# MGTF 413: Computational Finance Methods Lecture Notes

Jiahui Shui

January 13, 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 Gradient Descent

### 1.3 Newton's Method

## 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

$$\mathrm{d}f = \sum_{i=1}^{m}\sum_{j=1}\frac{\partial f}{\partial X_{ij}}\mathrm{d}X_{ij} = \mathrm{tr}\left(\frac{\partial f}{\partial X}^{T}\mathrm{d}X\right) \tag{7}$$

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. $\mathrm{d}(\mathrm{tr}(X)) = \mathrm{tr}(\mathrm{d}X)$
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**: $\mathrm{d}|X| = \mathrm{tr}(X^*\mathrm{d}X)$, where $X^*$ is the adjugate matrix of $X$. When $X$ is invertible, then $\mathrm{d}|X| = |X|\mathrm{tr}(X^{-1}\mathrm{d}X)$.
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**: $\operatorname{tr}(AB) =$ BA, where $A$ has the same size of $B^T$.
5. $\operatorname{tr}(A^T(B \odot C)) = \operatorname{tr}((A \odot B)^T C)$, where $A, B, C$ has the same dimension.

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

**Example 1** *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