

Non-Linear Programming

Basics of Mathematical Programming and unconstrained optimization

Anand Jacob Abraham

Department of Industrial & Systems Engineering
IIT Kharagpur

Table of contents

1. Basics of mathematical programming
2. Feasible directions and FONC for Maxima
3. FONC and SONC for set constrained maximizers
4. FONC and SOSC for interior point/set constrained maximizers

Basics of mathematical programming

Goal of optimization

A (single objective) optimization is concerned with computing

$$\max_{\mathbf{x} \in S} f(\mathbf{x}).$$

The solution (values of \mathbf{x} which accomplish this are called maxima¹ (minima) or maximizers (minimizers). The function value at the maxima is called the *maximum value*.

- Generally computing the maxima is not computationally easy. Even by exploiting some characteristics.
- Often we settle for local maxima. Solutions that perform well in a certain vicinity and not everywhere. They are easier to compute/

¹This is the global optima or global optimal solution

Local maxima

$\mathbf{x}^* \in S \subseteq \mathbb{R}^n$ is a local optima for the problem $\max_{\mathbf{x} \in S} f(\mathbf{x})$ if $f(\mathbf{x}^*) \geq f(\mathbf{x}) \forall \mathbf{x} : \|\mathbf{x}^* - \mathbf{x}\| < \varepsilon$ for some $\varepsilon > 0$.

- There are computationally efficient methods to find the local maxima.
- For special optimization problem classes like Concave Maximization Problems (CMP) local and global optima are the same.
- All global maxima are also local maxima.
- Local maxima enjoy special properties.

Mathematical basics: Jacobian

Let $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$. The function is a vector function with m

components $\mathbf{f}: \mathbb{R}^n \mapsto \mathbb{R}^m$. The derivative matrix or Jacobian for \mathbf{f} is given by

$$D\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}.$$

For a univariate function, the transpose of the derivative matrix is called the gradient vector.

Mathematical basics: Gradient (contd.)

For the transpose of the derivative matrix is called the gradient vector $\nabla f(\mathbf{x})$. Suppose the function is a scalar function instead of a vector function.

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_m}(\mathbf{x}) \end{pmatrix}.$$

Mathematical basics: Hessian matrix (contd.)

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable. The Hessian matrix (\mathbf{H} or $\nabla^2 f(x)$) of f at $x \in \mathbb{R}^n$ is defined as

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

Multivariate Taylor Series

Given $f: \mathbb{R}^n \mapsto \mathbb{R}$ and its derivatives at a point \mathbf{x} , we can approximate the function value at another point $\mathbf{x} + \mathbf{h}$ as a polynomial below.

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}) + \frac{1}{2} \mathbf{h}^T \mathbf{H} \mathbf{h} + o(\|\mathbf{h}\|^2)$$

- This approximation works so long as the function is smooth.
- The last few terms are jointly represented as $o(\|\mathbf{h}\|^2)$. They go to zero faster than $\|\mathbf{h}\|^2$.
 - You could ignore these terms if $\|\mathbf{h}\|^2 \approx 0$
 - However ignoring some these terms means that the polynomial series becomes an approximation which works well in the vicinity of \mathbf{x} .

Vectors, Matrices, and Negative Semi-definiteness

A vector is a member of the real vector space, i.e., $\mathbf{v} \in \mathbb{R}^n$. A matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ is a row vector of several column vector elements. i.e.,

$$\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_m]$$

Negative (Positive) semi-definite matrix

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is negative (positive) semi-definite if $\mathbf{v}^T \mathbf{A} \mathbf{v} \leq (\geq) 0 \forall \mathbf{v} \in \mathbb{R}^n$.

- In case the inequality is strict, we use the term negative (positive) definite and avoid the usage of term phrase semi.
- In this course, we are more concerned with negative semi-definiteness and negative definiteness, since these are associated with properties enjoyed by maximizers (as we shall see subsequently).

Alternate definition for negative definiteness

A matrix \mathbf{A} is negative definite if and only if all its Eigen values are negative. In case of semi-definite the Eigen values should be of \leq type.

Sylvester's criterion for Negative definiteness

Sylvester's criterion

A symmetric square matrix is a negative definite if its leading principal minors $\Delta_1, \Delta_2, \dots, \Delta_n$ are such that $(-1)^i \Delta_i > 0$. That is the leading principal minors have alternating signs with the first principal minor being negative.

- $A \succ 0$ if $\Delta_i > 0 \forall i = 1, 2, \dots, n$.
- $A \prec 0$ If $(-1)^i \Delta_i > 0 \forall i = 1, 2, \dots, n$.
- Indefinite if the first minor that breaks the pattern is of opposite sign.
- Inconclusive if the first minor which breaks the pattern is 0. In this case, apply the definition to check for definiteness (semi-definiteness).

Directional derivative

Let $f: \mathbb{R}^n \mapsto \mathbb{R}$ and $\mathbf{d} \in \mathbb{R}^n$ be an arbitrary direction of unit magnitude $\|\mathbf{d}\| = 1$.

$$\frac{\partial f}{\partial \mathbf{d}} = \lim_{\alpha \rightarrow 0} \left(\frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha} \right)$$

Directional derivative quantifies the level of change in a function if a unit distance is moved in a particular direction.

Directional derivative: A simpler equation

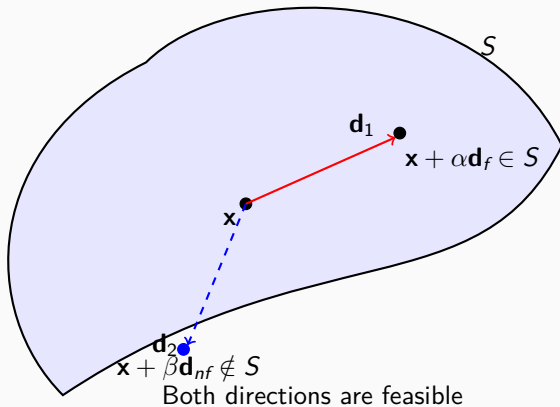
$$\begin{aligned}\frac{\partial f}{\partial \mathbf{d}} &= \lim_{\alpha \rightarrow 0} \left(\frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha} \right) \\ &= \left(\frac{\nabla f(\mathbf{x} + \alpha \mathbf{d})^T \mathbf{d}}{1} \right)_{\alpha=0} \\ \therefore \frac{\partial f}{\partial \mathbf{d}} &= \mathbf{d}^T \nabla f(\mathbf{x})\end{aligned}$$

Feasible directions and FONC for Maxima

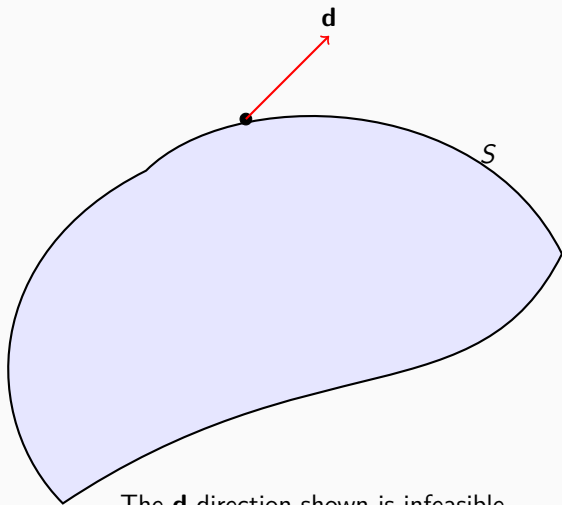
Feasible directions

Consider the maximization problem $\max_{\mathbf{x} \in S} f(\mathbf{x})$. Then a feasible direction $\mathbf{d} : \|\mathbf{d}\| = 1$ is unit vector such that $\exists \alpha > 0 : \mathbf{x} + \alpha \mathbf{d} \in S$.

Let \mathcal{D} denote set of feasible directions.



Feasible direction (contd.)



The d direction shown is infeasible

FONC and SONC for set constrained maximizers

Lemma

If \mathbf{x}^ is a local maximizer, then $\mathbf{d}^T \nabla f(\mathbf{x}^*) \leq 0$ for all feasible directions.*

Let \mathbf{d} be some feasible direction at the local maximizer \mathbf{x}^* .

$$f(\mathbf{x} + \alpha \mathbf{d}) = f(\mathbf{x}) + \alpha \mathbf{d}^T \nabla f(\mathbf{x}) + o(\|\alpha \mathbf{d}\|)$$

At $\alpha \rightarrow 0$, we have $f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x}) = \alpha \mathbf{d}^T \nabla f(\mathbf{x})$. Since \mathbf{x}^* is a local maximizer, this difference must not be positive. Thus, $\mathbf{d}^T \nabla f(\mathbf{x}^*) \leq 0$ \square

SONC for local maximizer

Lemma

If \mathbf{x}^ is a local maximizer, and for some $\mathbf{d} \in \mathcal{D}$, $\mathbf{d}^T \nabla f(\mathbf{x}) = 0$, then $\mathbf{d}^T \mathbf{H} \mathbf{d} \leq 0$ for all feasible directions.*

Proof by contradiction. Assume this claim is false. i.e., let \mathbf{d} be a feasible direction such that $\mathbf{d}^T \nabla f(\mathbf{x}) = 0$ and \mathbf{x}^* is a local maximizer. However, assume that $\mathbf{d}^T \mathbf{H} \mathbf{d} > 0$.

From Multivariate Taylor series up to two terms we have,

$$f(\mathbf{x}^* + \alpha \mathbf{d}) = f(\mathbf{x}^*) + \alpha \mathbf{d}^T \nabla f(\mathbf{x}^*) + \frac{\alpha^2}{2} \mathbf{d}^T \mathbf{H} \mathbf{d} + o(\|\alpha \mathbf{d}\|^2)$$

$$\alpha \rightarrow 0, \mathbf{d}^T \nabla f(\mathbf{x}^*) = 0 \implies f(\mathbf{x}^* + \alpha \mathbf{d}) - f(\mathbf{x}^*) = \frac{\alpha^2}{2} \mathbf{d}^T \mathbf{H} \mathbf{d} > 0$$

This implies that \mathbf{x}^* cannot be optimal. Therefore, optimality of \mathbf{x}^* implies that for $\mathbf{d} : \mathbf{d}^T \nabla f(\mathbf{x}^*) = 0$, $\mathbf{d}^T \mathbf{H} \mathbf{d} \leq 0$ must hold. □

FONC and SOSC for interior point/set constrained maximizers

FONC for unconstrained maximizers and interior points

Lemma (FONC for interior /unconstrained optima)

For an interior point maximizer \mathbf{x}^ , $\nabla f(\mathbf{x}^*) = \mathbf{0}$*

$\forall \mathbf{d} \in \mathbf{R}^n$, $-\mathbf{d}$ is also feasible.

$$\mathbf{d}^T \nabla f(\mathbf{x}^*) = -\mathbf{d}^T \nabla f(\mathbf{x}^*) = 0$$

$$\implies \nabla f(\mathbf{x}^*) = \mathbf{0}$$



SOSC for interior point/ unconstrained optimizer

Lemma (SOSC for interior point / unconstrained optima)

If \mathbf{x}^* is an interior point and

- $\nabla f(\mathbf{x}^*) = \mathbf{0}$
- \mathbf{H} is negative definite

Then, \mathbf{x}^* is a strict local maximizer of f .

Given that $\nabla f(\mathbf{x}^*) = \mathbf{0}$, we perform a Taylor series expansion around \mathbf{x}^* along any direction $\mathbf{d} \in \mathbb{R}^n$ by moving $\alpha \rightarrow 0$ units.

$$f(\mathbf{x}^* + \alpha \mathbf{d}) - f(\mathbf{x}^*) = \frac{\alpha^2}{2} \mathbf{d}^T \mathbf{H} \mathbf{d} + o(\|\alpha \mathbf{d}\|^2) < 0 \forall \mathbf{d} \in \mathbb{R}^n$$

Therefore \mathbf{x}^* is a local optimizer.

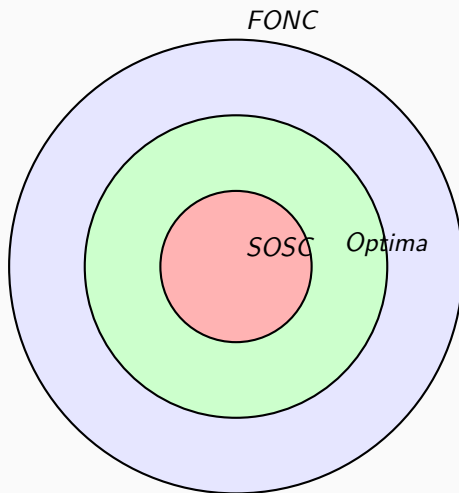


Figure 1: Necessary and sufficient conditions

Summary

- All set constrained maximizers² need to satisfy non-positivity of directional derivative in all feasible directions. This is the general FONC.
 - For interior point maxima or unconstrained maxima, this condition simplifies to $\nabla f(\mathbf{x}^*) = \mathbf{0}$.
- In addition to FONC, every maximizer must satisfy the SONC.
 - The SONC for set constrained maximizers says that if there is a zero directional derivative in any direction \mathbf{d} , then the $\mathbf{d}^T \mathbf{H} \mathbf{d} \leq 0$ along that direction when evaluated at that point.
 - The SONC for interior local maximizers says that the Hessian matrix should be negative semi definite at that maximizer point.
- There is a SOSC for interior point maximizers which says that the point should be a stationary point (gradient=0) and Hessian matrix should be negative definite.
- The FONC, SONC, and SOSC for interior point maximizers are applicable for unconstrained optimization problems.

²We mean local maximizers throughout

Example problems

1. Find the value of $x \in \mathbb{R}$ that minimizes the function

$$f(x) = x^2 - 6x + 13.$$

2. Find the value of $x \in \mathbb{R}$ that maximizes the function

$$f(x) = xe^{-x}.$$

3. Find the values of $(x, y) \in \mathbb{R}^2$ that minimize the function

$$f(x, y) = 3x^2 + 2y^2 - 4x - 8y + 10.$$

4. Find the values of $(x, y) \in \mathbb{R}^2$ that minimize the function

$$f(x, y) = x^2 + xy + y^2 - 6x - 4y.$$

5. Find all critical points and classify them for the function

$$f(x, y) = x^3 - 3xy^2.$$

6. Find the values of $(x, y) \in \mathbb{R}^2$ that maximize the function

$$f(x, y) = xye^{-(x^2+y^2)}.$$

7. Find the values of $(x, y) \in \mathbb{R}^2$ that maximize the function

$$f(x, y) = \ln(1 + x^2) + \ln(1 + y^2) - (x^2 + y^2).$$

8. Find the values of $(x_1, x_2, x_3) \in \mathbb{R}^3$ that minimize the function

$$f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 3x_3^2 - 4x_1 - 8x_2 + 6x_3.$$

9. Find the values of $(x, y) \in \mathbb{R}^2$ that minimize the function

$$f(x, y) = (x - y^2)^2 + (1 - x)^2.$$

10. Find the value of $q \in \mathbb{R}$ that maximizes the profit function

$$\pi(q) = 50q - 2q^2 - 10 \ln(1 + q).$$