

INTRODUCTION TO NONLINEAR OPTIMIZATION¹

Nonlinear optimization is concerned with methods for finding the minimum or maximum value of nonlinear function $F(\mathbf{x})$ of any number of independent variables x_1, x_2, \dots, x_n . The basic problem is to find vector \mathbf{x}^* that minimizes $F(\mathbf{x})$ where $F(\mathbf{x})$ is called the *Objective Function* or *Performance Index*. The variables x_1, x_2, \dots, x_n are called *Decision Variables*. The value $F(\mathbf{x}^*)$ is called *minimum* of the problem while \mathbf{x}^* is called a *minimizer*.

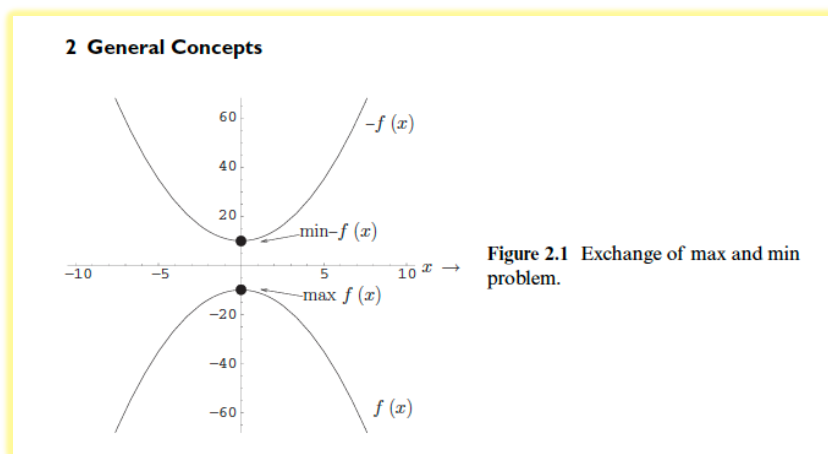
Unconstrained optimization: there are no special conditions that the independent variables are allowed to be subject to.

Constrained optimization: there are special conditions that the independent variables are allowed to be subject to:

- *Equality constraints*: $c_i(\mathbf{x})=0, i=1,2,\dots,m$
- *Inequality constraints*: $c_i(\mathbf{x})\geq 0, i=1,2,\dots,k$

Note: The set of all minimizers of $F(\mathbf{x})$ is denoted by $\text{argmin } F(\mathbf{x})$ with $\mathbf{x} \in \mathbf{S}$. Note: *Feasible region* or *feasible set* is a subset \mathbf{S} of the set of real numbers, \mathbf{R} .

Note. The maximization of $F(\mathbf{x})$ =minimization ($-F(\mathbf{x})$):



¹ Scales, L.E., *Introduction to Non-linear Optimization*, Springer-Verlag, New York, NY, 1985.

¹ Gill, P.E., W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, UK, 1981.

¹ Edgar, T.F. and D.M. Himmelblau, *Optimization of Chemical Processes*, 2nd ed McGraw-Hill, New York, 2001

¹ Rao, S.S., *Engineering Optimization*, 5th ed, Wiley, Hoboken, NJ, 2020.

¹Theodore, L., K. Behan, *Introduction to Optimization for Chemical and Environmental Engineers*, CRC Press, Boca Raton, FL, 2018.

E-HJ

First, we discuss Unconstrained optimization (minimization). In this case $S=R$. For example, consider the objective $F(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 3)^2$ defined $\forall x \in R$. Then $\min F(x) = 0$ at $x_1^* = 1$ and $x_2^* = 3$ or $\text{argmin } F(x) = (1, 3)^T$.

Solution methods. There are two categories of methods to solve the problem:

1. *Gradient methods* (require derivatives of the objective functions). It is assumed that the objective function has continuous second derivatives, whether or not these are explicitly available. *Gradient methods* are still efficient if there are some discontinuities in the derivatives.
2. *Direct search methods* (derivative-free). *These methods* use function values and are more efficient for highly discontinuous functions.

Applications: Nonlinear Parameter Estimation or Nonlinear Regression.

Suppose that we have a mathematical model describing of physical / chemical / biological process. A special class of optimization problems arises when the objective function is a suitable measure of the overall departure of the model calculated values from the experimental measurements obtained from the process. The objective function is of the form²

$$S(\mathbf{x}) = \sum_{i=1}^N \mathbf{e}_i^T \mathbf{Q}_i \mathbf{e}_i$$

This is the *parameter estimation* or *nonlinear regression problem*. Vector \mathbf{e}_i is the vector of residuals from the i^{th} experiment and \mathbf{Q}_i is a user supplied square weighting matrix. $\mathbf{e}_i = \hat{y}_i - f(\mathbf{x}_i, \mathbf{k})$. If $\mathbf{Q} = \mathbf{I}$ then we minimize the sum of squares of errors (SSE) and we have the *least squares* (LS) estimation problem. The objective function is given by

$$S(\mathbf{x}) = \sum_{i=1}^N \mathbf{e}_i^T \mathbf{e}_i$$

The mathematical model describing the process can be linear (w.r.t. parameters) or nonlinear algebraic equations, ordinary differential equation (ODE), or Partial Differential Equation (PDEs).

² Englezos, P., and N.E. Kalogerakis, *Applied Parameter Estimation for Chemical Engineers*, Marcel-Dekker, 2001

BASIC CONCEPTS (Matrices, Vectors,

From EH

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \quad (n \times m) \text{ matrix}$$

$a_{ij} \quad i=1,2,\dots,n$
 $j=1,2,\dots,m$

- If $n=m \rightarrow A$ is square matrix.
- If all $a_{ij}=0$ except diagonal $\rightarrow A$: diagonal matrix.
- If A is a diagonal matrix with $a_{ii}=1$
 \rightarrow Identity matrix (I) $A=I$

VECTORS are a special type of matrix.
(one column and n rows)

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad n \times 1 \text{ matrix} : \text{vector}$$

TRANSPOSE of A : A^T ($m \times n$) matrix

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix} \quad (m \times n) \text{ matrix}$$

if A is a square matrix : ($m=n$) and $A=A^T$
 $\rightarrow A$ is called a symmetric matrix

BASIC MATRIX OPERATIONS

Equality $A = B$ if $a_{ij} = b_{ij}$ for all i and j

Addition $A + B = C$ $c_{ij} = a_{ij} + b_{ij}$ for all i and j

Multiplication $A \cdot B = C$ this requires $m = q$
 $(n \times m)$ $(q \times r)$ $(n \times r)$

$$c_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$$

matrix multiplication
is NOT commutative
as in scalars

$$AB \neq BA.$$

Multiplication of A by a scalar λ

$$\lambda \cdot A = B \quad b_{ij} = \lambda \cdot a_{ij}$$

COMMONLY ENCOUNTERED VECTOR-MATRIX PRODUCTS

- IF x and y are n -dimensional vectors

$$1. \quad x^T x = \sum_{i=1}^n x_i^2 \quad (\text{scalar})$$

$$2. \quad x^T \cdot y = \sum_{i=1}^n x_i y_i$$

$$3. \quad x^T A x = \sum_{i=1}^n a_{ii} x_i^2 \quad A \text{ is } (n \times n) \text{ matrix}$$

$(1 \times n) \quad (n \times n) \quad (n \times 1)$ if A is diagonal

$x^T A x$ is a scalar

$$4 \quad X X^T = \begin{bmatrix} x_1 x_1 & x_1 x_2 & \dots & x_1 x_n \\ x_2 x_1 & x_2 x_2 & \dots & x_2 x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n x_1 & x_n x_2 & \dots & x_n x_n \end{bmatrix}$$

(n \times n)

THE INVERSE OF MATRIX A (A^{-1})

A^{-1} is defined such that

$$A \cdot A^{-1} = A^{-1} \cdot A = I$$

$$A \cdot B = C \quad \therefore \quad A^{-1} A B = A^{-1} C \quad \therefore \quad I \cdot B = A^{-1} C$$

$$B = A^{-1} C \quad \text{because} \quad I B = B$$

$$A \cdot X = b \quad \therefore \quad A^{-1} A \cdot X = A^{-1} b \quad \therefore \quad \underline{\underline{X = A^{-1} b}}$$

if A is a diagonal matrix: $A^{-1} = [a_{ji}^{-1}]$

DETERMINANT

The determinant of a matrix A, denoted $\det[A]$ or $|A|$

In general we define a Cofactor as

follows $C_{ij} = (-1)^{L+j} \det[M_{ij}]$

M_{ij} is obtained by deleting from A the i^{th} row and j^{th} column.

Then the determinant is calculated by either

1. $\sum_{j=1}^n a_{ij} C_{ij}$ (row expansion)

2. $\sum_{i=1}^n a_{ij} C_{ij}$ (column expansion)

Example $\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{21}a_{12}$

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$$

BASIC CONCEPTS (Gradient vector, Hessian Matrix, Jacobian matrix)

Let $F(\mathbf{x})$ be a typical function where \mathbf{x} is the real n -dimensional vector $(x_1, x_2, \dots, x_n)^T$. **Vectors** are denoted by using lower case **boldface** letters. Vectors are considered column vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1, x_2, \dots, x_n)^T \quad \checkmark$$

The gradient vector $\mathbf{g}(\mathbf{x})$ and the Hessian matrix $\mathbf{G}(\mathbf{x})$ are given by

$$\nabla F(\mathbf{x}) \equiv \frac{\partial F}{\partial \mathbf{x}} = \mathbf{g}(\mathbf{x}) = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots, \frac{\partial F}{\partial x_n} \right]^T$$

$$\mathbf{G}(\mathbf{x}) \equiv \nabla^2 F(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & \dots & \frac{\partial^2 F}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix}$$

Vector sizes³. Given an n -dimensional vector the **norms of vector** (infinity norm, Norm-1, Euclidian or Norm-2 and the Norm- k) offer a measure of the "length" of the vector.

$\| \cdot \|_\infty$ $\| \cdot \|_2$

The infinity norm is $\|\mathbf{x}\|_\infty = \max(x_i)$

Norm-1 is $\|\mathbf{x}\|_1 = [\sum_{i=1}^n |x_i|]$

Norm- k is $\|\mathbf{x}\|_k = [\sum_{i=1}^n |x_i|^k]^{1/k}$

The Euclidean norm is $\|\mathbf{x}\|_2 = (x_1^2 + x_2^2 + \dots + x_n^2)^{1/2} = (\mathbf{x}^T \mathbf{x})^{1/2}$.

³ Vasiliadis et al. Optimization for Chemical and Biochemical Engineering, CUP, 2020.

Jacobian Matrix.

Let $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_m(\mathbf{x})]$ (a vector-valued function) where $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$.

Then the **Jacobian matrix**, is given by

$$\mathbf{J}(\mathbf{x}) = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \dots & \frac{\partial h_1}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \frac{\partial h_m}{\partial x_1} & \frac{\partial h_m}{\partial x_2} & \dots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}$$

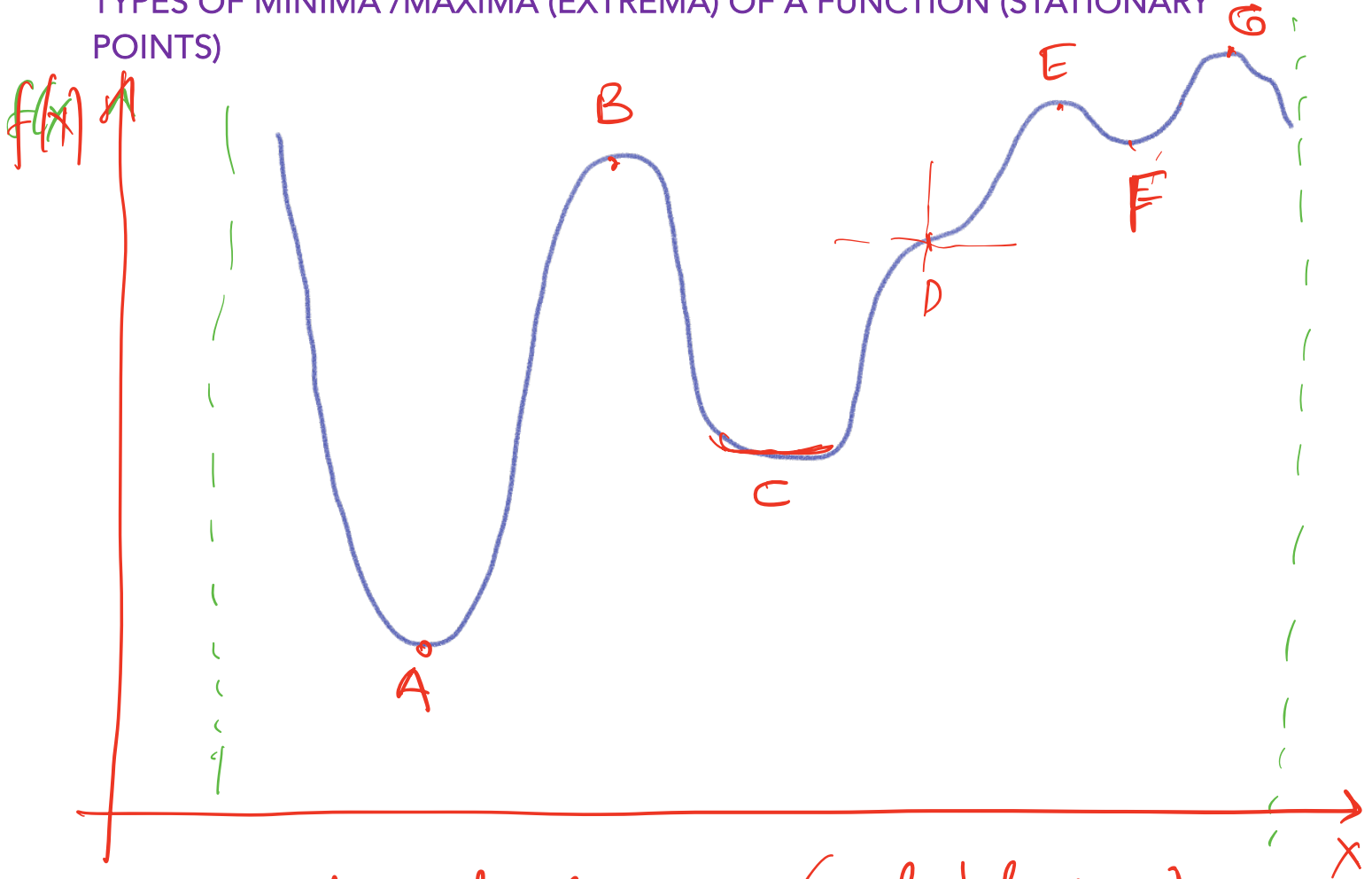
The **Hessian matrix** of a scalar function $F(\mathbf{x})$ is the **Jacobian matrix** of the vector function $\mathbf{g}(\mathbf{x})$, the gradient vector i.e.

$$\mathbf{g}(\mathbf{x}) = \nabla F(\mathbf{x}) = \left[\frac{\partial F}{\partial x_1} = h_1, \frac{\partial F}{\partial x_2} = h_2, \dots, \frac{\partial F}{\partial x_n} = h_n \right]^T \text{ and}$$

$$\mathbf{J}(\mathbf{x}) = \partial \mathbf{g} / \partial \mathbf{x} = \nabla^2 F(\mathbf{x}) = \mathbf{G}(\mathbf{x})$$

NOTE: For continuously differentiable functions, H is symmetric

TYPES OF MINIMA / MAXIMA (EXTREMA) OF A FUNCTION (STATIONARY POINTS)



A : strong local min (global min)

B strong local max

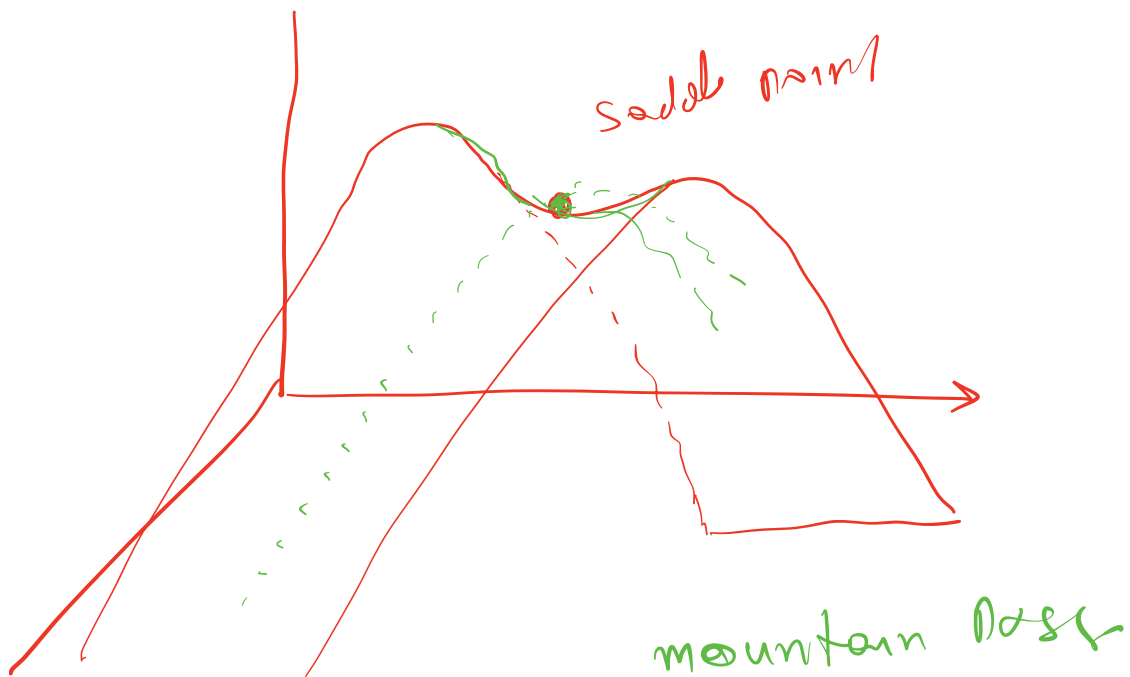
E strong local max

G strong local max (global max)

C weak local min.

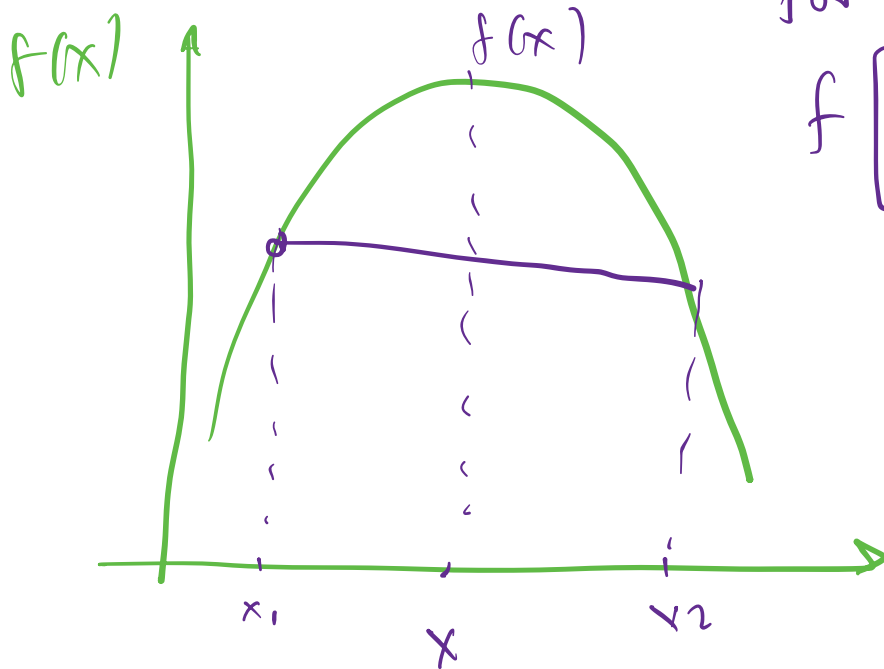
D inflection point

F strong local min.



CONVEX and CONCAVE FUNCTIONS. The concept of convexity of functions is important in Optimization

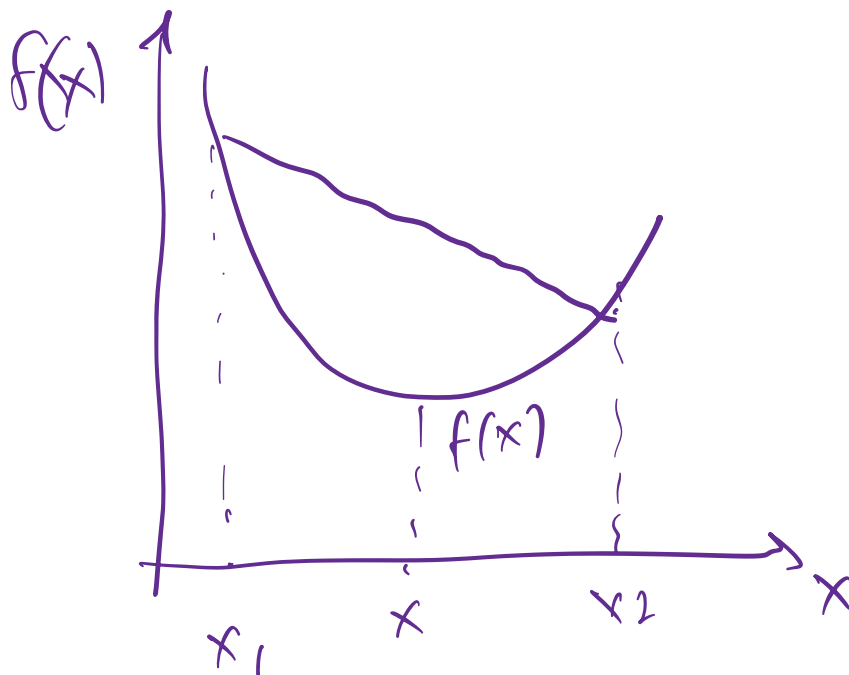
CONCAVE function



for all $\theta \in [0, 1]$

$$f\left[(1-\theta)x_1 + \theta x_2\right] \geq (1-\theta)f(x_1) + \theta f(x_2)$$

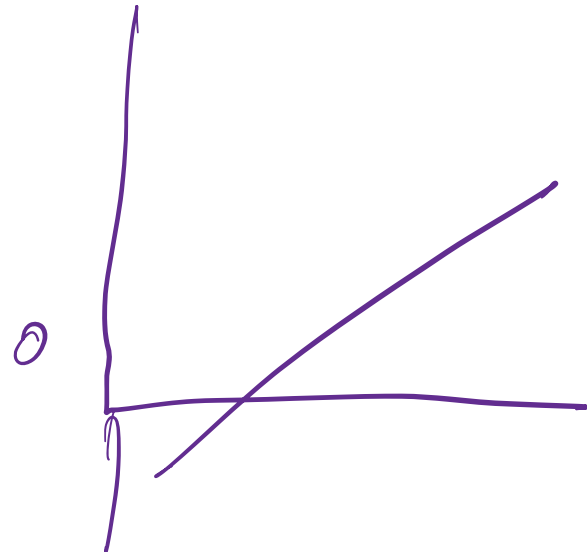
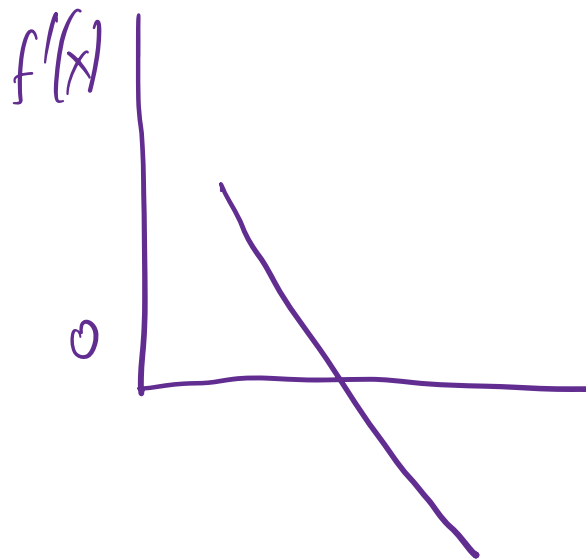
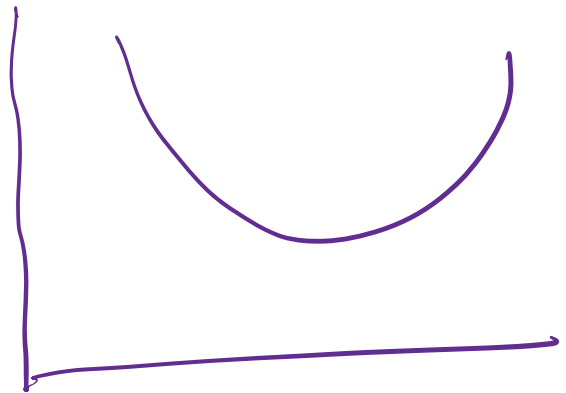
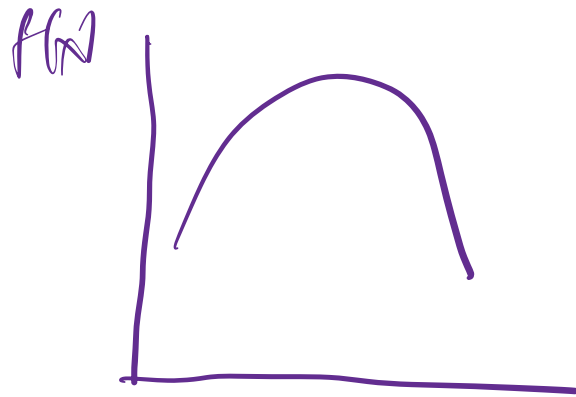
CONVEX function



$$f\left[(1-\theta)x_1 + \theta x_2\right] \leq (1-\theta)f(x_1) + \theta f(x_2)$$

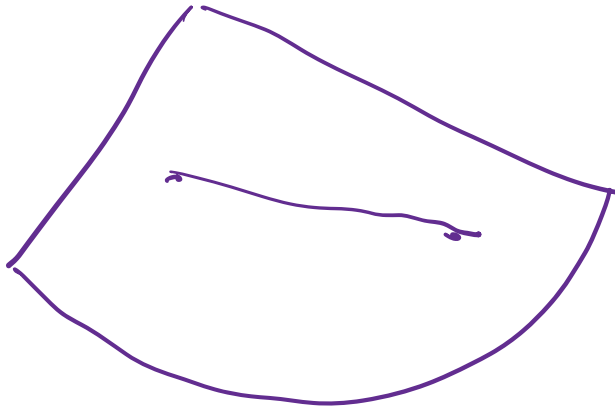
Convex and Concave Functions (single variable) and their 1st and 2nd derivatives

quadratics function:

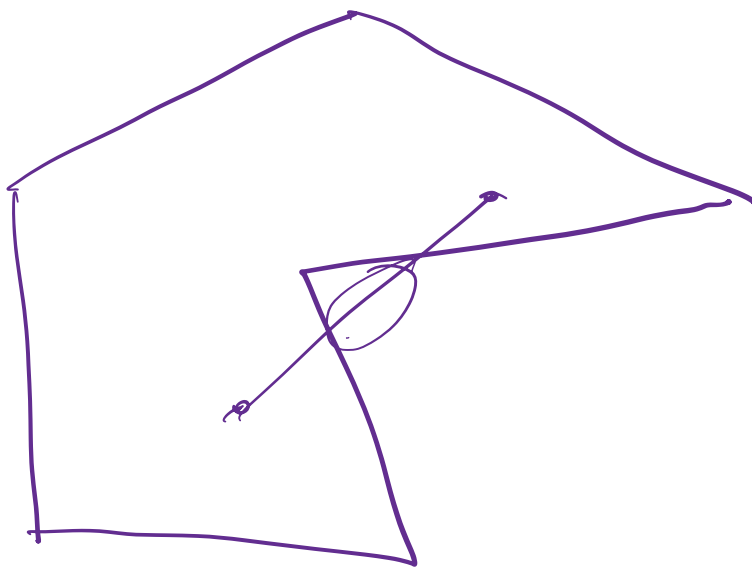


if $f''(x) \leq 0 \rightarrow$ concave
 $f''(x) \geq 0 \rightarrow$ convex.

CONVEX SETS. For global optimality it is not sufficient that the function $F(\mathbf{x})$ be convex. The set where the vector \mathbf{x} of variables takes its values is also a *convex* set. Otherwise, there is no guarantee of global optimality. A convex set is a set such that any chord connecting two points in the set lies entirely in the set. This is shown below⁴ (



convex set



non-convex set

⁴ Vasiliadis et al. Optimization for Chemical and Biochemical Engineering, CUP, 2020; Boyd, S.P. and Vandenberghe L. (2004). Convex Optimization, CUP.

CONVEXITY AND CONCAVITY (multivariable functions)

Consider the multivariable function $F(\mathbf{x})$ which has continuous second partial derivatives

- $F(\mathbf{x})$ is **concave** if and only if the Hessian matrix, $\mathbf{G}(\mathbf{x})$, is negative semi-definite. For $F(\mathbf{x})$ to be *strictly concave*, $\mathbf{G}(\mathbf{x})$ must be negative definite. $\mathbf{G}(\mathbf{x})$ is negative definite if and only if $\mathbf{x}^T \mathbf{G} \mathbf{x}$ is < 0 for all $\mathbf{x} \neq \mathbf{0}$.
- $F(\mathbf{x})$ is **convex** if $\mathbf{G}(\mathbf{x})$ is positive semidefinite. For $F(\mathbf{x})$ to be *strictly convex*, $\mathbf{G}(\mathbf{x})$ must be positive definite. $\mathbf{G}(\mathbf{x})$ is positive definite if and only if $\mathbf{x}^T \mathbf{G} \mathbf{x}$ is > 0 for all $\mathbf{x} \neq \mathbf{0}$.
- Note that $\mathbf{G}(\mathbf{x})$ is indefinite if and only if $\mathbf{x}^T \mathbf{G} \mathbf{x}$ is < 0 for some \mathbf{x} and > 0 for other \mathbf{x} .

$$(1 \times n) (n \times n) (n \times 1) = 1 \times 1$$

Note: It is noted that for *convexity* and *concavity*, the strict inequalities $>$ or $<$, respectively, in the tests are replaced by \geq or \leq , respectively.

Test for strict concavity: All the **eigenvalues** of $\mathbf{G}(\mathbf{x})$ are negative (< 0).

Test for strict convexity: All the **eigenvalues** of $\mathbf{G}(\mathbf{x})$ are positive (> 0).⁵

Note: If a function has a *stationary point* where the Hessian has eigenvalues of mixed signs, the function is neither convex nor concave.

⁵ Bellman, R Introduction to Matrix Analysis, 2nd edition, SIAM < 1997

Eigenvalues and Singular Value decomposition (SVD)⁶:

Let \mathbf{A} be a *square* ($n \times n$) matrix. If there exists a scalar λ and a nonzero vector \mathbf{v} such that $\mathbf{A}\mathbf{v}=\lambda\mathbf{v}$

Then λ is called an *eigenvalue* of \mathbf{A} and \mathbf{v} is the corresponding *eigenvector*. All eigenvalues λ_i (some of which may be equal) can be obtained by solving the characteristic equation of \mathbf{A} i.e. $\det(\mathbf{A}-\lambda\mathbf{I})=0$.

If the matrix \mathbf{A} is *normal* i.e. $\mathbf{A}\mathbf{A}^T=\mathbf{A}^T\mathbf{A}$, then it can be factorized into

$$\mathbf{A}=\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

Where \mathbf{V} is an ($n \times n$) orthogonal matrix and $\mathbf{\Lambda}$ is a diagonal matrix which has all the eigenvalues of \mathbf{A} as its diagonal elements. We say that matrix \mathbf{A} is orthogonally diagonalizable and $\mathbf{\Lambda} = \mathbf{V}^T\mathbf{A}\mathbf{V}$.

- The set of all eigenvalues of a matrix is its *spectrum*.
- The *spectral radius* of matrix \mathbf{A} is defined as $\rho(\mathbf{A})=\max |\lambda_i(\mathbf{A})|$.
- If matrix \mathbf{A} is *symmetric* ($\mathbf{A}=\mathbf{A}^T$) then all eigenvalues are real numbers.
- If matrix \mathbf{A} is *non-singular* (its inverse \mathbf{A}^{-1} exists) then all its eigenvalues are nonzero and the eigenvalues of \mathbf{A}^{-1} are the reciprocals of the eigenvalues of \mathbf{A} .

Note: A *square matrix* with *Determinant* equal to zero is a singular matrix.

⁶ Cichoki A. and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, Wiley, 1993, New York, NY.

Singular Value Decomposition (SVD). Let A be an $(m \times n)$ matrix. Then real *orthogonal matrices* U ($m \times m$) and V ($n \times n$) exist such that

$$(m \times m) (m \times n) (n \times n) = (m \times n)$$

$$U^T A V = S$$

where S is a diagonal matrix $\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$ and $p = \min(m, n)$.

The real non-negative numbers σ_i are called *singular values* of A and matrix A can be written as $A = U S V^T$. It is noted that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

The singular values of matrix A are the square roots of the nonzero eigenvalues of $A^T A$ or $A A^T$.

A square ($n \times n$) matrix is called *orthogonal* if $Q^T Q = Q Q^T = I$. For $m \neq n$, an $(m \times n)$ matrix is called orthogonal if $Q^T Q = I$. All the diagonal elements of the identity matrix (I) are equal to 1 and the rest are zeros.

A specific and important case of the singular value decomposition (SVD) is obtained when A is a *symmetric* ($A = A^T$) nonnegative definite matrix. In this case the matrix S is $S = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$. are the real eigenvalues of A .

One important use of the SVD is the solution of a system of linear equations: $Ax = b$ where A is an $(m \times n)$ matrix, x an n -dimensional vector and b an m -dimensional vector. Substituting the SVD of A we obtain

$$U S V^T x = b. \text{ Then } x = V S^{-1} U^T b.$$

$$\underbrace{U S V^T}_A \quad \uparrow \quad \underbrace{V S^{-1} U^T}_{U^T}$$

Example: Consider the following objective function: is it convex? Use eigenvalues in the analysis. $f(\mathbf{x}) = 2x_1^2 + 2x_1x_2 + 1.5x_2^2 + 7x_1 + 8x_2 + 24$.

Solution

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} = 4, \quad \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} = 3, \quad \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} = \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} = 2$$

$$\text{Hessian matrix, } \mathbf{G}(\mathbf{x}) = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$$

$$\mathbf{G} - \lambda \mathbf{I} = \begin{bmatrix} 4-\lambda & 2 \\ 2 & 3-\lambda \end{bmatrix} \quad \text{and} \quad \det \begin{bmatrix} 4-\lambda & 2 \\ 2 & 3-\lambda \end{bmatrix} = (4-\lambda)(3-\lambda) - 4 = 0 \quad \text{or} \quad \lambda^2 - 7\lambda + 8 = 0.$$

Hence, $\lambda_1 = 5.56$ and $\lambda_2 = 1.44$.

Both eigenvalues are positive and hence the function is strictly convex.

Example: Consider the following objective function $f(\mathbf{x}) = 2x_1^2 - 3x_1x_2 + 2x_2^2$. Is it convex?

$$\frac{\partial f}{\partial x_1} = 4x_1 - 3x_2$$

$$\frac{\partial f}{\partial x_2} = -3x_1 + 4x_2$$

$$\frac{\partial^2 f}{\partial x_1^2} = 4$$

$$\frac{\partial^2 f}{\partial x_2^2} = 4$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = -3$$

$$\mathbf{G}(\mathbf{x}) = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}, \quad [\mathbf{G} - \lambda \mathbf{I}] = \begin{bmatrix} 4-\lambda & -3 \\ -3 & 4-\lambda \end{bmatrix}$$

$$\det [\mathbf{G}(\mathbf{x}) - \lambda \mathbf{I}] = 0 \quad \therefore \quad \det \begin{bmatrix} 4-\lambda & -3 \\ -3 & 4-\lambda \end{bmatrix} = 0 \quad \therefore \quad (4-\lambda)^2 = 9$$

$$(4-\lambda) = \pm 3 \quad \therefore \quad \lambda_1 = 4-3 = 1$$

$$\lambda_2 = 4+3 = 7$$

$$\rightarrow \mathbf{G}(\mathbf{x}) = \text{P.D.}$$

$f(\mathbf{x}) = \text{strictly convex}$

Example: Consider the following objective function $f(x) = x_1^2 + x_1x_2 + 2x_2 + 4$. Is it convex?

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2 \quad \frac{\partial f}{\partial x_2} = x_1 + 2 \quad \frac{\partial^2 f}{\partial x_1^2} = 2 \quad \frac{\partial^2 f}{\partial x_2^2} = 0$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 1 \quad G = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}, [G - \lambda I] = \begin{bmatrix} 2-\lambda & 1 \\ 1 & -\lambda \end{bmatrix}$$

$$\det[G - \lambda I] = 0 \Rightarrow (2-\lambda)(-\lambda) - 1 = 0 \Rightarrow \lambda^2 - 2\lambda - 1 = 0$$

$$\therefore \lambda_{1,2} = \frac{2 \pm \sqrt{4+4}}{2} = \lambda_1 = 1 + \sqrt{2}, \lambda_2 = 1 - \sqrt{2} < 0$$

$$\therefore \begin{pmatrix} \lambda_1 > 0 \\ \lambda_2 < 0 \end{pmatrix} \rightarrow f(x) : \text{neither convex or concave,} \\ \text{mixed signs}$$

NECESSARY AND SUFFICIENT CONDITIONS FOR AN EXTREMUM

(EH P. 135-142)

Now Consider a multivariable function $F(\mathbf{x})$. If $F(\mathbf{x})$ has continuous second partial derivatives,

The **Necessary** and **Sufficient Conditions** for an *extremum* (\mathbf{x}^*) of **the** unconstrained Function $F(\mathbf{x})$ where $\mathbf{x}=(x_1, x_2, \dots, x_n)^T$ are as follows.^{7,8,6}

Necessary Conditions:

1. $F(\mathbf{x})$ is twice differentiable at \mathbf{x}^*
2. $\nabla F(\mathbf{x}^*)=0$, that is, a *stationary point* exists at \mathbf{x}^* .

Sufficient Condition:

$H(\mathbf{x}^*)=\nabla^2 F(\mathbf{x}^*)$ is positive definite for a minimum to exist at \mathbf{x}^* or negative definite for a maximum to exist at \mathbf{x}^* .



$$\underline{x=x}$$

⁷Bertsekas, D.P. Nonlinear Programming, 3rd edition, 2016. Pages 6-12

⁸Gill, P.E., W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, UK, 1981 p. 59-82

⁶Scales, L.E., *Introduction to Non-linear Optimization*, Springer-Verlag, New York, NY, 1985. P 14-23

Example 1. Identify the stationary points of the following function and determine if any extrema exist.

$$f(x) = 4 + 4.5x_1 - 4x_2 + x_1^2 + 2x_2^2 - 2x_1x_2 + x_1^4 - 2x_1^2x_2 \quad x_1, x_2$$

Example 2. Consider $f(x)=x^4$. Does this function have an extremum? If so, what is the value of x^* and $f(x^*)$ at the extremum.

Ex. 1 $\frac{\partial f}{\partial x_1} = 0 \therefore 4.5 + 2x_1 - 2x_2 + 4x_1^3 - 4x_1x_2 = 0 \quad (a)$

$$\frac{\partial f}{\partial x_2} = 0 \therefore -4 + 4x_2 - 2x_1 - 2x_1^2 = 0 \quad (b)$$

from (b): $x_2 = \frac{2x_1 + 2x_1^2 + 4}{4} \quad (c)$ Subst. into (a)

$$\therefore 4.5 + 2x_1 - 2\left(\frac{x_1}{2} + \frac{x_1^2}{2} + 1\right) + 4x_1^3 - 4x_1\left(\frac{x_1}{2} + \frac{x_1^2}{2} + 1\right) = 0$$

$$\therefore 4.5 + 2x_1 - x_1 - x_1^2 - 2 + 4x_1^3 - 2x_1^2 - 2x_1^3 - 4x_1 = 0$$

$$\therefore 4.5 - 3x_1 - 3x_1^2 + 2x_1^3 = 0 \rightarrow \text{solve for } x_1$$

Substitute in (c) + get x_2

STATIONARY POINTS	$f(x)$	λ	CLASSIFICATION
$(1.941, 3.854)$	0.9855	37.03 0.97	Local Min
$(-1.053, 1.028)$	-0.5134	10.15 3.8	Local Min
$(0.0117, 1.4929)$	2.83	7.0 -2.56	Saddle Point

$$f(x) = x^4$$

$$f'(x) = 4x^3$$

$$f''(x) = 12x^2$$

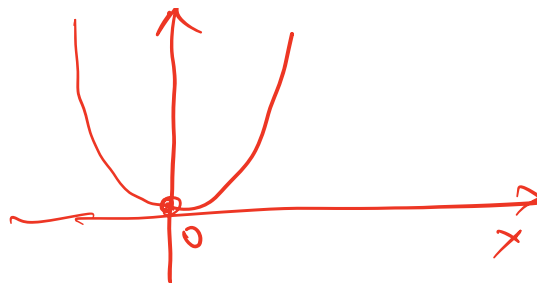
$$f'(x) = 0 \therefore x = 0$$

$$f''(x) = 0 \text{ at } x = 0,$$

$$f'''(x) = 24x$$

$$f^{(4)}(x) = 24$$

$$n = 4$$



$$f^{(n)}(x^*) \neq 0$$

IF n is even then

$f(x)$ has a MIN if $f^{(n)}(x) > 0$

$f(x)$ has a MAX if $f^{(n)}(x) < 0$

IF n is odd then

$f(x)$ has a saddle point

IN our case $n = 4$, $f^{(4)}(x) = 24 > 0 \rightarrow \underline{\underline{\text{MINIMUM}}}$