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

Basics of mathematical programming

Goal of optimization

A (single objective) optimization is concerned with computing

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

The solution (values of $\boldsymbol{x}$ which accomplish this are called maxima[1] (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/

[1] This is the global optima or global otpimal solution

Local maxima

Local maxima

$\boldsymbol{x^*} \in S \subseteq \mathbb{R}^n$ is a local optima for the problem $\max\limits_{\boldsymbol{x} \in S} f(\boldsymbol{x})$ if $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}) \forall \boldsymbol{x} : ||\boldsymbol{x}^* - \boldsymbol{x}|| < \varepsilon$ 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 $\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$. The function is a vector function with m components $\boldsymbol{f} : \mathbb{R}^n \mapsto \mathbb{R}^m$. The derivative matrix or Jacobian for $\boldsymbol{f}$ is given by

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

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 $\nabla f(\boldsymbol{x})$. Suppose the function is a scalar function instead of a vector function.

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

Mathematical basics: Hessian matrix (contd.)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. The Hessian matrix ($\boldsymbol{H}$ 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 $\boldsymbol{x}$, we can approximate the function value at another point $\boldsymbol{x} + \boldsymbol{h}$ as a polynomial below.

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

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

Vectors, Matrices, and Negative Semi-definiteness

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

$$\boldsymbol{A} = [\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \ldots \boldsymbol{v}_m]$$

Negative (Positive) semi-definite matrix

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

- 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 $\boldsymbol{A}$ 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.

Sylvestor's criterion for Negative definiteness

Sylvestor's criterion

A symmetric square matrix is a negative definite if its leading principal minors $\Delta_1, \Delta_2, \ldots, \Delta_n$ 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 $\Delta_i > 0 \forall i = 1, 2, \ldots, n$.
- $A \prec 0$ If $(-1)^i\Delta_i > 0 \forall i = 1, 2, \ldots, n$.
- 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 $\boldsymbol{d} \in \mathbb{R}^n$ be an arbitrary direction of unit magnitude $||\boldsymbol{d}|| = 1$.

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

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

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

$$= \left(\frac{\nabla f(\boldsymbol{x} + \alpha \boldsymbol{d})^T \boldsymbol{d}}{1} \right)_{\alpha = 0}$$

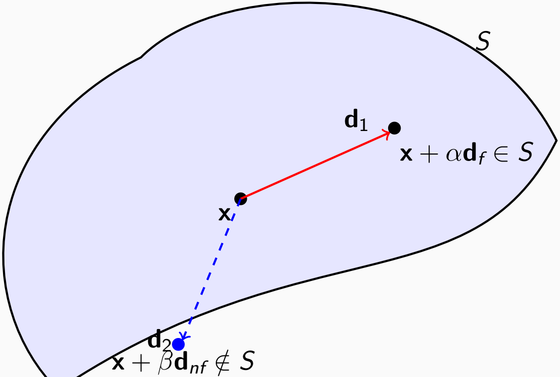
$$\therefore \frac{\partial f}{\partial \boldsymbol{d}} = \boldsymbol{d}^T \nabla f(\boldsymbol{x})$$

Feasible directions and FONC for Maxima

Feasible directions

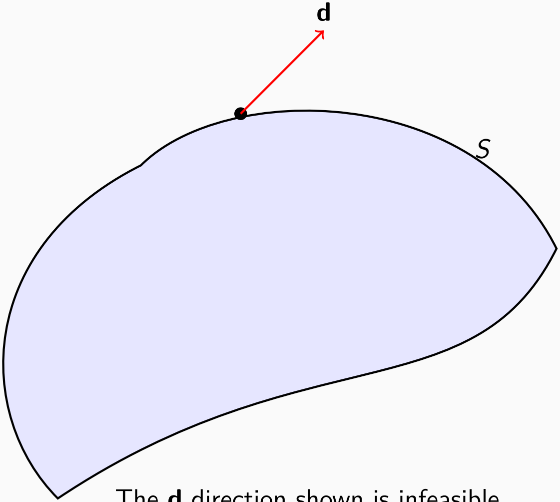
Consider the maximization problem $\max\limits_{\boldsymbol{x} \in S} f(\boldsymbol{x})$. Then a feasible direction $\boldsymbol{d} : ||\boldsymbol{d}|| = 1$ is unit vector such that $\exists \alpha > 0 : \boldsymbol{x} + \alpha \boldsymbol{d} \in S$.

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



Both directions are feasible

Feasible direction (contd.)



The **d** direction shown is infeasible

FONC for local maximizer

Lemma

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

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

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

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

SONC for local maximizer

Lemma

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

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

From Multivariate Taylor series up to two terms we have,

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

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

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

FONC for uncaonstrained maximizers and interior points

Lemma (FONC for interior /unconstrained optima)

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

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

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

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

$\square$

SOSC for interior point/ unconstrained optimizer

Lemma (SOSC fo interior point / unconstrained optima)

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

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

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

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

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

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

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