Chapter 2

Examples of Applications

The purpose of this chapter is to provide some examples of optimization where the previously mentioned basic principles can be applied. Each of the following subjects has been a major mathematical study. Only the ideas are outlined in this chapter. More details will be discussed later.

- Nonlinear Programming
- Linear Programming
- Calculus of Variation
- Optimal Control

Nonlinear Programming

A main tool used in nonlinear programming is the theory of Lagrange multipliers.

- The geometry of Lagrange multipliers is easy to understand in finite dimensional space, but the entire notion can be generalized to Banach space.
- An equality constrained optimization problem can be reformulate via the notion of Lagrange multipliers as an unconstrained problem. As such, necessary and sufficient conditions for constrained problems are similar to those of unconstrained problems.
- An inequality constrained problem can be reformulated as an equality constrained problem by adding extra variables or inequalities. As such, the idea of *active* constraints, where the equality holds, become important.

Equality Constrained Problem

- Consider the model problem

$$\begin{array}{rc} \text{Minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{g}(\mathbf{x}) = 0, \end{array}$$

  - ◇ $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$.
  - ◇ $\mathbf{g} : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $m < n$.

- A necessary condition that $\mathbf{x}_0$ be a relative minimum is that there exists $\lambda_0 \in \mathbb{R}$, $\lambda \in \mathbb{R}^m$ with $(\lambda_0, \lambda) \neq \mathbf{0}$ such that

  - ◇ If

$$F(\mathbf{x}, \lambda) := \lambda_0 f(\mathbf{x}) + \lambda^\top \mathbf{g}(\mathbf{x}),$$

  - ◇ Then

$$\nabla F(\mathbf{x}_0, \lambda) = 0.$$

- The proof is quite insightful!
  - Consider the artificial function $H : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}^m$,

$$H(\mathbf{x}, u) := (f(\mathbf{x}) - f(\mathbf{x}_0) - u, \mathbf{g}(x)).$$

  - Obviously $H(\mathbf{x}_0, 0) = 0$.
  - Note that

$$\left.\frac{\partial H}{\partial \mathbf{x}}\right|_{(\mathbf{x}_0,0)} = \left[\nabla f(\mathbf{x}_0), \nabla g_1(\mathbf{x}_0), \ldots, \nabla g_m(\mathbf{x}_0)\right].$$

  - If the columns are linear independent, then by the *implicit function theorem* there exists a neighborhood of $(\mathbf{x}_0, 0)$ in which $H(\mathbf{x}, u) \equiv 0$.
  - In particular, there exist $\mathbf{x}$ and $u < 0$ such that $f(\mathbf{x}) = f(\mathbf{x}_0) + u < f(\mathbf{x}_0)$.
- If $\nabla g(\mathbf{x}_0)$ has linearly independent "rows", then $\lambda_0 \neq 0$.
- What is the geometric meaning of the Lagrange multipliers?

Inequality Constrained Problems

- Consider the model problem

$$\begin{array}{rll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & \mathbf{g}(\mathbf{x}) & = 0, \\ & \mathbf{h}(\mathbf{x}) & \leq 0, \end{array}$$

  - ◇ $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$.
  - ◇ $\mathbf{g} : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $m < n$.
  - ◇ $\mathbf{h} : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^p$.

- A necessary condition that $\mathbf{x}_0$ be a relative minimum is that there exists vector $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that
  - ◇ $\nabla f(\mathbf{x}_0) + \lambda^\top \nabla \mathbf{g}(\mathbf{x}_0) + \mu^\top \nabla \mathbf{h}(\mathbf{x}_0) = \mathbf{0}$.
  - ◇ $\mu^\top \mathbf{h}(\mathbf{x}_0) = \mathbf{0}$.
    - ▷ If $h_j(\mathbf{x}_0) < 0$, then $\mu_j = 0$.
    - ▷ If $\mu_j > 0$, then $h_j(\mathbf{x}_0) = 0$.
  - ◇ $\mu_j \geq 0$, $j = 1, \ldots, m$.

- Prove the above conditions. (The Kuhn-Tucker Theorem)

Linear Programming

The development of linear programming has been ranked among the most important scientific advances of the mid-twentieth century.

- Linear programming typically deals with the problem of allocating *limited resources* among *competing activities* in the best possible, i.e., optimal, way.
- The adjective “linear” means that all the mathematical functions in the underling model are required to be linear functions.
- The most noted techniques for solving linear programming problem have been the *simplex method* and then the *interior point method*.

Discrete Model

- Suppose that there are m limited resources to be allocated among n competing activities.
- Define the following notation:
  - ◇ x_j = the level of activity j.
  - ◇ c_j = the "increase" in profit Z resulted from each unit increase in x_j.
  - ◇ b_i = the amount of resource i available for allocation.
  - ◇ a_{ij} = the amount of resource i consumed by each unit of activity j.
- Standard form:

$$\begin{array}{rll}
\text{Maximize} & Z = c_1x_1 + \ldots c_nx_n & \\
\text{subject to} & a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & \leq b_1 \\
& a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & \leq b_2 \\
& \vdots & \\
& a_{m1}x_2 + a_{m2}x_2 + \ldots a_{mn}x_n & \leq b_m \\
& x_1, x_2, \ldots, x_n & \geq 0.
\end{array}$$

- The feasible solutions, that is, the vector $\mathbf{x} = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$ whose components satisfy all the constraints, form a convex set cut out by hyperplanes.

- Some critical observations:
  - The optimal solution occurs necessarily at "corner-point" feasible solutions.
  - There are only finitely many corner-point feasible solutions.
  - If a corner-point feasible solution is better (as measured by Z) than all its *adjacent* corner-point feasible solutions, then it is better than all other corner-point feasible solutions, i.e., it is optimal.
- The main idea in the simplex method is to identify the corner-point solutions and to check the optimality among adjacent corner-point solutions effectively.
  - Details of the simplex method will not be reviewed in this note.

Dual Problem

- Denote the *primal* linear programming problem in matrix form

$$\begin{array}{rll} \text{Maximize} & \mathbf{c}^\top \mathbf{x} & \\ \text{subject to} & A\mathbf{x} & \leq \mathbf{b} \\ & \mathbf{x} & \geq \mathbf{0}. \end{array}$$

- The associated *dual* problem is

$$\begin{array}{rll} \text{Minimize} & \mathbf{y}^\top \mathbf{b} & \\ \text{subject to} & \mathbf{y}^\top A & \geq \mathbf{c}^\top \\ & \mathbf{y} & \geq \mathbf{0}. \end{array}$$

  - ◇ What is the meaning of the dual problem?

- Define

  - ◇ y_i = the "shadow" price (affected by the "market" price of products) of each unit of resource i.

- Want to

$$\begin{array}{rll} \text{Minimize} & y_1b_1 + \ldots + y_mb_m & \text{(Total value of resources)} \\ \text{subject to} & y_1a_{11} + \ldots + y_ma_{m1} & \geq c_1 \\ & y_1a_{12} + \ldots + y_ma_{m2} & \geq c_2 \\ & \vdots & \\ & y_1a_{1n} + \ldots + y_ma_{mn} & \geq c_n \\ & y_1, \ldots, y_m & \geq 0 \end{array}$$

  - ◇ One unit of activity j will consume, respectively, $a_{1j}, \ldots, a_{mj}$ units of resources.
  - ◇ The expense of one unit of activity j, i.e., $\sum_{i=1}^{m} y_i a_{ij}$ must generate at least the value c_j in profit Z.

- Weak Duality Theorem:

$$\mathbf{c}^\top \mathbf{x} \leq \mathbf{y}^\top A\mathbf{x} \leq \mathbf{y}^\top \mathbf{b}.$$

- Duality Theorem:

$$\max \mathbf{c}^\top \mathbf{x} = \min \mathbf{y}^\top \mathbf{b},$$

  if the feasible set is not empty.

An Illustration of the Dual Problem

- Assumptions:
  - ◇ The minimal daily requirement of calories and vitamins for an adult are 750 calories and 400 units of vitamins.
  - ◇ There are 5 categories of food to choose from with the following nutrition content:

	Calories	Vitamins	Market Prices
A	1	0	2
B	0	1	20
C	1	0	3
D	1	1	11
E	1	1	12

	Calories	Vitamins	Market Prices
Dentice alle olive	1	0	2
Polipetti in casseruola	0	1	20
Taralli	1	0	3
Fettuccine al frutto di riccio	1	1	11
Orecchiette con le braciole	1	1	12

- To find the optimal diet,
  - ◇ Let x_i, $i = 1, \ldots, 5$, denote the respective amount of food in a diet.
  - ◇ Want to minimize

$$W = 2x_1 + 20x_2 + 3x_3 + 11x_4 + 12x_5,$$

    subject to

$$\begin{aligned} x_1 + x_3 + x_4 + x_5 &\geq 750, \\ x_2 + x_4 + x_5 &\geq 400, \\ x_1, \ldots, x_5 &\geq 0. \end{aligned}$$

  - ◇ W is the total price paid for food.

- The dual problem:
  - ◇ Let

$$\begin{aligned} y_1 &= \text{the unit price of calorie,} \\ y_2 &= \text{the unit price of vitamins.} \end{aligned}$$

  - ◇ Want to maximize

$$Z = 750y_1 + 400y_2,$$

subject to

$$\begin{aligned} y_1 &\le 2, \\ y_2 &\le 20, \\ y_1 + y_2 &\le 3, \\ y_1 + y_2 &\le 11, \\ y_1 + y_2 &\le 12, \\ y_1, y_2 &\ge 0. \end{aligned}$$

  - ◇ Z is the total price paid for nutrition.

Convex Programming

- A more general setting that includes the linear programming problem is the so called *convex programming* problem.

$$\begin{array}{rll} \text{Minimize} & f(\mathbf{x}), & \\ \text{subject to} & G(\mathbf{x}) & \leq \theta, \\ & \mathbf{x} & \in \Omega \end{array}$$

  where

  - ◇ Ω is a convex set of a vector space X.
  - ◇ $f : \Omega \longrightarrow \mathbb{R}$ is a convex functional.
  - ◇ $G : \Omega \longrightarrow Z$ is a convex map into a normed vector space Z with positive cone.

A Continous Case

- Suppose a single goods is being produced which is either sold for reinvestment or kept in the inventory for future usage.
- Assume
  - ⋄ $x_1(t)$ = rate of production.
  - ⋄ $x_2(t)$ = rate of reinvestment.
  - ⋄ $x_3(t)$ = rate of storage.
- Whereas
  - ⋄ $x_1(t) = x_2(t) + x_3(t)$.
  - ⋄ The reinvestment increases the production rate via $\dot{x}_1(t) = x_2(t)$ with initial rate $x_1(0) = x_0 > 0$.
  - ⋄ $x_1(t), x_2(t), x_3(t) \geq 0$ for all $t \geq 0$.
- Want to maximize the storage over a period of time $[0, T]$.
- Express the problem in terms of $x_2(t)$ as

$$\begin{array}{ll} \text{Maximize} & \int_0^T \left\{ x_0 + \int_0^t x_2(\tau)d\tau - x_2(t) \right\} dt \\ \text{subject to} & x_0 + \int_0^t x_2(\tau)d\tau \geq x_2(t) \\ & x_2(t) \geq 0 \end{array}$$

  - ⋄ This is an infinite-dimensional "linear" programming problem.
  - ⋄ The feasible solutions are defined "implicitly" by the constraints.
  - ⋄ The problem can be handled by the theory of Lagrange multiplier.

Calculus of Variations

The earliest work of optimization, called the *calculus of variations*, began in 1969 with the Brachistochrone problem. It then developed into the *optimal control theory*, mainly by the governments for military usage, in the 1950's. The theory of *nonlinear programming*, characterized by the use of Lagrange multiplier principles, came into play only in the last forty years.

- Calculus of Variations and optimal control theory involve *functionals* as opposed to real-valued functions involved in nonlinear programming.
- The idea of differentiation from calculus can still be used in the space of functionals. This is the basis of the *Euler-Lagrange equation*.

The Brachistochrone Problem

- Assume that a particle move on a vertical plane under the influence of gravity only. Find the path so that the particle moves between two points in the shortest time.
- Assume the two points are located at $(a, 0)$ and (b, B) with $a < b$ and $B > 0$.
- Some physics:
  - ◇ Energy transformation: $\frac{1}{2}mv^2(t) = mgy(x)$.
  - ◇ $v(t) = \frac{ds}{dt}$ where $s(t) =$ distance travelled.
  - ◇ Travel time versus the current logistics:

$$dt = \frac{ds}{v} = \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{2gy(x)}} dx.$$

- Formulation:

$$\begin{array}{rl} \text{Minimize} & t = \int_a^b \frac{\sqrt{1+(y'(x))^2}}{\sqrt{2gy(x)}} dx \\ \text{subject to} & y(a) = 0, y(b) = B. \end{array}$$

  - ◇ Find the optimal solution $y(x)$.

Euler-Lagrange Equation

- Consider the general problem

$$\begin{array}{rl} \text{Minimize} & I(\mathbf{y}) = \int_a^b f(x, \mathbf{y}(x), \mathbf{y}'(x))dx, \\ \text{subject to} & \mathbf{y}(a) = \mathbf{y}_a, \mathbf{y}(b) = \mathbf{y}_b, \end{array}$$

  where $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ is smooth and $\mathbf{y} : [a, b] \longrightarrow \mathbb{R}^n$ is piecewise smooth.

- Define

$$F(\epsilon) := \int_a^b f(x, \mathbf{y}(x, \epsilon), \mathbf{y}'(x, \epsilon))dx,$$

  - ⋄ $\mathbf{y}(x, \epsilon) := \mathbf{y}(x) + \epsilon \mathbf{z}(x)$.
  - ⋄ $\mathbf{z}(x)$ is piecewise continuous with $\mathbf{z}(a) = \mathbf{z}(b) = \mathbf{0}$.

- Expand $F(\epsilon)$ around $\epsilon = 0$ to obtain

$$\begin{aligned} F(\epsilon) \;&=\; F(0) + \epsilon \underbrace{\int_a^b \left(\mathbf{z}^\top(x) \frac{\partial f}{\partial \mathbf{y}} + \mathbf{z}'(x)^\top \frac{\partial f}{\partial \mathbf{y}'} \right) dx}_{F'(0)} \\ &+ \frac{1}{2}\epsilon^2 \underbrace{\int_a^b \left(\mathbf{z}^\top(x) \frac{\partial^2 f}{\partial \mathbf{y}^2} \mathbf{z}(x) + 2\mathbf{z}(x)^\top \frac{\partial^2 f}{\partial \mathbf{y} \partial \mathbf{y}'} \mathbf{z}'(x) + \mathbf{z}'(x)^\top \frac{\partial^2 f}{\partial \mathbf{y}' \partial \mathbf{y}'} \mathbf{z}'(x) \right) dx}_{F''(0)} + O(\epsilon^3). \end{aligned}$$

- From calculus, the necessary condition of being a minimizer is $F'(0) = 0$ and $F''(0) \geq 0$.
  - ◇ Upon integration by parts,

$$0 = F'(0) = \mathbf{z}(x)^\top \int_a^x \frac{\partial f}{\partial \mathbf{y}} ds \Big|_a^b + \int_a^b \mathbf{z}'(x)^\top \left(-\int_a^x \frac{\partial f}{\partial \mathbf{y}} ds + \frac{\partial f}{\partial \mathbf{y}'} \right) dx$$

    for all piecewise continuous function $\mathbf{z}(x)$ with zero boundary values.
  - ◇ It follows that at the critical point $\mathbf{y}_0$, the Euler-Lagrange equation (in integral form)

$$\frac{\partial f}{\partial \mathbf{y}'}(x, y_0(x), y_0'(x)) = \int_a^x \frac{\partial f}{\partial \mathbf{y}}(s, y_0(s), y_0'(s)) ds + c$$

    must be satisfied.
  - ◇ Equivalently, the Euler-Lagrange equation in differential form is

$$\frac{\partial f}{\partial \mathbf{y}}(x, y_0(x), y_0'(x)) = \frac{d}{dx} \frac{\partial f}{\partial \mathbf{y}'}(x, y_0(x), y_0'(x)).$$

- Use the Euler-Lagrange equation to solve the brachistochrone problem.

Optimal Control

- The calculus of variations is a special case of optimal control theory achieved by considering the control variable $\mathbf{u}(t)$ to be the derivative of the state variable $\mathbf{x}'(t)$.
- The first necessary condition in optimal control theory is the Pontryagin Maximum Principle.
- The idea of Lagrange multipliers in calculus can be carried over to abstract function space with the notion of Fréchet derivatives.

Basic Formulation

- Consider the optimization problem

$$\begin{aligned} \text{Minimize} \quad J(\mathbf{x},\mathbf{u}) &= h(b,\mathbf{x}(b)) + \int_a^b f(t,\mathbf{x},\mathbf{u})dt \\ \text{subject to} \quad \mathbf{x}'(t) &= g(t,\mathbf{x},\mathbf{u}) \\ \mathbf{x}(a) &= \xi, \\ \mathbf{x}(b) &= \eta, \\ \mathbf{u} &\in \mathcal{U}. \end{aligned}$$

  - ◇ $\mathbf{x}(t) = [x_1(t), \ldots, x_n(t)]^\top$ is the *state variable.*
  - ◇ $\mathbf{u}(t) = [u_1(t), \ldots, u_m(t)]^\top$ is the *control variable.*
  - ◇ $\mathcal{S}(b)$ is a surface in $\mathbb{R}^n$.
  - ◇ $\mathcal{U} \subset \mathbb{R}^m$ is the set of controls.
  - ◇ f, g, h are sufficiently smooth functions.
- Find the optimal control $\mathbf{u}_0(t)$ from $\mathcal{U}$ that drives the trajectory of $\mathbf{x}_0(t)$ starting from (a, ξ) according to the trajectory equation to the point $(b, \mathbf{x}(b))$ with $\mathbf{x}(b) \in \mathcal{S}$ so that the objective functional $J(\mathbf{x}, \mathbf{u})$ is minimized.

Hamilton-Jacobi Equations

- The above problem can be case as an equality constrained optimization of the form

$$\begin{array}{rl} \text{Minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{g}(\mathbf{x}) = 0, \end{array}$$

  - ◇ $f : X \longrightarrow \mathbb{R}$.
  - ◇ $\mathbf{g} : X \longrightarrow Z$.
  - ◇ Both X and Z are some Banach spaces.

- Being able to measure the norm of a vector, we should be able to define the notion of Fréchet derivative and smoothness on the abstract functions f and g.

- The notion of Lagrange multiplier can be generalized which leads to the necessary conditions for the optimal control.

  - ◇ For $\mathbf{p}(t) = [p_1(t), \ldots, p_n(t)]^\top$, define the *Hamiltonian function*

$$H(t, \mathbf{x}, \mathbf{u}, \mathbf{p}) = f(t, \mathbf{x}, \mathbf{u}) + \mathbf{p}^\top g(t, \mathbf{x}, \mathbf{u}).$$

  - ◇ $(\mathbf{x}_0, \mathbf{u}_0)$ is an optimal control only if there exists a function $\mathbf{p}_0(t)$ such that
    - ▷ $\mathbf{x}'(t) = H_\mathbf{p}(t, \mathbf{x}, \mathbf{u}, \mathbf{p})$.
    - ▷ $\mathbf{p}'(t) = -H_\mathbf{x}(t, \mathbf{x}, \mathbf{u}, \mathbf{p})$.
    - ▷ $H_\mathbf{u}(t, \mathbf{x}, \mathbf{u}, \mathbf{p}) = 0$.
    - ▷ $\mathbf{x}(a) = \xi$.
    - ▷ $\mathbf{x}(b) = \eta$.