# INTRO/REFRESHER – OPTIMISATION

MATHEMATICAL OPTIMISATION : It is the problem of finding the best, or optimal, solution among a set of decisions

OPTIMALITY IS DEFINED BY USING AN OBJECTIVE FUNCTION

SOME SOLUTION CANDIDATES ARE SAID TO BE FEASIBLE
⟶ OTHERS ARE NOT (FEASIBLE)

FEASIBILITY IS VERIFIED BY USING CONSTRAINT FUNCTIONS

— // — // —

There exists a number of classes of optimisation classes
→ WE REVIEW THE BASIC CONCEPTS
→ WE REVIEW THE BASIC CLASSES OF PROBLEMS

Important to choose the right algorithm to solve a certain task



# INTRO/REFRESHER OPTIMIZATION

Elements of an optimisation problem (3 MAIN INGREDIENTS)
- OBJECTIVE FUNCTION $f(x)$
- DECISION VARIABLES $x$
- CONSTRAINT FUNCTIONS: $g(x) = 0$, $h(x) \geq 0$

The objective function is to be optimised (maximised or minimised) by chosing 'optimal' values of the decision variables that satisfy the constraints

STANDARD MATHEMATICAL FORM

$$\min_{x \in \mathbb{R}^N} f(x)$$
$$\text{subject to} \quad g(x) = 0$$
$$h(x) \geq 0$$

$f: \mathbb{R}^N \to \mathbb{R}$ (scalar)

$g: \mathbb{R}^N \to \mathbb{R}^p$ (there are $p$ equality constraints)

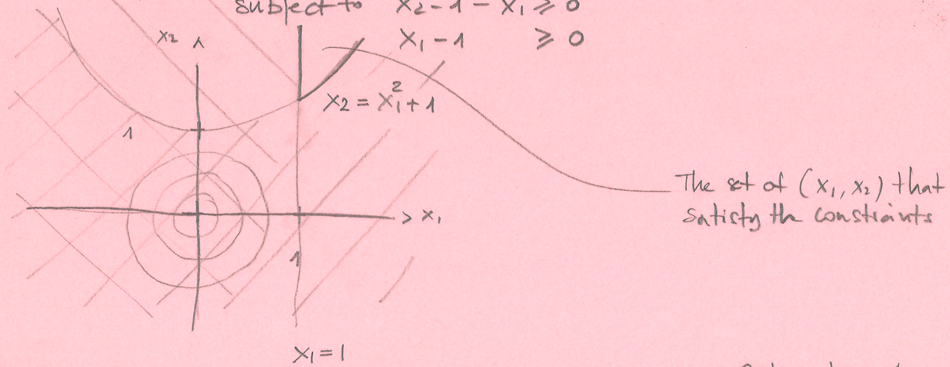
$h: \mathbb{R}^N \to \mathbb{R}^d$ (there are $d$ inequality constraints)

ALL FUNCTIONS ARE ASSUMED TO BE DIFFERENTIABLE (at least)

THE INEQUALITIES $h(x) \geq 0$ HOLD FOR ALL THE COMPONENTS
⟶ $h(x) \geq 0$ — $h_i(x) \geq 0$ for all $i = 1, \dots, q$

**Example**

$$\text{minimize} \quad x_1^2 + x_2^2$$
$$\text{subject to} \quad x_2 - 1 - x_1^2 \geq 0$$
$$x_1 - 1 \geq 0$$



**Def.** LEVEL SETS: $\{x \in \mathbb{R}^N \mid f(x) = c\}$ for some constant $c$

**Def.** FEASIBLE SET: $\Omega = \{x \in \mathbb{R}^N \mid g(x) = 0 , h(x) \geq 0\}$

**Def.** GLOBAL MINIMIZER: $x^*$ IFF $x^* \in \Omega$ and $\forall x \in \Omega : f(x) \geq f(x^*)$

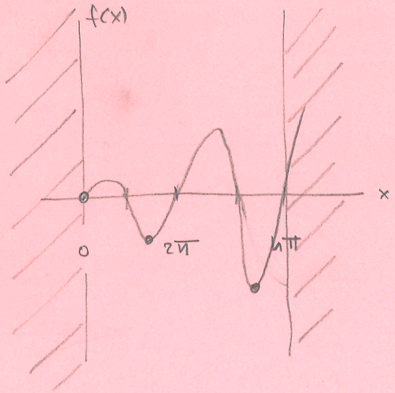
**Def.** STRICT GLOBAL MINIMIZER: $x^*$ IFF $x^* \in \Omega$ and
$\forall x \in \Omega \setminus \{x^*\} : f(x) > f(x^*)$

**Def.** LOCAL MINIMISER $x^*$ IFF $x^* \in \Omega$ and there exists a neighbourhood around $x^*$ such that $\forall x \in \Omega \cap N(x^*) : f(x) \geq f(x^*)$

**Def.** STRICT LOCAL MINIMISER $x^*$ IFF $x^* \in \Omega$ and
$\forall x \in \Omega / \{x^*\} : f(x) > f(x^*)$

**Example**

$$\underset{x \in \mathbb{R}}{\text{minimise}} \quad \sin(x)\exp(x)$$
$$\text{subject to} \quad x \geq 0, \quad x \leq 4\pi$$




$\Omega = \{x \in \mathbb{R} \mid x \geq 0 , x \leq 4\pi\}$

GLOBAL MINIMISER ?

LOCAL MINIMISER ?



**Theorem** IF $\Omega \in \mathbb{R}^N$ IS NONEMPTY AND COMPACT (BOUNDED AND CLOSED) AND IF $f : \Omega \to \mathbb{R}$ IS CONTINUOUS THEN THERE EXISTS A GLOBAL MINIMISER THAT SOLVES

$$\underset{x \in \mathbb{R}^N}{\text{minimise}} \quad f(x) \quad \text{subject to } x \in \Omega$$

---

Thus minimisers exist under fairly mild conditions and we are interested in finding them using computer algorithms

SOME NOTATIONAL DEVICES

$\mathbb{R}$ : real numbers
$\mathbb{R}_+$ : non negative real numbers
$\mathbb{R}_{++}$ : positive real numbers
$\mathbb{Z}$ : integers
$\mathbb{N}$ : natural numbers including zero $(\mathbb{N} = \mathbb{Z}_+)$
$\mathbb{R}^N$ : real valued N-dimensional vectors
$\mathbb{R}^{N \times M}$: real valued $N \times M$ matrices
$\mathbb{R}^{N \times 1}$: column vectors
$\mathbb{R}^{N+M}$: concatenation of vectors $[y^T x^T]^T = \begin{bmatrix} y \\ x \end{bmatrix}$

- Arbitrary norm of a vector $x$ is denoted by $\|x\|$, the Euclidean norm is denoted by $\|x\|_2$ and we have that $\|x\|_2^2 = x^T x$
- The weighted Euclidean norm $\|x\|_Q = x^T Q x$ for some positive weighting matrix $Q$
- The L1-norm $\|x\|_1 = \sum_{i=1}^{N} |x_i|$
- The $\infty$-norm $\|x\|_\infty = \max(\{|x_1|, \dots, |x_N|\})$
- Matrix norms are induced operators if not stated otherwise
- The Frobenius norm of matrix $A$ is denoted as $\|A\|_F$ so that $\|A\|_F^2 = \text{trace}(AA^T) = \sum_i \sum_j A_{ij} A_{ij}$

FUNCTIONS OF SEVERAL INPUTS AND WITH SEVERAL OUTPUTS ARE

$$f : \mathbb{R}^N \to \mathbb{R}^M , \; x \mapsto f(x)$$

$$f = \begin{bmatrix} f_1(x_1, \_\_\_, x_N) \\ f_2(x_1, \_\_\_, x_N) \\ \vdots \\ f_M(x_1, \_\_\_, x_N) \end{bmatrix}$$

WE DEFINE THE JACOBIAN MATRIX $\left.\frac{\partial f}{\partial x}\right|_x$ AS THE $\mathbb{R}^{M \times N}$

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_M}{\partial x_1} & \frac{\partial f_M}{\partial x_2} & \cdots & \frac{\partial f_M}{\partial x_N} \end{bmatrix}_x$$

FOR SCALAR FUNCTIONS THE JACOBIAN IS A 1-ROW MATRIX (A ROW VECTOR), WE DEFINE THE GRADIENT AS ITS TRANSPOSE

$$f : \mathbb{R}^N \to \mathbb{R} , \; x \mapsto f(x)$$

$$\nabla f(x) = \left.\frac{\partial f}{\partial x}\right|_x^T$$

FOR VECTOR-VALUED FUNCTION, THE NOTATION REMAINS VALID

USING THIS NOTATION, WE DEFINE THE TAYLOR EXPANSION OF FIRST ORDER

$$f(x) = f(\bar{x}) + \nabla f(x)^T (x - \bar{x}) + \text{higher order terms}$$



FOR SCALAR FUNCTIONS WE DEFINE THE HESSIAN MATRIX AS THE MATRIX WHOSE ELEMENTS ARE THE SECOND DERIVATIVES OF $f$, $\nabla^2 f(x)$

(FOR VECTOR VALUED FUNCTIONS, IT BECOMES A TENSOR)

CONSIDER A SYMMETRIC MATRIX $A$

- $A \succeq 0$ DENOTES A POSITIVE-SEMI-DEFINITE MATRIX (all its evalues are non-negative)
- $A \succ 0$ DENOTE A POSITIVE-DEFINITE MATRIX (all its evalues are positive)

The same notation is used for matrix inequalities

e.g. $A \succeq B$ means $A - B \succeq 0$

WHY B IS CALLED THE 'HESSIAN' MATRIX IN A QP ?

$$\frac{\partial J}{\partial x} = \partial/\partial x \left( c^T x + \frac{1}{2} x^T B x \right)$$
$$= c + \frac{1}{2} 2 x^T B \quad \text{(BECAUSE B IS SYMMETRIC)}$$
$$\frac{\partial^2 J}{\partial x \partial x} = \frac{\partial}{\partial x}\left( c + x^T B \right) = B$$

EXAMPLE (Non convex QP) minimize $[0\ 2]x + \frac{1}{2}x^T \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} x$

subject to $x_1 \in [-1, 1]$
$x_2 \in [-1, 10]$

EXAMPLE (STRICTLY CONVEX QP) minimize $[0\ 2]x + \frac{1}{2}x^T \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} x$

subject to $x_1 \in [-1, 1]$
$x_2 \in [-1, 10]$

## CLASSES of OPTIMISATION PROBLEMS

**Nonlinear programs**

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(x) \quad \text{subject to} \quad g(x) = 0, \ h(x) \geq 0$$

$f: \mathbb{R}^n \to \mathbb{R}$, $g: \mathbb{R}^n \to \mathbb{R}^p$, $h: \mathbb{R}^n \to \mathbb{R}^q$

CONTINUOUSLY DIFFERENTIABLE (AT LEAST ONCE, OFTEN TWICE OR MORE)

From differentiability we derive algorithms based on derivatives $\underbrace{\text{(NEWTON-TYPE OPTIMIZATION)}}_{\text{This are the workhorse algorithms}}$

**Linear programs**

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \ \underbrace{c^T x}_{f(x)} \ (+c_0) \quad \text{subject to} \quad \underbrace{Ax - b}_{g(x)} = 0, \ \underbrace{Cx - d}_{h(x)} \geq 0$$

$f: \mathbb{R}^N \to \mathbb{R}$, $g: \mathbb{R}^N \to \mathbb{R}^p$, $h: \mathbb{R}^N \to \mathbb{R}^q$

CONTINUOUSLY DIFFERENTIABLE AND AFFINE

$c \in \mathbb{R}^N$
$A \in \mathbb{R}^{p \times N}$
$b \in \mathbb{R}^p$
$C \in \mathbb{R}^{q \times N}$
$d \in \mathbb{R}^q$

1940 the simplex method for efficient solution (ACTIVE SET TYPE OF METHOD)

**Quadratic programs**

$$\text{minimize} \ c^T x + \frac{1}{2} x^T B x \quad \text{subject to} \quad Ax - b = 0, \ Cx - d \geq 0$$

$f: \mathbb{R}^N \to \mathbb{R}$, $g: \mathbb{R}^N \to \mathbb{R}^p$, $h: \mathbb{R}^N \to \mathbb{R}^q$

g, h are affine, f is linear-quadratic

$B \in \mathbb{R}^{N \times N}$ ('Hessian matrix', because $\nabla^2 f(x) = B$)

$\nabla^2 f(x) = B$ for $f(x) = c^T x + \frac{1}{2} x^T B x$

If $B \geq 0$ then the QP problem is convex
If $B > 0$ then the QP problem is strictly convex

CONVEXITY ⇒ GLOBAL OPTIMALITY

## GENERAL CONVEX OPTIMISATION PROBLEMS

LINEAR PROGRAMS AND SOME QUADRATIC PROGRAMS ARE CONVEX OPTIMISATION PROBLEMS

We recall the definiton of convex set and convex fuction

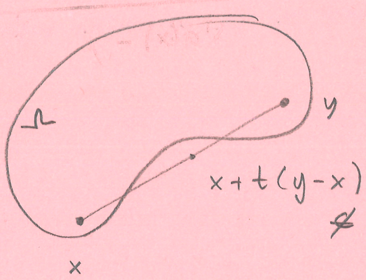
**Def. CONVEX SET:** A set $\Omega \subset \mathbb{R}^n$ is convex IFF

$$\forall (x,y) \in \Omega,\ t \in [0,1]:\ x + t(y-x) \in \Omega$$

(ALL CONNECTING LINES LIE INSIDE THE SET


CONVEX SET

$x + t(y-x),\ t \in [0,1]$
$\in \Omega$

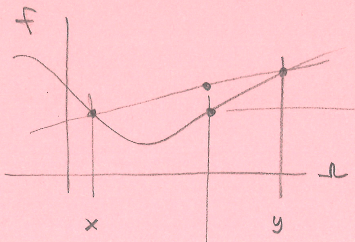



$x + t(y-x),\ t \in [0,1]$
$\notin \Omega$

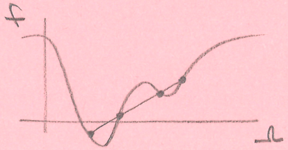
**Def. CONVEX FUNCTION:** A function $f: \Omega \to \mathbb{R}$ is convex IFF

$$\forall (x,y) \in \Omega,\ t \in [0,1]:\ f(x+t(y-x)) \leq f(x) + t(f(y)-f(x))$$

(ALL SECANT LINES/PLANES ARE ABOVE THE GRAPH)




$f(x+t(y-x)) < f(x) + t(f(y)-f(x))$






The Lagrangian function is a function that allow us to find a lower bound on the optimal value

# THE LAGRANGIAN FUNCTION AND DUALITY

Def. (PRIMAL FORMULATION)

$$\boxed{\begin{aligned}&\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x)\\ &\text{subject to}\quad g(x)=0\\ &\qquad\qquad\quad h(x)\geq 0\end{aligned}}$$

The globally optimal value of the objective function subject to the constraints is called the 'PRIMAL OPTIMAL VALUE' $p^*$

$$p^* = \left(\min_{x\in\mathbb{R}^n} f(x),\ \text{s.t. } g(x)=0,\ h(x)\geq 0\right)$$

The minimal function value

Def (LAGRANGIAN FUNCTION + LAGRANGIAN MULTIPLIERS)

We define $\boxed{\mathcal{L}(x,\lambda,\mu) = f(x) - \lambda^T g(x) - \mu^T h(x)}$ LAGRANGIAN FUNCTION

$\lambda\in\mathbb{R}^p$, $\mu\in\mathbb{R}^q$ } DUAL VARIABLES (OR LAGRANGIAN MULTIPLIERS)

We typically require the inequality multipliers to be positive or zero

$\rightsquigarrow \mu \geq 0$

We do not restrict the sign of the equality multipliers

$\rightsquigarrow \lambda \in \mathbb{R}^p$

Lemma (THE LOWER BOUND OF THE LAGRANGIAN)

If $\bar{x}$ is a feasible point and $\mu\geq 0$, then $\mathcal{L}(\bar{x},\lambda,\mu)\leq f(\bar{x})$

As $\mathcal{L}(\bar{x},\lambda,\mu) = f(\bar{x}) - \lambda^T \underbrace{g(\bar{x})}_{=0} - \underbrace{\mu^T}_{\geq 0}\underbrace{h(\bar{x})}_{\geq 0} \leq f(\bar{x})$

The Lagrangian is a lower bound on $f$ at any feasible point $\bar{x}\in\Omega$



If we minimize the Lagrangian function, we know that we will have a lower bound for the optimal value

—//—//—

WE SET $\lambda$ AND $\mu$ IN $\mathcal{L}(x,\lambda,\mu)$ TO SOME VALUES, THEN WE MINIMISE IT WRT $x$

$\rightsquigarrow$ WE GET A FUNCTION $q(\lambda,\mu)$ WHICH DEPENDS ONLY ON $\lambda$ AND $\mu$ (AS $x$ HAS BEEN MINIMISED AWAY)

—//—//—

For any feasible point $\tilde{x}\in\Omega$, we have that $q(\lambda,\mu)\leq f(\tilde{x})$

$$q(\lambda,\mu) = \inf_x \mathcal{L}(x,\lambda,\mu) \leq f(\tilde{x})$$

For $\tilde{x}=x^*$, we have that $f(x^*) = p^*$, hence $q(\lambda,\mu)\leq p^*$

## LAGRANGE DUAL FUNCTION and WEAK DUALITY

**Def.** (LAGRANGE DUAL FUNCTION)

We define the Lagrange dual function as the unconstrained infimum of the lagrangian over x, for fixed values of the multipliers $\lambda, \mu$

$$\boxed{q(\lambda,\mu) = \inf_{x \in \mathbb{R}^N} \mathcal{L}(x,\lambda,\mu)}$$

THIS FUNCTION WILL OFTEN TAKE THE VALUE $-\infty$, IN WHICH CASE WE SAY THAT THE PAIR $\lambda, \mu$ IS DUAL INFEASIBLE

**Lemma** (LOWER BOUND OF THE LAGRANGE DUAL FUNCTION)

If $\mu \geq 0$, then $\boxed{q(\lambda,\mu) \leq p^*}$ whatever the value of $\lambda$

**Theorem** (CONCAVITY OF THE LAGRANGE DUAL FUNCTION)

Function $q: \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ is concave (even if the original NLP is convex)

QUESTION: WHAT IS THE BEST LOWER BOUND THAT WE CAN GET FROM THE LAGRANGE DUAL FUNCTION?

THIS CAN BE OBTAINED BY MAXIMISING THE LAGRANGE DUAL FUNCTION OVER ALL POSSIBLE VALUES OF THE MULTIPLIERS $\rightsquigarrow$ THIS YIELDS THE 'DUAL PROBLEM'

$\rightarrow$ IT CAN BE SHOWN BY SHOWING THAT $-q$ IS CONVEX



**Defi** (DUAL PROBLEM)

The DUAL PROBLEM, with DUAL OPTIMAL VALUE $d^*$ is defined as the convex maximisation problem

$$d^* = \left( \max_{\substack{\lambda \in \mathbb{R}^p \\ \mu \in \mathbb{R}^d}} q(\lambda,\mu) \text{ s.t. } \mu \geq 0 \right)$$

**Thm** (WEAK DUALITY) $\quad d^* \leq p^*$

## OPTIMALITY CONDITIONS

Consider the general unconstrained optimisation problem

$$\min_{x \in \Omega} f(x) \qquad (f \in C^1) \qquad \text{and } (\Omega \subseteq \mathbb{R}^N)$$

## NECESSARY OPTIMALITY CONDITIONS

**Th.** (FIRST ORDER NECESSARY CONDITIONS)

If $x^* \in \Omega$ is a local minimiser of $f : \Omega \to \mathbb{R}$ and $f \in C^1$, then

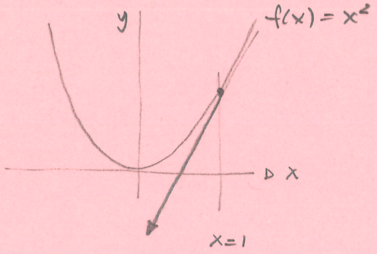
$$\nabla f(x^*) = 0$$

**Def.** (STATIONARY POINT)

A point $\bar{x}$ such that $\nabla f(\bar{x}) = 0$ is called a stationary point of f

**Def.** (DESCENT DIRECTION)

A vector $p \in \mathbb{R}^N$ with $\nabla f(x)^T p < 0$ is called a descent direction at $x$




$-\nabla f(x) = p = 2$

**Th** (SECOND ORDER NECESSARY CONDITIONS)

If $x^* \in \Omega$ is a local minimiser of $f : \Omega \to \mathbb{R}$ and $f \in C^2$, then

$$\nabla^2 f(x^*) \succeq 0$$

(It is not sufficient for $x^*$ to be a minimiser, ⇝ saddle points)

## NEWTON-TYPE OPTIMISATION

Consider the general unconstrained nonlinear optimisation problem

$$\min_{x \in \mathbb{R}^N} f(x) \qquad (f \in C^2)$$

An iterative algorithm generates a sequence of solutions $x_0, x_1, \ldots$ such that $x_k \to x^*$

## EXACT NEWTON METHOD

Newton's (or Newton-Raphson's) methods are methods for finding the root of equations in one ore more dimensions

Consider the equation $\nabla f(x^*) = 0$ with $\nabla f : \mathbb{R}^N \to \mathbb{R}^N$

Newton's methods idea consists of linerizing the nonlinear equations at $x_k$ to find $x_{k+1} = x_k + p_k$

$$\nabla f(x_k) + \frac{\partial}{\partial x}\left[\nabla f(x_k)\right] p_k = 0$$

$$-\nabla^2 f(x_k)^{-1} \nabla f(x_k) = p_k$$

$p_k$ is called the Newton's step and $\nabla^2 f(x_k)$ is the Hessian

## NEWTON-TYPE METHODS

Any iteration of the general form $x_{k+1} = x_k - B_k^{-1} \nabla f(x_k)$ is called a NEWTON-TYPE ITERATION FOR OPTIMIZATION

(B must clearly be invertible)

For $B_k = \nabla^2 f(x_k)$, we have the exact Newton's method

Usually we prefer $B_k \approx \nabla^2 f(x_k)$