Introduction to Neural Networks

Johns Hopkins University
Engineering for Professionals Program
605-447.71/625-438.71
Dr. Mark Fleischer

Module 2.3: Mathematical Review-Calculus Based Optimization

This Sub-Module Covers …

- Calculus-based optimization methods and related material:
  - First order necessary conditions.
  - Second order sufficiency conditions.
  - Definition of convexity.
- Sets the stage for further mathematical review by exploring Metric Spaces in the next sub-module.

First-Order Necessary Conditions

TFAE

$$\frac{d f(x^*)}{dx} = 0$$

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

$$\forall \mathbf{d}, \quad \nabla f(\mathbf{x}^*) \cdot \mathbf{d} = 0$$

There exists a $t' > 0$, such that for all t, where $0 < t < t'$, and for all non-zero vectors $\mathbf{d}$

$$(\exists t' > 0, \ni \forall 0 < t < t' \wedge \mathbf{d} > \mathbf{0}),$$

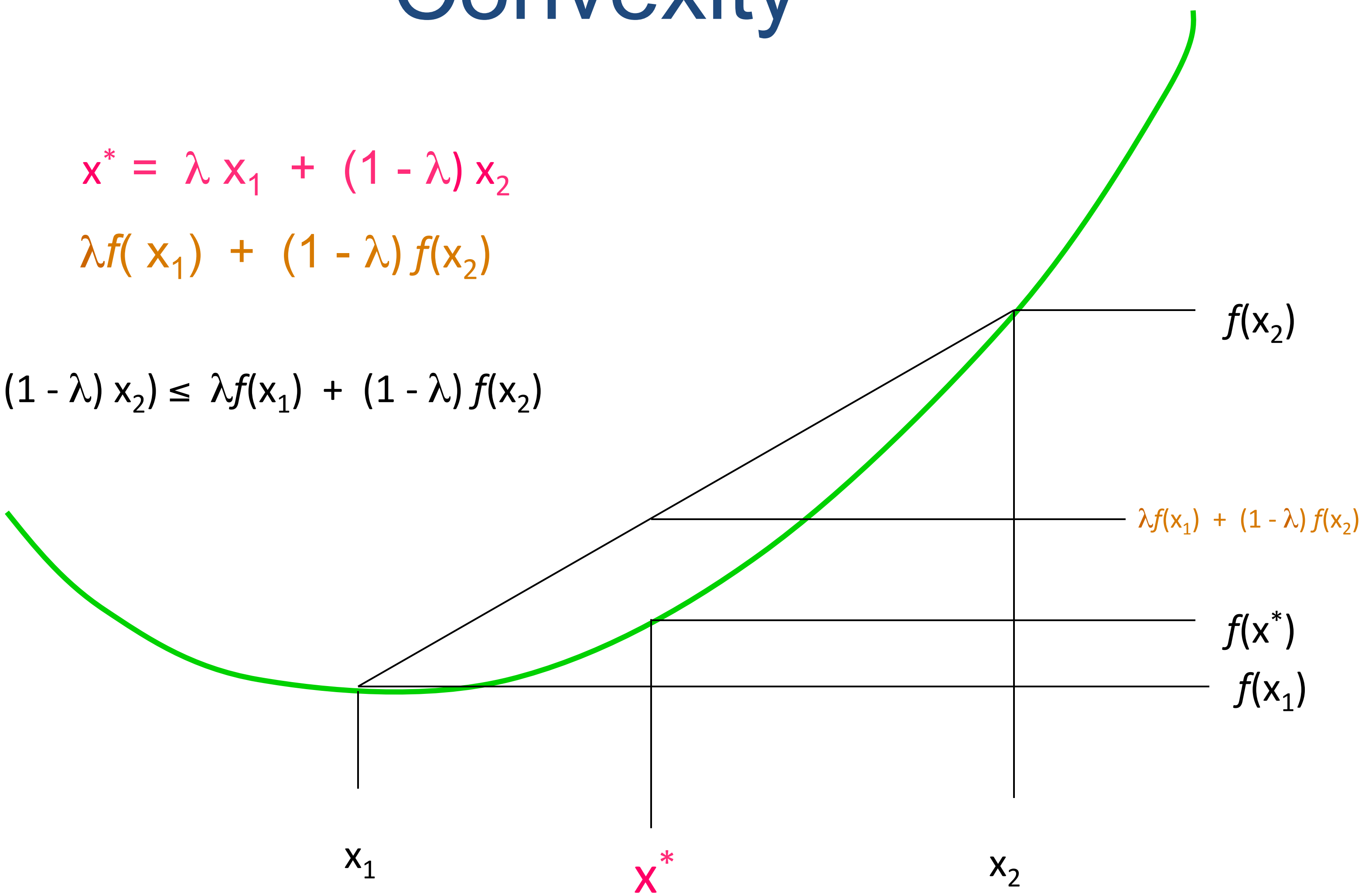
$$f(\mathbf{x}^*) < f(\mathbf{x}^* + t\mathbf{d})$$

Convexity

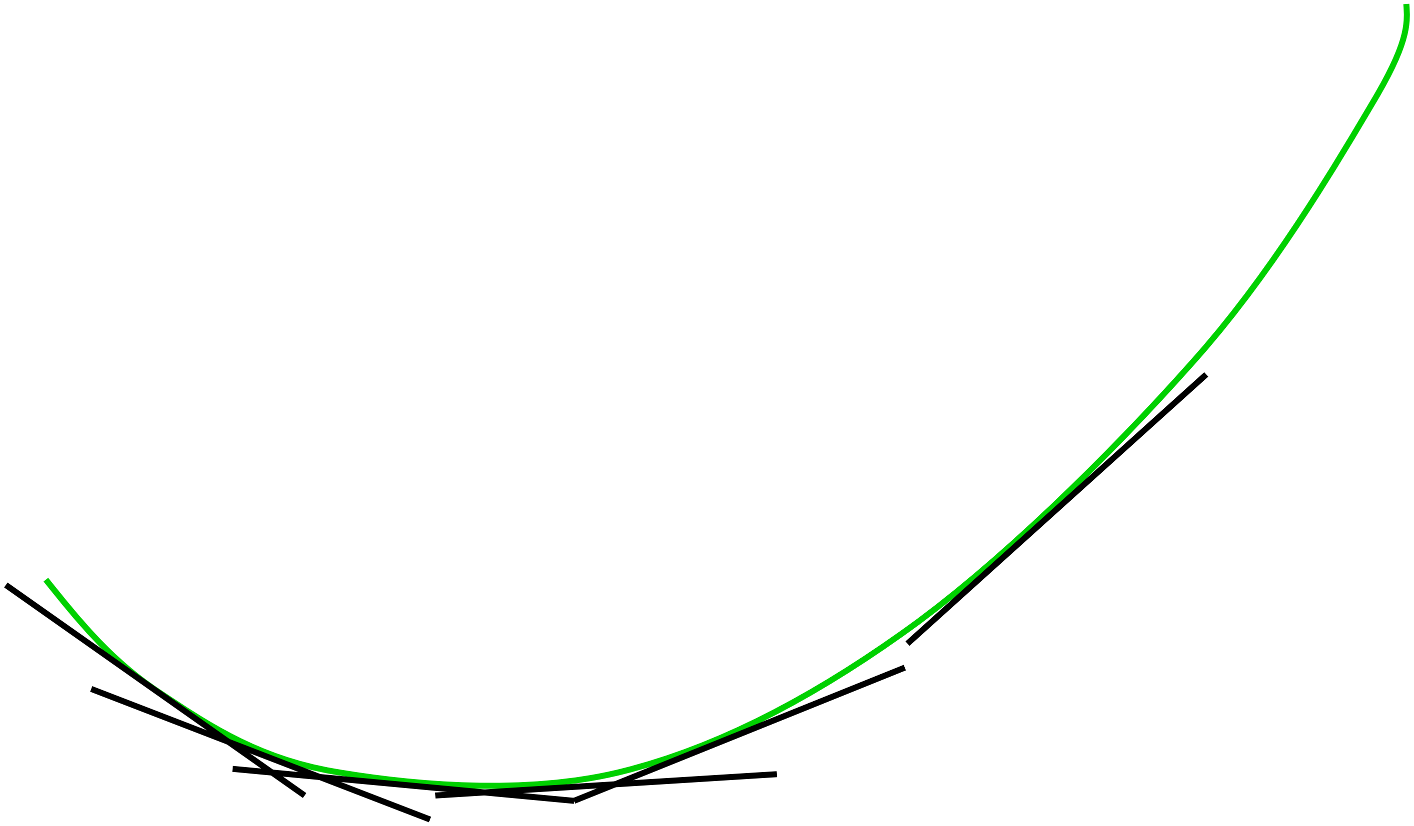
$x^* = \lambda x_1 + (1 - \lambda) x_2$

$\lambda f(x_1) + (1 - \lambda) f(x_2)$

$f(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)$



Second Derivatives



Second-Order Sufficiency Conditions

TFAE (for determining minima)

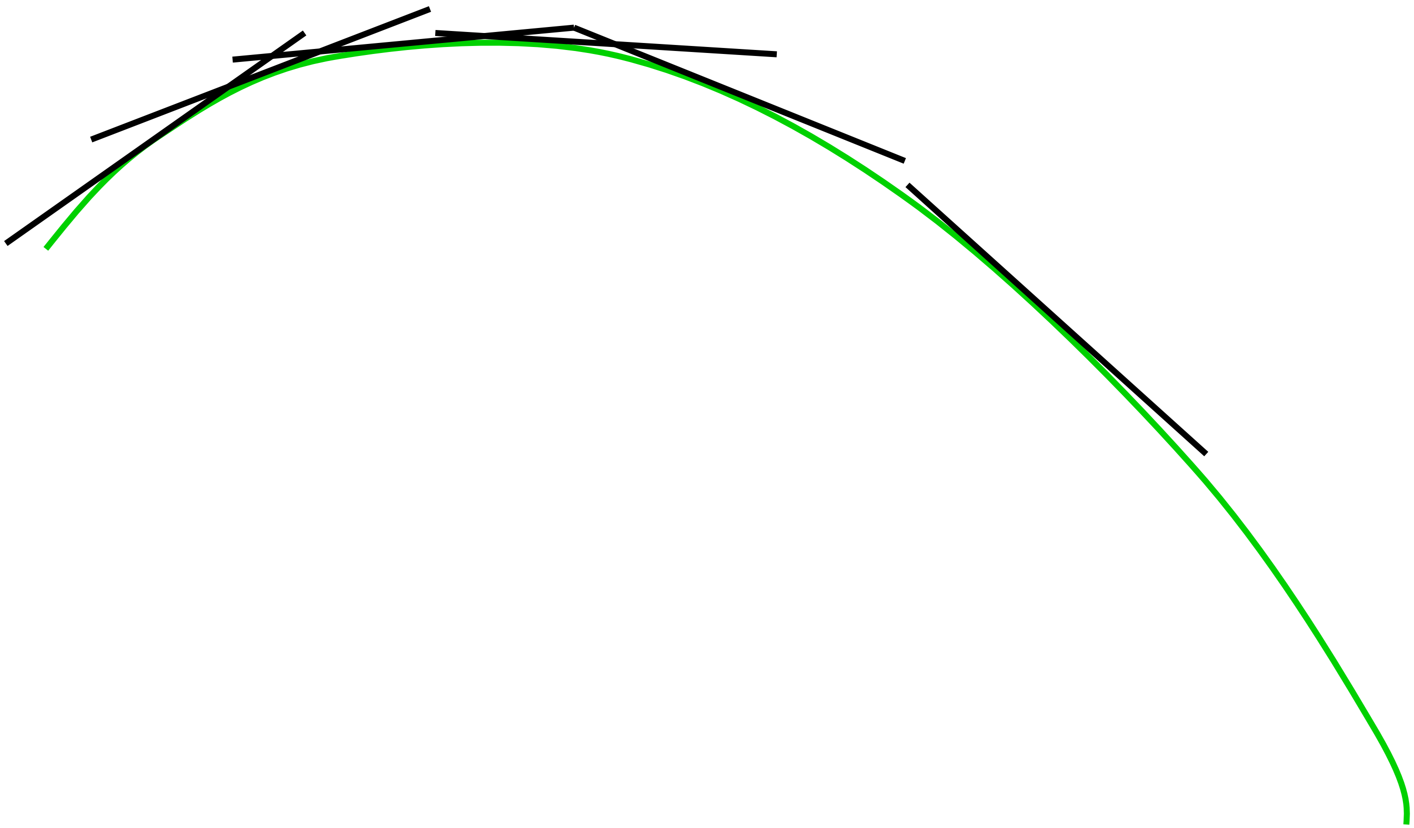
1. All second derivatives are positive
2. All points in a tangent plane (or hyperplane) have function values less than or equal to the objective function value.
3. The function is convex.

Item 1:

$$d^2y/dx^2 = f''(x^*) > 0$$

The Hessian Matrix $\mathbf{H}(\mathbf{x})$ is positive definite,
i.e., for all R^n, $\mathbf{x}^T\mathbf{H}\mathbf{x} \geq 0$:

Second Derivatives



Second-Order Sufficiency Conditions
TFAE (for determining minima)

1. All second derivatives are positive
2. All points in a tangent plane (or hyperplane) have function values less than or equal to the objective function value.
3. The objective function is convex.

Item 2:

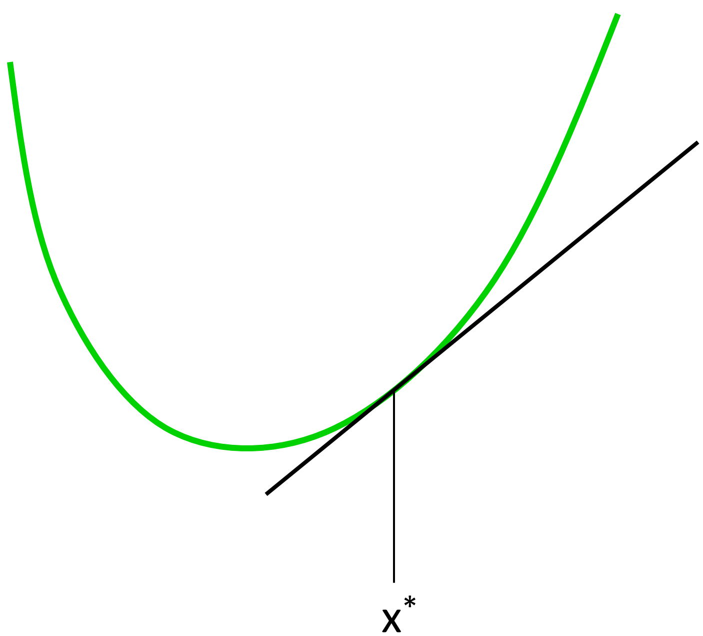
2. $\forall \mathbf{x}, \mathbf{x}^* \in R^n$,

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*).$$

Item 2: Tangent Plane

2. $\forall \mathbf{x}, \mathbf{x}^* \in R^n$,

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*).$$



$$f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$$

$= f(x^*) + m(x - x^*)$
$= f(x^*) + mx - mx^*$
$= mx + [f(x^*) - mx^*]$
$= mx + b$

Second-Order Sufficiency Conditions
TFAE (for determining minima)

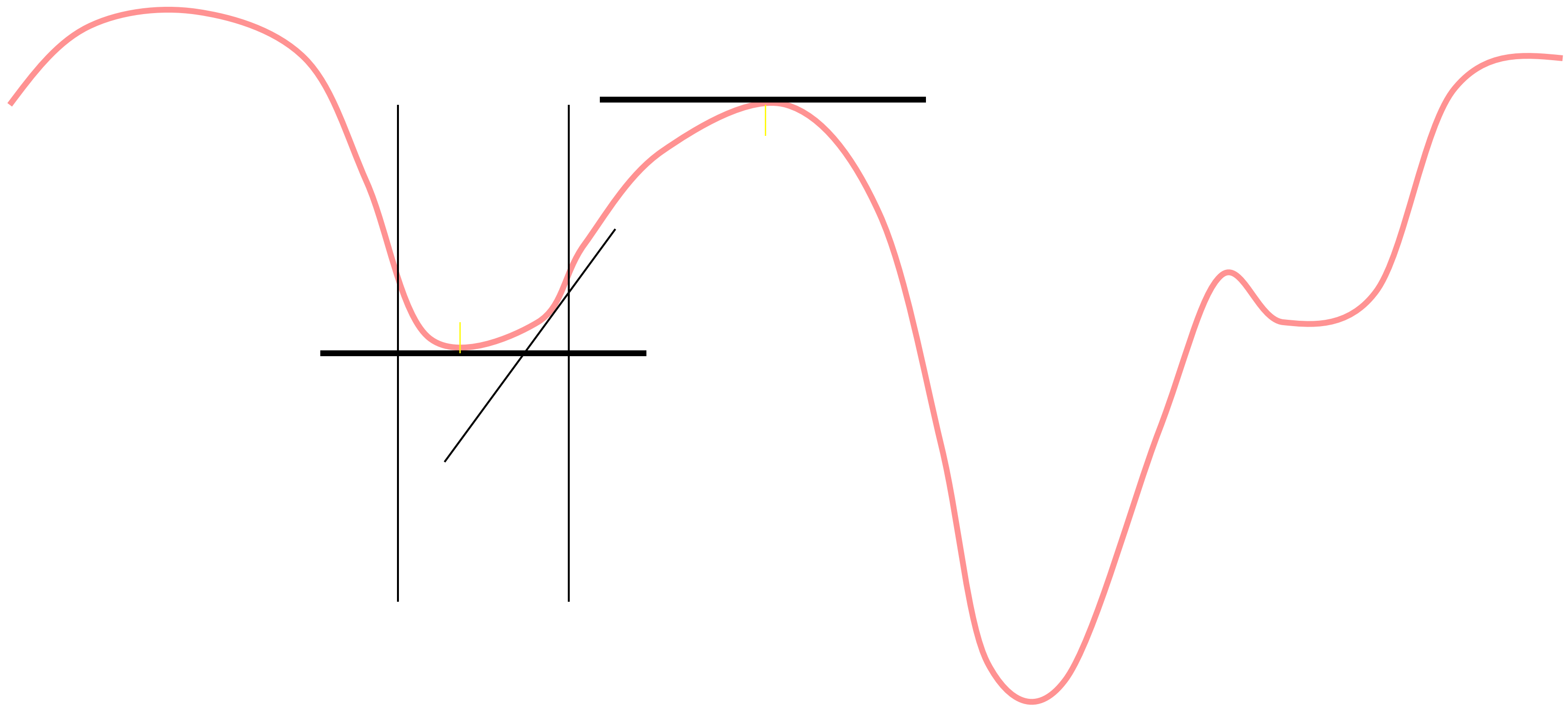
1. All second derivatives are positive
2. All points in a tangent plane (or hyperplane) have function values less than or equal to the objective function value.
3. The objective function is convex.

Item 3:

3. $\forall \mathbf{x}, \mathbf{x}^* \in R^n$ and $0 \leq \lambda \leq 1$,

$$\lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{x}^*) \geq f(\lambda \mathbf{x} + (1-\lambda)\mathbf{x}^*).$$

Calculus Based Optimization



Example

Suppose we want to minimize the function:

$$f(x_1, x_2) = x_1 x_2 - 6x_1 - 3x_2 + 18$$

Taking all partial derivatives, we see that:

$$\frac{\partial f}{\partial x_1} = x_2 - 6 = 0$$

$$\frac{\partial f}{\partial x_2} = x_1 - 3 = 0$$

$$x_1 = 3; x_2 = 6$$

Example

To show that it is a minimum, it is *sufficient* to show that one of the three second-order sufficiency conditions holds. Using the condition in item 2, we see that the equation for the tangent plane at say point (4,3) is:

Remembering that

$$\frac{\partial f}{\partial x_1} = x_2 - 6 = 0$$

$$\frac{\partial f}{\partial x_2} = x_1 - 3 = 0$$

$$\begin{aligned}
f_P(x_1, x_2) &= f(4,3) + \nabla f(4,3) \begin{pmatrix} x_1 - 4 \\ x_2 - 3 \end{pmatrix} \\
&= -3 + (-3, 1) \begin{pmatrix} x_1 - 4 \\ x_2 - 3 \end{pmatrix} \\
&= -3 - 3(x_1 - 4) + 1(x_2 - 3) \\
&= -3x_1 + x_2 + 6
\end{aligned}$$

Example

Now consider any point (x_1, x_2) and compare the value of f_P with the objective function value *f*. For example, at point (0,0) the value of f_p = 6. For the objective function, *f*(0,0) = 18. Since the relationship between the objective function and tangent plane holds as item 2 above, it *suggests* that the second-order conditions hold. To establish this however requires that we prove this relation holds for *all* points (x_1, x_2).
Can you prove that they do?

Remembering that $f(x_1, x_2) = x_1 x_2 - 6x_1 - 3x_2 + 18$

$$
\begin{aligned}
f_P(x_1, x_2) &= f(4,3) + \nabla f(4,3) \begin{pmatrix} x_1 - 4 \\ x_2 - 3 \end{pmatrix} \\
&= -3 + (-3, 1) \begin{pmatrix} x_1 - 4 \\ x_2 - 3 \end{pmatrix} \\
&= -3 - 3(x_1 - 4) + 1(x_2 - 3) \\
&= -3x_1 + x_2 + 6
\end{aligned}
$$

Taylor's Theorem

Single variable case

$$
\begin{aligned}
f(x) &= a_0 + a_1(x - x_0) + a_2(x - x_0)^2 \ldots \\
&= \boxed{f(x_0) + (x - x_0)f'(x_0)} + \frac{1}{2!}(x - x_0)^2 f''(x_0) \ldots
\end{aligned}
$$

Multi - variable case

$$
\begin{aligned}
f(\mathrm{x}) &= a_0 + a_1(\mathrm{x} - \mathrm{x}_0) + a_2(\mathrm{x} - \mathrm{x}_0)^2 \ldots \\
&= \boxed{f(\mathrm{x}_0) + (\mathrm{x} - \mathrm{x}_0)\nabla f(\mathrm{x}_0)} + \frac{1}{2!}(\mathrm{x} - \mathrm{x}_0)^T \mathbf{H}(\mathrm{x} - \mathrm{x}_0) \ldots
\end{aligned}
$$

Mathematical Review

So far we've reviewed:

- Basic Vector/Matrix operations: inner, outer products
- Linear Independence
- Differential calculus
- Partial Differentiation,
- Directional Derivatives $\nabla f(x,y,z)\cdot\mathbf{d} = \|\nabla f(x,y,z)\| \times \|\mathbf{d}\| \times \cos\theta$
- Gradient vector $\nabla f(\mathbf{x}) = (\partial f(\mathbf{x})/\partial x_1, \partial f(\mathbf{x})/\partial x_2, \ldots, \partial f(\mathbf{x})/\partial x_n)$
- First order necessary conditions, Second order sufficiency conditions

In the next sub-modules we will review:

- Metric Spaces:
  - Distance and Magnitude of vectors, matrices
  - Definitional requirements of "norms"
    - **Positivity, homogeneity, Triangle Inequality**