Introduction to Optimization

QF607 Numerical Methods

Zhenke Guan
zhenkeguan@smu.edu.sg

Optimization

- From a mathematical perspective, optimization deals with finding the maxima or minima of a function that depends on one or more variables.
- Applications in QF:
  - Apply optimization methods to help making better decisions.
  - Make the best possible choice from a set of possible choices
- Examples are everywhere
  - Calibrating model parameters to fit the market observables
  - Investors select portfolios to maximize return or minimize risk

Optimization Process

- Understand and formulate the problem
  - The objective function — what to optimize?
  - The variables, or the model parameters
  - The constraints
- Solve the problem using numerical methods
  - Study the properties of the problem
  - Design and implement algorithm
  - Validate the solution

Problem Formulation

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{x} \in A \tag{1}$$

where $f : R^n \to R$ is the objective function, and A is the feasible region. Note that maximizing $f(\mathbf{x})$ is equivalent to minimizing $-f(\mathbf{x})$.

Types of Optimization Problems

Categorize by constraints:

- Unconstrained optimization — $A = R^n$
- Constrained optimizations

Types of Optimization Problems

Categorize by f and A:

- Linear programing (f is linear and A is a polyhedron),
- Quadratic programing (f is quadratic and A is a polyhedron),
- Convex problems (f is convex function and A is a convex set),
- General non-linear problems

Depending on the problem, solution can be:

Global vs Local

A global optimum represents the very best solution while a local optimum is better than its immediate neighbors.

We will discuss the general approach.

Unconstrained Optmization

Optimality Necessary Conditions

If f is sufficiently differentiable, and it has local minimum at a point $\hat{x}$, then

$$f'(\hat{x}) = 0 \quad \text{and} \tag{2}$$
$$f''(\hat{x}) \succcurlyeq 0 \tag{3}$$

Optimality Sufficient Conditions

The sufficient condition for $\hat{x}$ be the local minimum is

$$f'(\hat{x}) = 0 \quad \text{and} \tag{4}$$
$$f''(\hat{x}) \succ 0 \tag{5}$$

For convex problem, $f''(\hat{x}) \succ 0$ is guaranteed, so we need to consider only $f'(\hat{x}) = 0$

E.g. $f(x) = x^2$ — convex, $f(x) = x^3$ — non-convex, no local minimum

Optimaility Conditions

Rationale

Taylor expansion at neigbhour of $\hat{x}$:

$$f(\hat{x}+h) = f(\hat{x}) + f'(\hat{x})h + \frac{1}{2}f''(\hat{x})h^2 + O(h^3) \tag{6}$$

Eliminating the higher order terms when $h \to 0$, $f(\hat{x}+h) > f(\hat{x})$ if $f'(\hat{x}) = 0$ and $f''(\hat{x}) > 0$.

Optimization Solution using Root Finder

- The problem of optimization becomes root-finding — find x such that $f'(x) = 0$, and then verify the second derivative.
- For one dimensional problem, we can use our root-searcher of choice: bisection, secant, false position method, Brent, etc.
- For multi-dimensional problem, we can use Newton method:
  - Start with an initial guess at x_0, then iteratively:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \frac{1}{f''(\mathbf{x})} f'(\mathbf{x}) \tag{7}$$

    where $f'(\mathbf{x})$ is the gradient $[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}]$, and $f''(\mathbf{x})$ is the Jacobian of the gradient — the Hessian matrix, whose element H_{ij} is

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Unconstrained Optimization without Derivatives

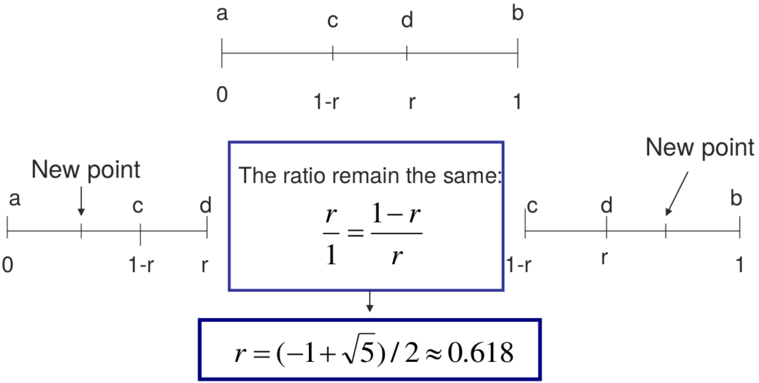
- Sometimes, the derivatives of f is not available, numerical methods are needed, approximating the derivatives — Quasi Newton method.

- Or, there is also methods that do not work with derivatives — Golden Section Search.

Golden Section Search

- Assume there is minimum within $[a, b]$
- Select two interior points c and d where $a < c < d < d$
- If $f(c) < f(d)$, choose $[a, d]$ as the next interval to search
- If $f(c) > f(d)$, choose $[c, b]$ as the next interval to search

Golden Section Search



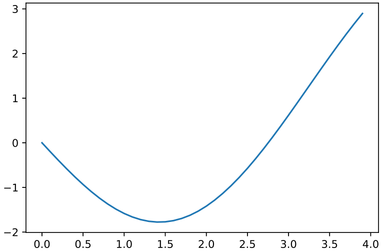
- The interior points c and d are selected so that the resulting interval $[a, c]$ and $[d, b]$ are symmetric: $b - d = c - a$. We denote $r = \frac{d-a}{b-a}$
- The interval width shrink by the same constant proportion in each step — the ratio r remain the same on each subinterval.
- For the next step, to save the computational cost, both c and d will be used, so we want $\frac{r}{1} = \frac{1-r}{r}$.

Example

- Find the minimum of

$$f(x) = \frac{x^2}{10} - 2\sin x \tag{8}$$

within the interval $[0, 4]$.



Other Iterative Methods

- Most optimization algorithms are iterative:

$$x_{i+1} = x_i + \alpha \mathbf{S} \tag{9}$$

  where **S** is the vector search direction, α is a scalar — search distance.

- Different algorithm determines the search direction S and the scalar according to different criteria, e.g.:
  - Gradient decent — search along the gradient direction, and choose α that minimize $f(\mathbf{x_i} + \alpha \mathbf{S})$ at each iteration
  - Conjugate gradient — http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf

Constrained Optimization

- A standard constrained optimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad \text{s.t. } g(\mathbf{x}) = 0. \tag{10}$$

- We just need to transform the problem to unconstrained optimization problem

- A constrained minimization can be written as an unconstrained minimization by defining the Lagrangian function:

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(x) \tag{11}$$

$L(\mathbf{x}, \lambda)$ is called the Lagrangian function, and λ the Lagrange multiplier.

Constrained Optimization

- Necessary condition for optimality: the gradient of the Lagrangian is 0:

$$\nabla f(\mathbf{x}) + \lambda \nabla g(\mathbf{x}) = 0 \tag{12}$$

$$\frac{\partial L}{\partial \lambda} = 0 \tag{13}$$

- Solve the system, we get the optimum.

Example

- Minimize $f(x, y) = x^2 + 3y^2$, s.t. $x + y = 1$

Inequality Constraints

- An optimization problem with inequality constrainst:

$$\min_{\mathbf{x}} f(x), \quad \text{s.t.} \quad g(\mathbf{x}) \geq a \tag{14}$$

- Introduce a "slack variable" s for each inequality, we then translate them to equality constraints:

$$p(x) = g(x) - a - s^2 = 0 \tag{15}$$

- We can then use Lagrange multpiler:

$$\nabla f(\mathbf{x}) + \lambda \nabla p(\mathbf{x}) = 0 \tag{16}$$

$$\frac{\partial L}{\partial \lambda} = 0 \tag{17}$$

$$\frac{\partial L}{\partial s} = 0 \tag{18}$$

Example

- Minimize $f(x, y) = x^2 + 3y^2$, s.t. $x + y \geq 1$

Application: Portfolio Optimization

- Consider a portfolio of n assets, the expected return and variance are given by

$$\mu_p = \mathbf{w}^\top \mu \tag{19}$$
$$\sigma_p^2 = \mathbf{w}^\top \sum \mathbf{w} \tag{20}$$

  where $\mu = [\mu_1 \dots \mu_n]^\top$ is the expected return of each asset, $\mathbf{w} = [w_1 \dots w_n]^\top$, $\sum_{w_i} = 1$ and $\sum$ is the covariance matrix.

- The goal is to choose a value for $\mathbf{w}$ which gives a large μ_p and a small σ_p^2.

Application: Portfolio Optimization

- A portfolio is efficient if for a fixed μ_p, there is no other portfolio which has a smaller variance σ_p^2
- A portfolio is efficient if for a fixed σ_p^2, there is no other portfolio which has a larger μ_p

Markowitz's Morden Portfolio Theorem

$$\min_{\mathbf{w}} \left[-\mathbf{w}^\top \mu + \frac{1}{2} \gamma \mathbf{w}^\top \sum \mathbf{w} \right], \quad \text{s.t. } \sum_i w_i = 1 \tag{21}$$

This model allows to build an efficient frontier (those portfolios that, for a given return, have a minimum risk-level)