MGTF 413: Computational Finance Methods Lecture Notes

Jiahui Shui

January 16, 2024

Some notes may deviate from what we learned in course. Use at your own risk.

1 Optimization

An optimization problem looks like:

$$\min_{x \in S} f(x) \tag{1}$$

f is called objective function. The components of $x \in \mathbb{R}^n$ are the decision variables. S is the constraint set or feasible set. $x^* = \text{argmin}_{x \in S} f(x)$ is called the minimizer.

Rather than writing in argmax/argmin form, I'll write the optimization into the following form:

$$\begin{aligned} \max_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) = 0, \quad i \in \mathcal{I} \\ & h_j(x) \geq 0, \quad j \in \mathcal{J} \end{aligned} \tag{2}$$

1.1 Linear Programming

A function $l(x)$ for $x \in \mathbb{R}^n$ is called linear if $l(x)$ is a linear combination of the components $x_1, \cdots, x_n$. That is, we can find a vector $c \in \mathbb{R}^n$ such that $l(x) = c^T x$. Property: $l(\alpha x) = \alpha l(x)$ and $l(x + y) = l(x) + l(y)$ for any $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

The graph of a linear function $l(x) = c^T x, x \in \mathbb{R}^n$ is an $n-$dimensional plane living in $\mathbb{R}^{n+1}$. For example, consider $x \in \mathbb{R}$, then $l(x) = cx$ is a line in $\mathbb{R}^2$.

Definition 1 (Level Sets) *We call $\{x | g(x) = \alpha\}$ the $\alpha-$level set of function $g(x)$.*

Definition 2 (Hyperplane) *We call $\{x | c^T x = \alpha\}, c \neq 0$ a hyperplane, which is a $n - 1$ dimensional hyperplanes in $\mathbb{R}^n$.*

Definition 3 (Half-Space) *We call $\{x | c^T x \geq \alpha\}, c \neq 0$ a half space. c is the **outer-norm** of the half-space.*

Standard form of LP:

$$\begin{aligned} \max_{x} \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{3}$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$. The constraint $x \geq 0$ denotes $x_i \geq 0$ for all $i = 1, \cdots, n$.

Now we might have a question, what if the given problem is not the standard form? For example, consider the following optimization problem:

$$\begin{aligned} \max_{x_1, x_2} \quad & c_1 x_1 + c_2 x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \leq 12 \end{aligned} \tag{4}$$

Then we can introduce four non-negative variables: y_1, z_1, y_2, z_2, such that

$$x_1 = y_1 - z_1, \quad x_2 = y_2 - z_2$$

Hence, we can rewrite the optimization problem (4) into the following form:

$$\begin{aligned} \max_{y_1, y_2, z_1, z_2} \quad & c_1 y_1 + c_2 y_2 - c_1 z_1 - c_2 z_2 \\ \text{s.t.} \quad & 2y_1 + y_2 - 2z_1 - z_2 \leq 12 \\ & y_1, y_2, z_1, z_2 \geq 0 \end{aligned} \tag{5}$$

Furthermore, introduce $w \geq 0$, then

$$\begin{aligned} \max_{y_1, y_2, z_1, z_2, w} \quad & c_1 y_1 + c_2 y_2 - c_1 z_1 - c_2 z_2 \\ \text{s.t.} \quad & 2y_1 + y_2 - 2z_1 - z_2 + w = 12 \\ & y_1, y_2, z_1, z_2, w \geq 0 \end{aligned} \tag{6}$$

That is, we can add more decision variables into the optimization problem to convert it into standard form. These additional variables are called surplus and slack variables. Summary of procedures:

1. Introduce non-negative variables ($x \geq 0$ in standard form)

2. Convert inequalities into equalities. $Ax \leq b$ can be converted into $Ax + y = b$ for $y \geq 0$. ($Ax \geq b$ can be written as $Ax = b + y$).

Skip this if you want. Simplex Method. For more details, take a look at chapter 2 and chapter 3 of [1]

Definition 4 *A point x in a convex set C is said to be an extreme point of C if there are **no** two distinct points $x_1, x_2 \in C$ such that $x = \alpha x_1 + (1 - \alpha) x_2$ for some $\alpha \in (0, 1)$.*

Definition 5 (Polytope, Polyhedron) *A set which can be expressed as the intersection of a finite number of closed half spaces is said to be a convex polytope. A nonempty bounded polytope is called a polyhedron.*

1.2 Non-linear Optimization

Example 1 (Markowitz Mean-Variance Optimization) *Let x_i be the proportion of the portfolio invested in asset i, and μ_i be the expected return of asset i. Moreover, let x and μ denote corresponding vector of x_i and μ_i. Σ is the covariance matrix of stocks, i.e.*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{pmatrix}$$

For simplicity, we denote

$$e = (1, 1, \cdots, 1)^T$$

Therefore, the portfolio has expectation and variance:

$$\mathbb{E}[x] = \mu^T x \quad \mathrm{Var}[x] = x^T \Sigma x$$

The optimization problem is (we allow short-sale here.)

$$\begin{aligned} \min_x \quad & x^T \Sigma x \\ \text{s.t.} \quad & \mu^T x \geq R \\ & e^T x = 1 \end{aligned} \tag{7}$$

Solution: Now we will use matrix calculus and KKT to help us to solve this problem: The Lagrangian is

$$\mathcal{L} = x^T \Sigma x + \lambda(e^T x - 1) + \nu(R - \mu^T x) \tag{8}$$

where λ, ν are Lagrange multipliers. We have

$$\frac{\partial \mathcal{L}}{\partial x} = 2\Sigma x + \lambda e - \nu_1 \mu = 0$$

Hence

$$x^* = \Sigma^{-1}\left(\frac{1}{2}\nu\mu - \frac{1}{2}\lambda e\right)$$

If $\mu^T x > R$, then $\nu = 0$ due to complementary slackness condition. However, since Σ is positive semi-definite, then so does Σ^{-1}. $\nu = 0$ will lead to $x^* \leq 0$, which is not feasible. So $\mu^T x = R$.

Therefore, we will get the following equation system from KKT:

$$\begin{pmatrix} 2\Sigma & e & \mu \\ e^T & 0 & 0 \\ \mu^T & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \\ \nu \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ R \end{pmatrix} \tag{9}$$

Solving this system we get the optimal x^*: ■

Remark 1 *When non-negative constraint is added, there is no closed-form solution for this problem.*

There are some (numerical) methods to handle non-linear optimization problem:

- (Steepest) descent method (calculus based)
- Newton's method (calculus based)
- Interior point methods
- Sequential quadratic programming

1.2.1 Gradient-Descent

The steepest descent direction for objective function is $-\nabla f(x)$, i.e. negative gradient direction. Steps:

1. Start with location that is a guess of the minimizer: x^0
2. Move a certain distance in $-\nabla f(x^0)$, call it x^1.
$$x^1 = x^0 - \alpha \nabla f(x^0) \tag{10}$$
3. Iterate by
$$x^k = x^{k-1} - \alpha \nabla f(x^{k-1}) \tag{11}$$

actually, α can be varying.

1.2.2 Newton's Method

Newton's method is to solve root for a (nonlinear) function $g(x)$. Recall the first order condition is just $\nabla f(x) = 0$. So we can use Newton's method to find root for gradient, therefore it might be possible minimum/maximum of objective function.

Algorithm for find root for $g(x)$ (univariate):

1. Start with x^0
2. Iterate
$$x^{k+1} = x^k - \frac{g(x^k)}{g'(x^k)} \tag{12}$$
3. Stop iterations if $|g(x^k)|$ is small or $|x^k - x^{k-1}|$ is small.

For univariate optimization problem: $\min_x \quad f(x)$:

1. Start with x^0

2. Iterate

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} \tag{13}$$

3. Stop iterations if $|f'(x^k)|$ is small or $|x^k - x^{k-1}|$ is small.

In multivariate case, first order derivative is gradient, second order derivative is Hessian matrix. Therefore, if we want to find root for $G(x), x \in \mathbb{R}^n$, we can do iteration:

$$x^{k+1} = x^k - [\nabla G(x^k)]^{-1} G(x^k) \tag{14}$$

For $\min_{x \in \mathbb{R}^n} F(x)$, we have

$$x^{k+1} = x^k - [\nabla^2 F(x^k)]^{-1} \nabla F(x^k) \tag{15}$$

where

$$\nabla^2 F(x) = \left(\frac{\partial^2 F}{\partial x_i \partial_j}\right)_{ij} = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & \cdots & \frac{\partial^2 F}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{pmatrix}$$

I wrote a tutorial for gradient descent and Newton's method. One can find it at Tutorial.

Example 2 (IRR) *(Find definition of IRR at your corporate finance textbook). The IRR of a bond is called its yield. Suppose that a non-callable bond has a maturity of 4 years. The par(face) value is 1000 and the price today is 900. The coupon rate is 10%, annually. Calculate the yield of this bond.*

Solution: We just have to calculate root for following function:

$$g(r) = \frac{100}{1+r} + \frac{100}{(1+r)^2} + \frac{100}{(1+r)^3} + \frac{1100}{(1+r)^4} - 900$$

Choose one programming language to do it! ■

The Quadratic Programming(QP) problem is a simple nonlinear constrained optimization problem. Standard form of QP:

$$\begin{aligned} \min_x \quad & \frac{1}{2} x^T Q x + c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{16}$$

where $Q \in \mathbb{R}^{n \times n}, A \in \mathbb{R}^{m \times n}$.

2 Partial Differential Equations

References

[1] David G Luenberger, Yinyu Ye, et al. *Linear and nonlinear programming*. Vol. 2. Springer, 1984.

Appendix A: Matrix Calculus

A.1 Scalar Function

Suppose that $f(X)$ is a scalar function of matrix X ($m \times n$). Then the total derivative of f is

$$\mathrm{d}f = \sum_{i=1}^{m} \sum_{j=1} \frac{\partial f}{\partial X_{ij}} \mathrm{d}X_{ij} = \operatorname{tr}\left(\frac{\partial f}{\partial X}^T \mathrm{d}X \right) \tag{17}$$

We can use this formula to find the derivative. Here are some properties:

1. $\mathrm{d}(X \pm Y) = \mathrm{d}X \pm \mathrm{d}Y$
2. $\mathrm{d}(XY) = (\mathrm{d}X)Y + X(\mathrm{d}Y)$
3. $\mathrm{d}(X^T) = (\mathrm{d}X)^T$
4. $\mathrm{d}(\operatorname{tr}(X)) = \operatorname{tr}(\mathrm{d}X)$
5. **Inverse**: $\mathrm{d}X^{-1} = -X^{-1}(\mathrm{d}X)X^{-1}$. Sketch of proof: Take differentiation at BHS of $XX^{-1} = I$.
6. **Determinant**: $\mathrm{d}|X| = \operatorname{tr}(X^* \mathrm{d}X)$, where X^* is the adjugate matrix of X. When X is invertible, then $\mathrm{d}|X| = |X| \operatorname{tr}(X^{-1}\mathrm{d}X)$.
7. $\mathrm{d}(X \odot Y) = \mathrm{d}X \odot Y + X \odot \mathrm{d}Y$, where $\odot$ denotes element-wise product, (or Hadamard product, etc.), i.e. $(A \odot B)_{ij} = (A)_{ij}(B)_{ij}$
8. Element-wise Function: suppose that $\sigma(X) := [\sigma(X_{ij})]$. $\sigma'(X) := [\sigma'(X_{ij})]$. Then $\mathrm{d}\sigma(X) = \sigma'(X) \odot \mathrm{d}X$. For example:

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}, \quad \mathrm{d}\sin(X) = \begin{pmatrix} \cos X_{11}\mathrm{d}X_{11} & \cos X_{12}\mathrm{d}X_{12} \\ \cos X_{21}\mathrm{d}X_{21} & \cos X_{22}\mathrm{d}X_{22} \end{pmatrix} = \cos(X) \odot \mathrm{d}X$$

Some tricks for **trace**:

1. For scalar, $a = \operatorname{tr}(a)$
2. $\operatorname{tr}(A^T) = \operatorname{tr}(A)$
3. Linearity: $\operatorname{tr}(A \pm B) = \operatorname{tr}(A) \pm \operatorname{tr}(B)$

4. **Multiplication**: $\mathrm{tr}(AB) = \mathrm{BA}$, where A has the same size of B^T.
5. $\mathrm{tr}(A^T(B \odot C)) = \mathrm{tr}((A \odot B)^T C)$, where A, B, C has the same dimension.

Ok now let's begin to look at some examples.

Example 3 *Suppose that $f = \boldsymbol{a}^T X \boldsymbol{b}$, where $\boldsymbol{a}$ is a $m \times 1$ vector while $\boldsymbol{b}$ is a $n \times 1$ vector. Find $\frac{\partial f}{\partial X}$*

Appendix B: Lagrange multiplier, KKT

2.1 B.1 Gradient