretical framework to solve convex optimization problems.
- Basic Idea: Convert the original optimization problem (called primal problem) into a dual problem.
  - The solution to the dual problem provides a lower bound to the solution of the primal problem.
  - Maximizing the objective of dual problem help us understand the optimal objective of the primal problem.

Lagrange Function

- Recall: the constrained optimization problem (Primal Problem)

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \; i = 1, ..., m, \end{aligned}$$

Lagrange Function

- Recall: the constrained optimization problem (Primal Problem)

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \; i = 1, ..., m, \end{aligned}$$

Definition (Lagrangian Function)

The Lagrangian function $L(\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is defined as

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) \triangleq f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}),$$

where $\boldsymbol{x}$ is of dimension n and $\boldsymbol{\lambda}$ is of dimension m.

- Intuitively, Lagrangian function is a weighted sum of the objective function $f(\boldsymbol{x})$ and the constraint functions $f_i(\boldsymbol{x})$.
- $\lambda_i \geq 0$: the weight (called Lagrange multiplier or dual variable) associated with each constraint $f_i(\boldsymbol{x}) \leq 0$.

Dual Function

Definition (Dual Function)

The Lagrange dual function $g : \mathbb{R}^m \to \mathbb{R}$ is defined as the minimum value of the Lagrangian function over $\boldsymbol{x}$:

$$g(\boldsymbol{\lambda}) \triangleq \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} \left(f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) \right).$$

Dual Function

Definition (Dual Function)

The Lagrange dual function $g : \mathbb{R}^m \to \mathbb{R}$ is defined as the minimum value of the Lagrangian function over $\boldsymbol{x}$:

$$g(\boldsymbol{\lambda}) \triangleq \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} \left(f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) \right).$$

- The dual function $g(\boldsymbol{\lambda})$ is always concave even if the primal problem is not convex.
- The dual function $g(\boldsymbol{\lambda})$ yields a lower bound of the optimal primal objective value $f(\boldsymbol{x}^*)$:

$$g(\boldsymbol{\lambda}) \leq f(\boldsymbol{x}^*), \quad \forall \boldsymbol{\lambda} \succeq \mathbf{0}$$

  - Proof: for any feasible solution $\tilde{\boldsymbol{x}}$ of the Primal Problem, we have

$$g(\boldsymbol{\lambda}) \triangleq \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \leq L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}) = f(\tilde{\boldsymbol{x}}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{\boldsymbol{x}}) \leq f(\tilde{\boldsymbol{x}})$$

Lagrange Dual Problem

- How far is the dual function $g(\boldsymbol{\lambda})$ away from the optimal $f(\boldsymbol{x}^*)$?

Definition (Lagrange Dual Problem)

Find the optimal dual variables $\boldsymbol{\lambda}^*$ that maximizes the dual function $g(\boldsymbol{\lambda})$:

$$\begin{aligned} \text{maximize} \quad & g(\boldsymbol{\lambda}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \succeq \mathbf{0}. \end{aligned}$$

Duality Gap

Definition (Duality Gap)

The gap between primal and dual objectives: $f(\boldsymbol{x}) - g(\boldsymbol{\lambda})$.

- The duality gap reflects how suboptimal a given point $\boldsymbol{x}$ is, without knowing the exact value of $f(\boldsymbol{x}^*)$:

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) - g(\boldsymbol{\lambda})$$

- Any primal-dual feasible pair $\{\boldsymbol{x}, \boldsymbol{\lambda}\}$ localizes the optimal primal and dual objectives to an interval $[g(\boldsymbol{\lambda}), f(\boldsymbol{x})]$, that is,

$$g(\boldsymbol{\lambda}) \leq g(\boldsymbol{\lambda}^*) \leq f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$$

- Weak duality (always true): $g(\boldsymbol{\lambda}^*) \leq f(\boldsymbol{x}^*)$. The difference $f(\boldsymbol{x}^*) - g(\boldsymbol{\lambda}^*)$ is called the optimal duality gap.
- Strong duality (sometimes is true): $g(\boldsymbol{\lambda}^*) = f(\boldsymbol{x}^*)$ if the optimality gap is zero.

KKT Optimality Conditions

Lemma (Karush-Kuhn-Tucker (KKT) Conditions)

Assume that the primal problem is strictly convex and the strong duality holds. A primal-dual feasible pair $\{\boldsymbol{x}^*, \boldsymbol{\lambda}^*\}$ *is optimal for both primal and dual problems, if and only if*

$$\begin{cases} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\boldsymbol{x}^*) = \boldsymbol{0} & \textit{(Stationarity)} \\ \lambda_i^* \cdot f_i(\boldsymbol{x}^*) = 0, \quad i = 1, ..., m & \textit{(Complementary Slackness)} \\ f_i(\boldsymbol{x}^*) \leq 0,\ \lambda_i^* \geq 0, \quad i = 1, ..., m & \textit{(Feasibility)} \end{cases}$$

Shadow Price

- Shadow Price: An interpretation of the Lagrange multipliers λ_i, $i = 1, ..., m$, in terms of economics.
  - Introduce perturbing parameters $\boldsymbol{u} \triangleq (u_i, i = 1, ..., m)$, and define a perturbed version of the original primal problem:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq u_i, \;\; i = 1, ..., m \end{aligned}$$

  - Denote the optimal perturbed objective as $p^*(\boldsymbol{u}) = \inf_{\boldsymbol{x}} f(\boldsymbol{x})$:

$$\left. \frac{\partial p^*(\boldsymbol{u})}{\partial u_i} \right|_{\boldsymbol{u}=\mathbf{0}} = -\lambda_i^*$$

    - ★ $f(\boldsymbol{x})$: total cost of the firm;
    - ★ $-f(\boldsymbol{x})$: total profit of the firm;
    - ★ u_i: the limit on resource i's investment;
    - ★ When $\boldsymbol{u}$ equals to $\mathbf{0}$, the λ_i^* reflects how much more profit the firm could make, for a small increase in the availability of resource i.

Subgradient

- Subgradient: A vector $\boldsymbol{d}$ is called a subgradient of a convex $f(\cdot)$ at a point $\boldsymbol{x}$, if

$$f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + \boldsymbol{d}^T(\boldsymbol{z} - \boldsymbol{x}), \quad \forall \boldsymbol{z} \in \mathcal{D}(f).$$

  - Generalization of the gradient for the non-differentiable functions

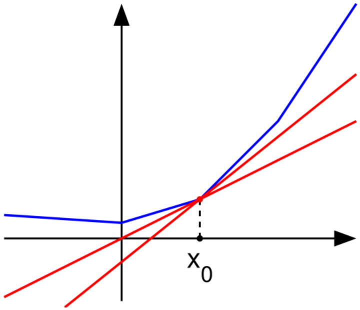


Figure: A convex function (blue) and subgradients at x_0 (red) ©Wikipedia

Solving Dual Problem

- Subgradient method for solving the due problem
  - A subgradient $\boldsymbol{d}$ of the concave dual function $g(\boldsymbol{\lambda})$ at a point $\boldsymbol{\lambda}$ satisfies:
  $$g(\boldsymbol{\mu}) \leq g(\boldsymbol{\lambda}) + \boldsymbol{d}^T(\boldsymbol{\mu} - \boldsymbol{\lambda}), \quad \forall \boldsymbol{\mu} \in \mathcal{D}(g).$$
  - Subgradient Method for updating the values of $\boldsymbol{\lambda}$:
  $$\boldsymbol{\lambda}^{(k+1)} = \left[\boldsymbol{\lambda}^{(k)} + \gamma^{(k)} \boldsymbol{d}^{(k)}\right]^+$$

Solving Dual Problem

Lemma (4.10)

For every dual optimal solution $\boldsymbol{\lambda}^*$*, we have* $||\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^*|| < ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^*||$ *for all step-sizes* $\gamma^{(k)}$ *satisfying*

$$0 < \gamma^{(k)} < 2 \cdot \frac{g(\boldsymbol{\lambda}^*) - g(\boldsymbol{\lambda}^{(k)})}{||\boldsymbol{d}^{(k)}||^2}.$$

- The above range for $\gamma^{(k)}$ requires the dual optimal value $g(\boldsymbol{\lambda}^*)$, which is usually unknown.
- In practice, we can use the following approximate step-size formula

$$\gamma^{(k)} = \alpha^{(k)} \cdot \frac{g^{(k)} - g(\boldsymbol{\lambda}^{(k)})}{||\boldsymbol{d}^{(k)}||^2},$$

where $g^{(k)}$ is an approximation of $g(\boldsymbol{\lambda}^*)$, and $0 < \alpha^{(k)} < 2$.