MIE377 – Financial Optimization Models

Winter 2019

Lecture Notes

Prepared by Giorgio Costa

1 Mathematics of optimization

1.1 Linear and Non-Linear programs

What is optimization?

Mathematical optimization (or 'mathematical programming') consists of maximizing or minimizing an objective function to select the best solution from a set of available alternatives.

Examples of optimization problems

Mean–Variance Optimization:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} x_i x_j \\ \text{s.t.} \quad & \sum_{i=1}^{n} \mu_i x_i = R \\ & \sum_{i=1}^{n} x_i = 1 \\ & (x_i \geq 0, \quad i = 1,\ ...,\ n) \end{aligned}$$

Bond dedication:

$$\begin{aligned} \min_{x} \quad & P_1 x_1 + P_2 x_2 + P_3 x_3 & \\ \text{s.t.} \quad & F_1 x_1 + ¢_2 x_2 + ¢_3 x_3 & \geq L_1 \\ & (F_2 + ¢_2) x_2 + ¢_3 x_3 & \geq L_2 \\ & (F_3 + ¢_3) x_3 & \geq L_3 \\ & x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0 & \end{aligned}$$

In general, an optimization problem has the following standard form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(x) && \big\} \text{ Objective function} \\ \text{s.t.} \quad & g(x) \leq 0 && \big\} \text{ Inequality constraints} \\ & h(x) = 0 && \big\} \text{ Equality constraints} \end{aligned}$$

In finance, the criteria used to formulate an optimization problem is defined by the investor, i.e., investors must determine what is their objective and what are their constraints. For this course, we will focus on applying mathematical programming to solve asset allocation problems to create optimal portfolios.

How do we determine if a problem is linear or non-linear?

The degree of an optimization problem is determined by the highest degree between its objective function and constraints.

Example: Consider the following problem

$$
\begin{array}{lll}
\min\limits_{\boldsymbol{x}} & x_1 + 2x_2 - x_1x_2 & \Big\} \text{ Quadratic} \\
\text{s.t.} & x_1 - x_2 \leq 0 \quad \Big\} & \text{linear} \\
 & x_1^4 - x_2 \geq 0 \quad \Big\} & \text{4th degree}
\end{array}
$$

Therefore, we have a non-linear (4th degree) problem.

1.2 General non-linear optimization

$$
\begin{array}{ll}
\min\limits_{\boldsymbol{x}} & f(x) \\
\text{s.t.} & x \in S \subseteq \mathbb{R}^n
\end{array}
$$

where S is some set defined by the collection of equality and inequality constraints.

Figure 1: function with local and global minima

1.2.1 Algebraically characterizing local minima

A minimum point of a function $f(x)$ can be characterized as points where the 1st order derivative is zero and where the 2nd order derivative is positive.

Now, Consider a scalar-valued function $f(x)$ with input vector $x \in \mathbb{R}^n$, i.e.,

$$
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
$$

1st order derivative: The gradient, $\nabla f(x)$, is the column vector of first-order partial derivatives of the scalar-valued function $f(x)$,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

<u>2nd order derivative</u>: The Hessian matrix, $H(x)$, is the square matrix of second-order partial derivatives of the scalar-valued function $f(x)$,

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Note that $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

<u>Conditions for a local min</u>:

1. If x^* is a local min, then $\nabla f(x^*) = 0$

   → This is known as a First Order Necessary Condition (FONC).

   <u>Example</u>: Let $f(x) = 2x_1^2 + x_1x_2 + x_1x_3 + x_2^2 + x_2x_3 + x_3^2$

$$\nabla f(x) = \begin{bmatrix} 4x_1 + x_2 + x_3 \\ x_1 + 2x_2 + x_3 \\ x_1 + x_2 + 2x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \rightarrow \quad \text{Solve for} \quad x^* = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

2. If x^* is a local min, then $\nabla f(x^*) = 0$ and $H(x^*)$ is positive semi-definite (PSD).

   → This is known as a Second Order Necessary Condition (SONC)

   → The SONC become <u>sufficient</u> if $\nabla f(x^*) = 0$ and $H(x^*)$ is positive definite (PD), i.e, we can say x^* is a local min.

   <u>Example</u>: We can find $H(x)$ by getting the partial derivatives of $\nabla f(x)$ from before,

$$H(x) = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

   If $H(x)$ is PD, then

$$x^* = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is a strict local minimum. How do we check if $H(x)$ is PD?

1.2.2 Checking for a PD/PSD matrix

Definition of positive definiteness:

- A square symmetric (or Hermitian) matrix $Q \in \mathbb{R}^{n\times n}$ is PSD if $x^T Q\ x \geq 0$ for all x
- Q is PD if $x^T Q\ x > 0$ for all x except x $= 0$

Conditions:

1. If all eigenvalues are ≥ 0 (> 0), then Q is PSD (PD), or
2. If all principal minors are ≥ 0 (> 0), then Q is PSD (PD)

Claim

We can determine if Q is PD by checking if the leading principal minors are positive, $\Delta_i > 0$.

$$
\begin{aligned}
Q &= \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \\
\Delta_1 &= \det(q_{11}) = q_{11} \\
\Delta_2 &= \det\left(\begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \right) = q_{11}q_{22} - q_{12}q_{21} \\
\Delta_3 &= \det(Q)
\end{aligned}
$$

However, if $\Delta_i = 0$ for any i, we must check all principal minors

$$
\begin{aligned}
\Delta_4 &= \det(q_{22}) = q_{22} \\
\Delta_5 &= \det\left(\begin{bmatrix} q_{22} & q_{23} \\ q_{32} & q_{33} \end{bmatrix} \right) = q_{22}q_{33} - q_{23}q_{32} \\
\Delta_6 &= \det(q_{33}) = q_{33}
\end{aligned}
$$

If $\Delta_i \geq 0$, then our matrix is PSD.

Example:

$$
Q = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}
$$

$\rightarrow$ 1st principal minor$\Delta_1 = \det(2) = 2 > 0$

$\rightarrow$ 2nd principal minor$\Delta_2 = \det\left(\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \right) = 3 > 0$

$\rightarrow$ 3rd principal minor$\Delta_3 = \det(Q) = 4 > 0 \quad \therefore Q$ is PD.

1.2.3 Characterizing global minima

Now that we have the FONC and SONC, how do we characterize a global min?

Determining global optimality relies on the nature of our objective function and feasible set (our constraints), i.e., we must identify whether the problem itself has the properties required to guarantee global optimality.

To characterize a global minima, we must have a convex objective function $f(x)$ and a convex set S

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(x) \\ \text{s.t.} \quad & x \in S \end{aligned}$$

Under these conditions any x^* such that $\nabla f(x^*) = 0$ is a global min.

1.3 Convexity

A set $S \subseteq \mathbb{R}^n$ is said to be convex if for any $x, y \in S$ we have that $\lambda x + (1-\lambda)y \in S$ for all $\lambda \in [0, 1]$.
Figure 2: Convex and Non-Convex sets

Convex functions

Figure 3: Definition of convexity

A function $f(x) : S \to \mathbb{R}$ for a convex set is said to be convex if:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \quad \forall \quad \lambda \in [0,1]$$

A quadratic function $f(x) = \frac{1}{2}x^T Q\ x + c^T x$ is convex iff (if and only if) $H(x) = Q$ is PSD over some set S.

1.4 Quadratic Programs

The general quadratic program (QP) is

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}x^T Q\ x + c^T x \\ \text{s.t.} \quad & Ax \leq b_1 \\ & Dx = b_2 \end{aligned}$$

where

- $x \in \mathbb{R}^n$ is our decision variable,
- $c \in \mathbb{R}^n$
- $Q \in \mathbb{R}^{n \times n}$ is a symmetric matrix
- $A \in \mathbb{R}^{m \times n}$
- $b_1 \in \mathbb{R}^m$
- $Ax \leq b_1$ describes our m linear inequality constraints
- $D \in \mathbb{R}^{l \times n}$
- $b_2 \in \mathbb{R}^l$
- $Dx \leq b_2$ describes our l linear equality constraints

<u>Note</u>: if $f(x)$ is a quadratic function, then its Hessian is equal to Q, $H(x) = Q$.

MVO is one example of a QP:

$$\begin{aligned}\min_{\boldsymbol{x}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}x_i x_j \\ \text{s.t.} \quad & \sum_{i=1}^{n}\mu_i x_i \geq R \\ & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0\end{aligned}$$

Numerical example:

$$\begin{aligned}\min_{\boldsymbol{x}} \quad & 4x_1^2 + x_1x_2 + 2x_1x_3 + 3.5x_2^2 + x_2x_3 + 3x_3^2 + 2x_1 - x_2 + x_3 \\ \text{s.t.} \quad & x_1 + 2x_3 \leq 5 \\ & 3x_1 + x_3 \leq 2 \\ & x_1 + x_3 = 2\end{aligned}$$

$$Q = \begin{bmatrix} 8 & 1 & 2 \\ 1 & 7 & 1 \\ 2 & 1 & 6 \end{bmatrix} \qquad c = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 0 & 1 \end{bmatrix}$$

$$D = [1\ 0\ 1] \qquad b_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \qquad b_2 = [2]$$

Claim

The tractability of solving a QP will depend on the matrix Q. The QP is considerably easier to solve (and well-behaved) if Q is PD or PSD.

Proposition

We want to solve a QP of the form

$$\begin{aligned}\min_{\boldsymbol{x}} \quad & \frac{1}{2}x^TQ\,x + c^Tx \\ \text{s.t.} \quad & Ax = b\end{aligned}$$

Fact: $S = \{x \mid Ax = b\}$ is a convex set (it is linear!)
Assuming that Q is PSD, the only complication we have is $Ax = b$. Key idea: remove the constraint by using the Lagrangian method.

This QP can be solved in closed-form by

$$\begin{bmatrix} x^* \\ \pi^* \end{bmatrix} = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} -c \\ b \end{bmatrix}$$

where π^* is the corresponding Lagrangian multiplier.

Proof: By Lagrangian method

$$\mathcal{L} = \frac{1}{2}x^T Q\ x + c^T x + \pi(Ax - b)$$
$$\frac{\partial \mathcal{L}}{\partial x} = Q\ x + c + A^T\pi = 0$$
$$\frac{\partial \mathcal{L}}{\partial \pi} = Ax - b = 0$$

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \pi \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

$x = -Q^{-1}(A^T\pi + c)$ plug into second equation

$AQ^{-1}(A^T\pi + c) = -b \quad \rightarrow$ solve for π, $\quad \pi^* = -(AQ^{-1}A^T)^{-1}(b + AQ^{-1}c)$

Solving for $x \quad \rightarrow \quad x^* = Q^{-1}(A^T(AQ^{-1}A^T)^{-1}(b + AQ^{-1}c) - c)$

Note: To arrive at this result we have assumed:

1. A has full row rank, which implies
$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$$
is invertible (i.e., non-singular)

2. Q is PD, which ensures x^* is the unique global solution of the problem. If Q is PD (or PSD) then the objective function $\frac{1}{2}x^T Q\ x + c^T x$ is convex. The minimum of a convex function is a global min. The extrema point of a convex function is guaranteed to be a min.

   If Q is PSD, the objective is also convex, and the min is a global min, but uniqueness is not guaranteed.

1.5 Constrained Non-Linear Optimization

How do we characterize optimization problems more generally?

Consider a problem in standard form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & h(x) = 0 \end{aligned}$$

where $x \in \mathbb{R}^n$ and we have m inequality constraints and l equality constraints,

$$g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}, \qquad h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_l(x) \end{bmatrix}$$

1.5.1 KKT Necessary Conditions

The Karush–Kuhn–Tucker (KKT) conditions are the first-order necessary conditions (FONCs) for a solution to a general non-linear program to be optimal. This is a generalization of the Lagrangian method.

Let $\bar{x}$ be a feasible solution of a problem and let $I = \{i : g_i(\bar{x}) = 0\}$ be the index indicating which constraints are active. Also, suppose $\nabla h_j(\bar{x})$ for $j = 1, ..., l$ and $\nabla g_i(\bar{x})$ for $i \in I$ are linearly independent.

If $\bar{x}$ is a local min, then there exists vectors u and v such that we comply with the following conditions:

1. Stationarity: $\nabla f(\bar{x}) + [J_g(\bar{x})]^T u + [J_h(\bar{x})]^T v = \bar{0}$
2. Dual feasibility: $u_i \geq 0$ for $i = 1, ..., m$
3. Complementary slackness: $u_i \cdot g_i(\bar{x}) = 0$ for $i = 1, ..., m$

1.5.2 Correct setup of KKT Condition 1

The first KKT condition is

$$\nabla f(x) + [J_g(x)]^T u + [J_h(x)]^T v = \bar{0}$$

To correctly set up this condition, we should first derive the gradient of $f(x)$ and the Jacobians of $g(x)$ and $h(x)$. Assume we have n decision variables (i.e., $x \in \mathbb{R}^n$), m inequality constraints, and l equality constraints.

$$\nabla f(x) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} \quad \Rightarrow \quad \nabla f(x) \in \mathbb{R}^{n\times 1}$$

$$J_g(x) = \begin{bmatrix} \nabla g_1(x)^T \\ \nabla g_2(x)^T \\ \vdots \\ \nabla g_m(x)^T \end{bmatrix} \quad \Rightarrow \quad [J_g(x)]^T = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_n} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}, \quad [J_g(x)]^T \in \mathbb{R}^{n\times m}$$

$$J_h(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \nabla h_2(x)^T \\ \vdots \\ \nabla h_m(x)^T \end{bmatrix} \quad \Rightarrow \quad [J_h(x)]^T = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_l}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_l}{\partial x_n} \end{bmatrix}, \quad [J_h(x)]^T \in \mathbb{R}^{n\times l}$$

For example, assume we have $n = 3$, $m = 2$, $l = 1$

$$\underbrace{\begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \partial f/\partial x_3 \end{bmatrix}}_{n\times 1} + \underbrace{\underbrace{\begin{bmatrix} \partial g_1/\partial x_1 & \partial g_2/\partial x_1 \\ \partial g_1/\partial x_2 & \partial g_2/\partial x_2 \\ \partial g_1/\partial x_3 & \partial g_3/\partial x_3 \end{bmatrix}}_{n\times m} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}_{n\times 1} + \underbrace{\underbrace{\begin{bmatrix} \partial h_1/\partial x_1 \\ \partial h_1/\partial x_2 \\ \partial h_1/\partial x_3 \end{bmatrix}}_{n\times l} v_1}_{n\times 1} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{n\times 1}$$

1.5.3 KKT Example

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & (x_1 - 4)^4 + (x_2 + 6)^2 \\ \text{s.t.} \quad & x_1^2 + 3x_1 + x_2^2 - 4.5x_2 \leq 6.5 \\ & (x_1 - 9)^2 + x_2^2 \leq 64 \\ & 8x_1 + 4x_2 = 20 \end{aligned}$$

Could $\bar{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ be a local min? Find u and v that satisfy the KKT conditions.

Check primal feasibility:

- $g_1(\bar{x}) = \bar{x}_1^2 + 3\bar{x}_1 + \bar{x}_2^2 - 4.5\bar{x}_2 - 6.5 = 0 \leq 0 \quad \therefore$ this constraint is active
- $g_2(\bar{x}) = (\bar{x}_1 - 9)^2 + \bar{x}_2^2 - 64 = -14 \leq 0 \quad \therefore$ this constraint is inactive
- $h(\bar{x}) = 8\bar{x}_1 + 4\bar{x}_2 - 20 = 0$

Thus, we have shown that $\bar{x}$ is feasible. However, we also found that constraint $g_2(\bar{x})$ is inactive. By complementary slackness, we must have that $u_2 = 0$.

To setup the first KKT condition, we must first find the gradient of $f(x)$ and the transpose of the Jacobians of $g(x)$ and $h(x)$

$$\nabla f(x) = \begin{bmatrix} 4(\bar{x}_1 - 4)^3 \\ 2\bar{x}_2 + 12 \end{bmatrix} = \begin{bmatrix} -32 \\ 14 \end{bmatrix}$$

$$[J_g(x)]^T = \begin{bmatrix} 2\bar{x}_1 + 3 & 2\bar{x}_1 - 18 \\ 2\bar{x}_2 - 4.5 & 2\bar{x}_2 \end{bmatrix} = \begin{bmatrix} 7 & -14 \\ -2.5 & 2 \end{bmatrix}$$

$$[J_h(x)]^T = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

And now we have that

$$\begin{bmatrix} -32 \\ 14 \end{bmatrix} + \begin{bmatrix} 7 & -14 \\ -2.5 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 8 \\ 4 \end{bmatrix} v = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Since we already determined that $u_2 = 0$, we have the following system of equations

$$-32 + 7u_1 + 8v = 0$$

$$14 - 2.5u_1 + 4v = 0$$

Solving this system, we find that $u_1 = 5$ and $v = -0.375$, which satisfies the dual feasibility condition $u_i \geq 0$ for $i = 1, ..., m$.

Therefore, we have determined that $\bar{x}$ may be a local min. The KKT conditions are just the First Order Necessary Condition (FONC), but they are not sufficient.

- To show $\bar{x}$ is a local minimum, we must show that the objective function and the active constraints are convex at $\bar{x}$.
- To show $\bar{x}$ is a global minimum, we must show that the objective function and the active constraints are convex at over the entire feasible set.

Convexity check

Check the convexity of the objective $f(x)$ by inspecting its Hessian,

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 12(x_1 - 4)^2 & 0 \\ 0 & 2 \end{bmatrix}$$

and, since $12(x_1 - 4)^2 \geq 0$ for $x_1 \in \mathbb{R}$, we have that the Hessian is PSD for all x.

We also have that the equality constraint $h_1(x)$ is linear, so, by definition, it is convex for all x. The second inequality constraint $g_2(x)$ is inactive, so we can ignore it.

Finally, we must check the convexity of the active inequality constraint

$$H_{g_1}(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

which means $g_1(x)$ is also convex for all $x \in \mathbb{R}^2$. Therefore, we can conclude that $\bar{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is not only a local minimum, but it is also a global minimum since we have a convex problem.

1.5.4 KKT conditions for a standard QP

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}x^T Q\, x + c^T x && \}\ f(x) \\ \text{s.t.} \quad & Ax = b && \}\ h(x) = Ax - b = 0 \\ & x \geq 0 && \}\ g(x) = -x \leq 0 \end{aligned}$$

1. $Q\,x + c \quad - I\,u \quad + A^T v = 0$
2. $u = [u_1\ ...\ u_n]^T \geq 0$
3. $u_i \cdot (-x_i) = 0$ for $i = 1, ..., n$

Note that the number of inequality constraints is the same as the number of decision variables, $m = n$.