NUMERICAL OPTIMIZATION

2b

OBJECTIVE FUNCTION OF w (SCALAR)

CONSIDER THE STANDARD NONLINEAR PROGRAM (NLP)

$$\boxed{\begin{aligned} &\underset{w \in \mathbb{R}^n}{\text{minimize}} \quad f(w) \\ &\text{subject to} \quad g(w) = 0 \\ &\qquad\qquad\quad h(w) \leq 0 \end{aligned} \quad \left\{ \begin{aligned} f &: \mathbb{R}^n \to \mathbb{R} \\ g &: \mathbb{R}^n \to \mathbb{R}^n \\ h &: \mathbb{R}^n \to \mathbb{R}^n \end{aligned} \right. \quad \text{ALL ASSUMED TO BE AT LEAST } C^2}$$

f - THE OBJECTIVE FUNCTION (SCALAR)

g - VECTOR OF EQUALITY CONSTRAINTS

h - VECTOR OF INEQUALITY CONSTRAINTS

} THEY CAN BE VECTOR VALUED

The set of points that satisfy all constraints : FEASIBLE SET

<u>Def (Feasible set)</u> THE <u>FEASIBLE SET Ω</u> IS THE SET

$$\Omega \equiv \{ w \in \mathbb{R}^n \mid g(w) = 0 , h(w) \leq 0 \}$$

SATISFY BOTH TYPE OF CONSTRAINTS

Interest is only on the points that satisfy the constraints and minimize the objective funtion

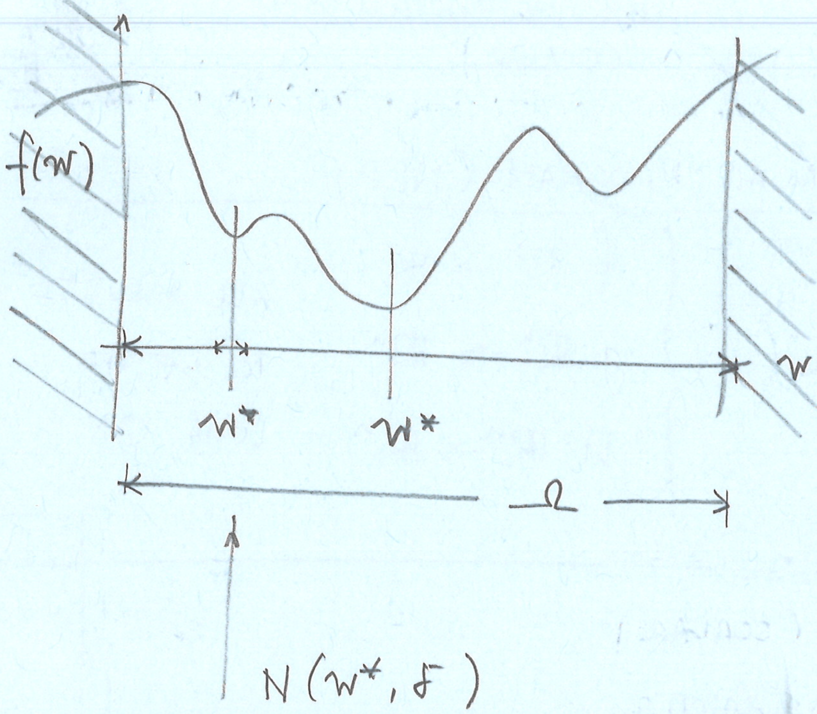
<u>Def. (Global minimum)</u> THE POINT $w^* \in \mathbb{R}^n$ IS A <u>GLOBAL MINIMIZER</u> IFF $w^* \in \Omega$ AND $\forall w \in \Omega$ WE HAVE THAT

$$f(w) \geq f(w^*)$$

The value $f(w^*)$ is the <u>global minimum</u>

→ THIS IS DIFFICULT TO FIND, IN GENERAL

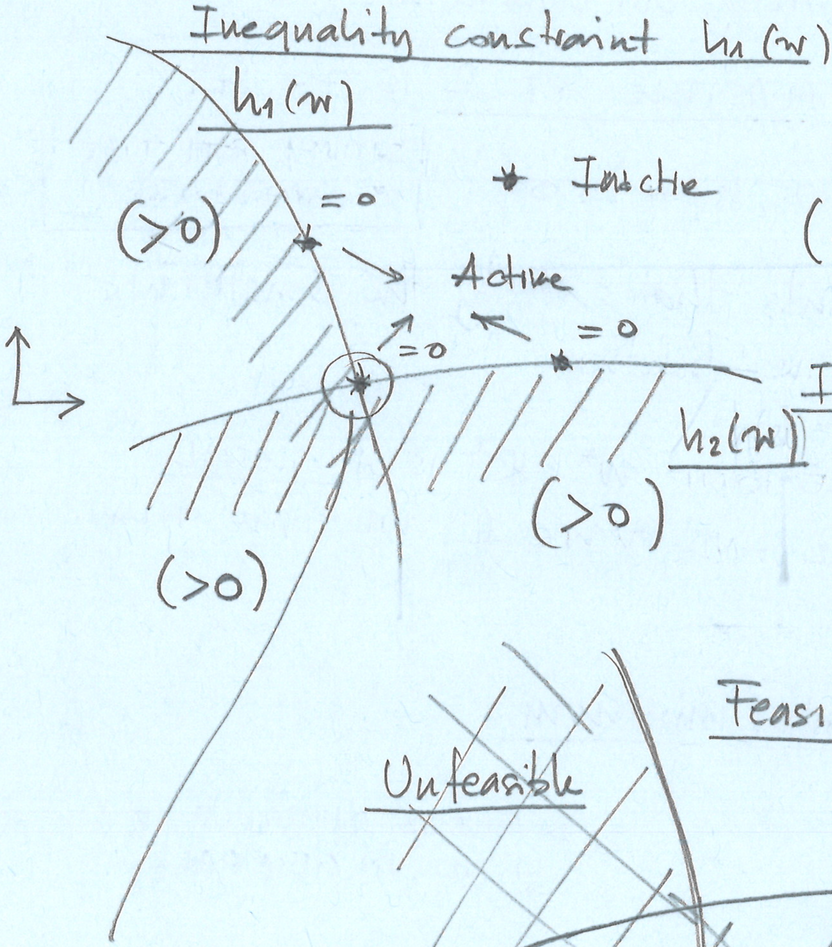
OPTIMIZATION IS THE TASK OF FINDING THE OPTIMAL POINT, w^*, WITHIN THE SET OF FEASIBLE POINTS



GLOBAL

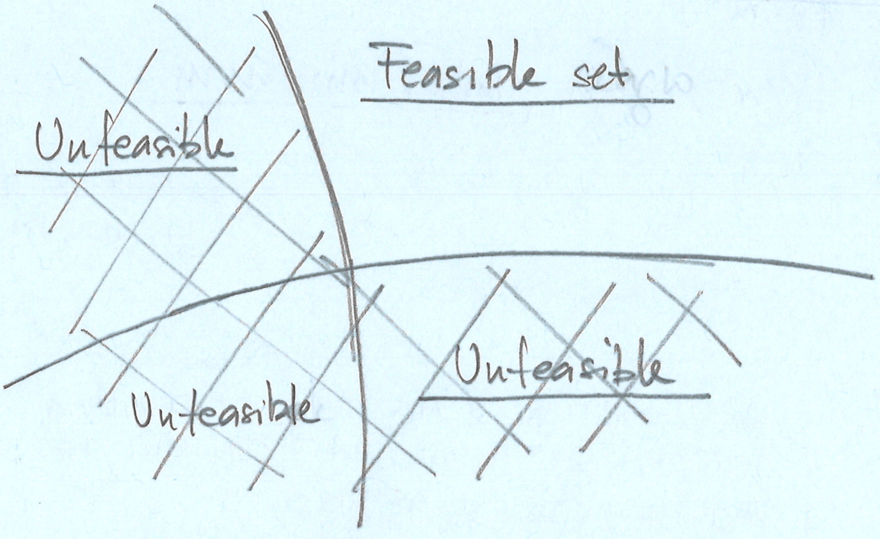
VS

LOCAL



$h_i(w) \leq 0$

$A(w) \subset \{1, 2, \dots, N_h\}$



Def. (LOCAL MINIMUM) THE POINT $w^* \in \mathbb{R}^n$ IS A LOCAL MINIMISER IFF $w^* \in \Omega$ AND THERE EXIST AN OPEN BALL $N(w^*, \varepsilon)$ CENTRED IN w^* AND WITH RADIUS ε SUCH THAT $\forall w \in \Omega \cap N$, WE HAVE THAT $f(w) \geq f(w^*)$

The value $f(w^*)$ is a local minimum.

THIS ARE EASIER TO FIND (IN GENERAL

Def. (Active constraints and active sets) AN INEQUALITY CONSTRAINT $h_i(w) \leq 0$ IS SAID TO BE ACTIVE AT $w^* \in \Omega$ IFF $h_i(w^*) = 0$ (OTHERWISE IT IS SAID TO BE INACTIVE)

THE INDEX SET $A(w^*) \subset \{1, 2, \dots, n_h\}$ OF ACTIVE INEQUALITY CONSTRAINT IS DENOTED AS THE ACTIVE SET

It often includes the equality constraints

Important classes of NLPs (overview)

- LINEAR OPTIMISATION (PROGRAMMING)
- QUADRATIC OPTIMIZATION (PROGRAMMING)
- CONVEX OPTIMISATION (PROGRAMMING)

REMEMBER THAT $h(w) = 0$ = $\begin{cases} h_1(w) \leq 0 \\ h_2(w) \leq 0 \\ \vdots \\ h_i(w) \leq 0 \\ \vdots \end{cases}$

(THE SAME IS ALSO VALID FOR $g(w) = 0$)

All functions, that is f, g, h, are all affine

linear + $\underbrace{\text{constant}}_{\text{it does not change the solution}}$

LINEAR OPTIMISATION (LINEAR PROGRAMMING)

$$\underset{w \in \mathbb{R}^n}{\text{minimize}} \quad f(w)$$

$$\text{subject to} \quad g(w) = 0$$

$$h(w) \leq 0$$

with f, g, h linear in w (affine)

$$\boxed{\begin{aligned} &\underset{w \in \mathbb{R}^n}{\text{minimize}} && c^T w \quad (+\text{ constants}) \\ &\text{subject to} && Aw - b = 0 \\ &&& Cw - d \leq 0 \end{aligned} \quad \text{with} \quad \begin{aligned} & c \in \mathbb{R}^n \\ & A \in \mathbb{R}^{N_g \times n}, \; b \in \mathbb{R}^{N_g} \\ & C \in \mathbb{R}^{N_h \times n}, \; d \in \mathbb{R}^{N_h} \end{aligned}}$$

constant wrt w

IF THE LINEAR PROGRAM HAS A SOLUTION AND IT IS NOT UNBOUNDED, ONE OPTIMAL SOLUTION IS ONE OF THE VERTICES OF THE POLYTOPE OF THE FEASIBLE POINTS

THE VERTICES CAN BE CALCULATED BY USING BASIS SOLUTION VECTORS, WITH BASIS OF ACTIVE INEQUALITY CONSTRAINTS

→ A FINITE NUMBER OF VERTICES EXIST

→ THE SIMPLEX ALGORITHM (SOLUTION COMPARISON)

B IS ALSO A SYMMETRIC MATRIX →

QUADRATIC OPTIMISATION

$$\underset{w \in \mathbb{R}^n}{\text{minimize}} \; f(w)$$

subject to $g(w) = 0$
$h(w) \le 0$

with f quadratic-linear
g, h affine (as in LP)

$\rightsquigarrow$ $\underset{w \in \mathbb{R}^n}{\text{minimise}} \; \frac{1}{2} w^T B w + c^T w$ (+ constant)

subject to $Aw - b = 0$
$Cw - d \le 0$

with $c \in \mathbb{R}^n$
$A \in \mathbb{R}^{N_g \times n}$, $b \in \mathbb{R}^{N_g}$
$C \in \mathbb{R}^{N_h \times n}$, $d \in \mathbb{R}^{N_h}$

MATRIX $B \in \mathbb{R}^{n \times n}$ IS KNOW AS THE HESSIAN MATRIX

- $\nabla^2 f(w) = B$

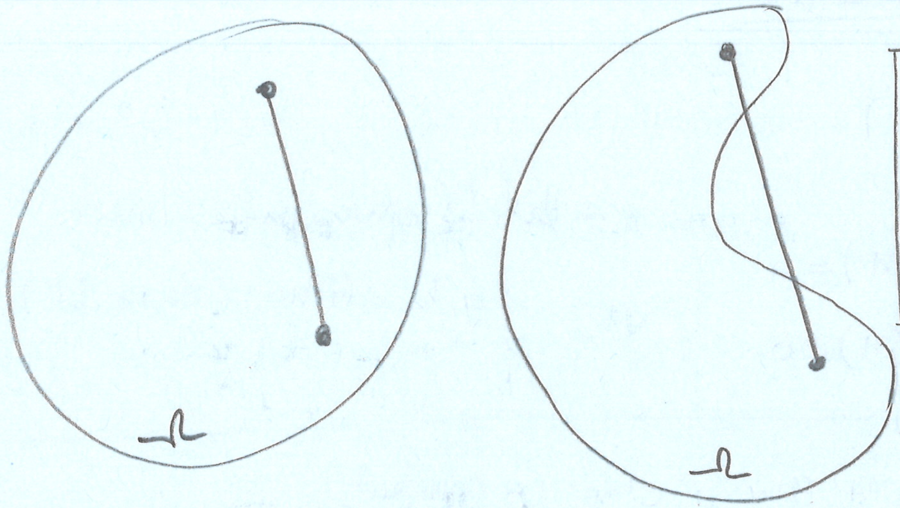
THE EIGENVALUES OF B DECIDE ON THE CONVEXITY OF THE QUADRATIC PROGRAM (POSSIBILITY TO SOLVE IN POLYNOMIAL TIME AND TO GLOBAL OPTIMALITY)

- $B \succcurlyeq 0$ (CONVEX)
- $B \succ 0$ (STRICTLY CONVEX) $\rightsquigarrow$ Unique minimizers

It is the second derivative of the cost function at any point (!)

WHEN $B = 0$, WE GET A LINEAR PROGRAM

Always symmetric for twice differentiable continuos $f(w)$

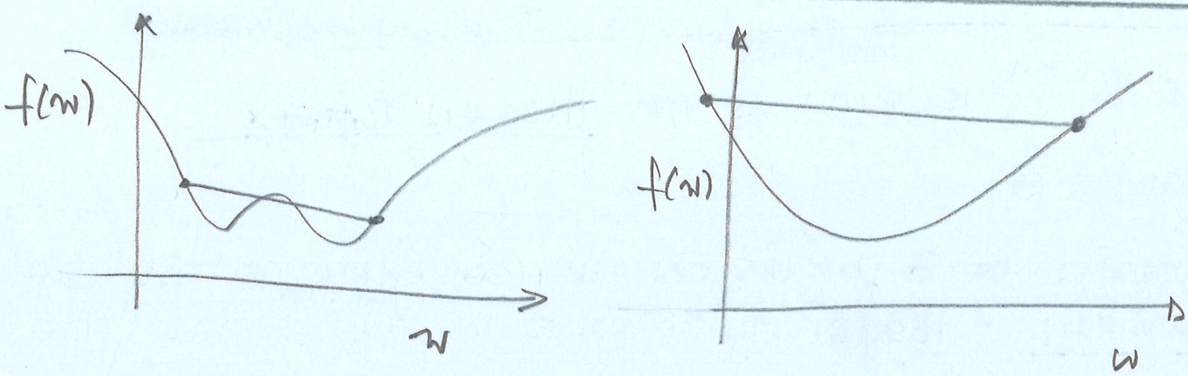


CONVEX SET

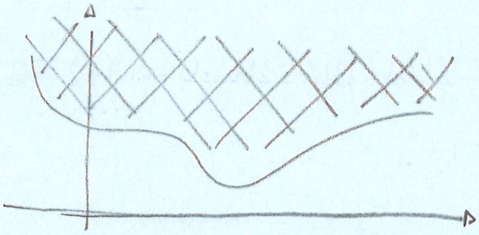
VS

NON CONVEX SET

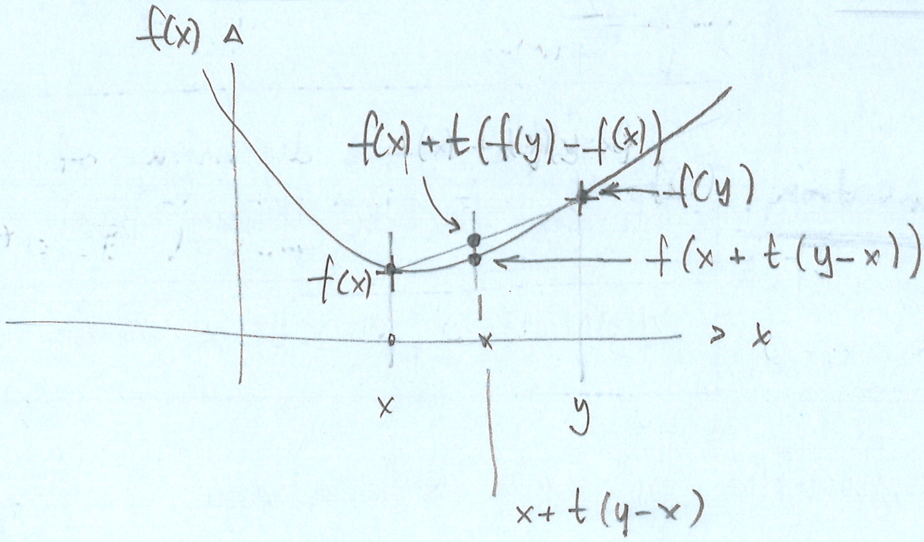
The all line is / is not contained in the set



CONVEX FUNCTION VS NON CONVEX FUNCTION



This is also known as the epigraph of the function (it is a set)



CONVEX OPTIMIZATION

Informally, a set is convex if all connecting lines belong to the set

Def. (Convex set) A SET $A \in \mathbb{R}^n$ IS CONVEX IF

$$\forall (x,y) \in \Omega, t \in [0,1], \text{ WE HAVE THAT } x + t(y-x) \in \Omega \text{ FOR ALL } t$$

Informally, a function is convex if all secants are above the graph

Def. (Convex function) A FUNCTION $f : \Omega \to \mathbb{R}$ IS CONVEX IF

Ω IS CONVEX AND IF $\forall (x,y) \in \Omega$, $t \in [0,1]$ WE HAVE

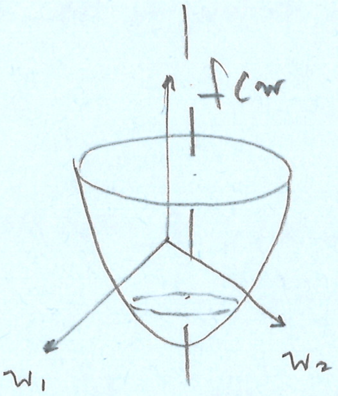
$$f(x + t(y-x)) \leq f(x) + t\,[f(y) + f(x)]$$

Def. (Concave function) A FUNCTION $f : \Omega \to \mathbb{R}$ IS CONCAVE IF FUNCTION $(-f)$ IS CONVEX

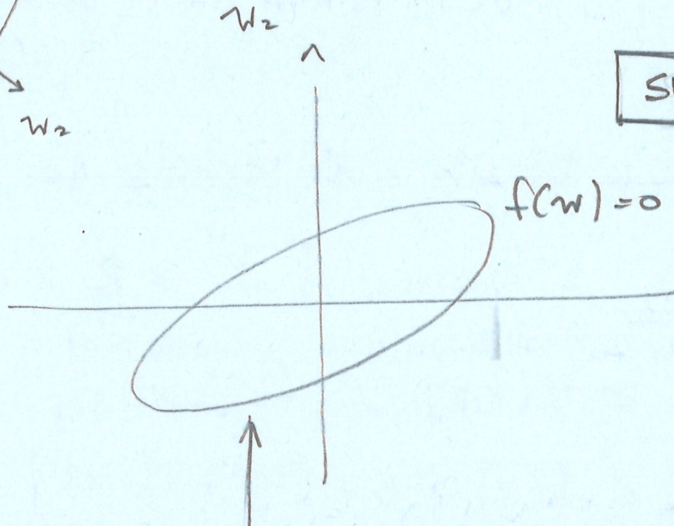
We can show that the feasible set of a NLP is convex if function g is affine and the functions h_i are convex

Th. (Convexity of sublevel sets) A SUBLEVEL SET $\{x \in \Omega \mid h(x) \leq 0\}$ OF A CONVEX FUNCTION $h : \Omega \to \mathbb{R}$ IS CONVEX

Def. (Convex optimisation) AN OPTIMIZATION PROBLEM WITH CONVEX FEASIBLE SET Ω AND CONVEX OBJECTIVE FUNCTION $f : \Omega \to \mathbb{R}$ IS CALLED A <u>CONVEX OPTIMIZATION PROBLEM</u>

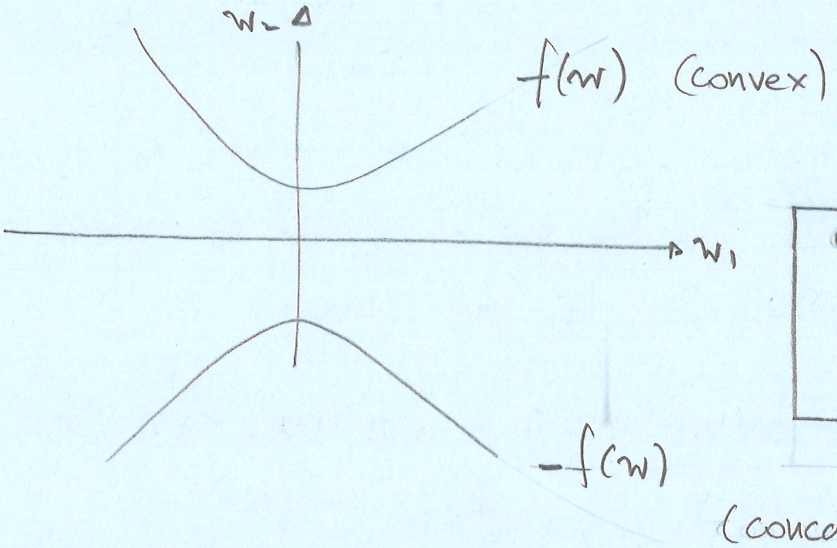


SUBLEVEL SET



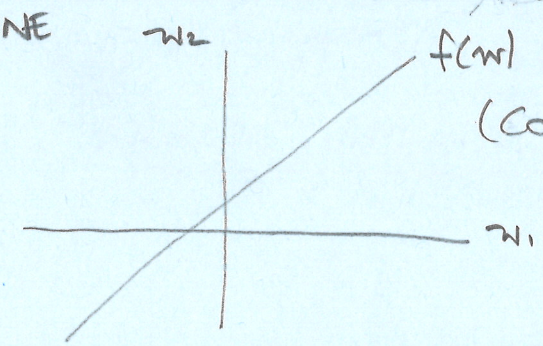
Sublevel set, at $f(w) = 0$

IT IS CONVEX IF $f(w)$ IS CONVEX



CONVEX vs CONCAVE FUNCTIONS

A FUNCTION THAT IS BOTH CONCAVE AND CONVEX IS SAID TO BE AFFINE



CERTAIN LINEAR TRANSFORMATIONS PRESERVE THE CONVEXITY OF FUNCTIONS AND SETS (SEE BENTAL/NEMIROVSKI AND BOYD/VANDERBERGHE)

Def. (General Inequality for Symmetric matrices)

LET $B = B^T$, $B \in \mathbb{R}^{n \times n}$ (Symmetric, square)

WE SAY $B \geq 0$ IFF B is POSITIVE SEMIDEFINITE, THAT IS

$$z^T B z \geq 0 \quad \text{for all } z \in \mathbb{R}^n$$

EQUIVALENTLY ALL REAL EIGENVALUES OF B (SYMMETRIC) ARE NON-NEGATIVE

$$B \geq 0 \iff \min \operatorname{eig}(B) \geq 0$$

NON-NEGATIVE

FOR SUCH MATRIXES, WE WRITE $B \geq A$ IFF $A - B \geq 0$

$B \leq A$ IFF $B \geq A$

$B > 0$ IFF B is POSITIVE DEFINITE, $z^T B z \geq 0$ for all $z \in \mathbb{R}^n \setminus \{\emptyset\}$

$$B > 0 \iff \min \operatorname{eig}(B) > 0$$

STRICTLY POSITIVE

Th. (Convexity of C^2 functions) LET $f : \Omega \to \mathbb{R}$ BE TWICE DIFFERENTIABLE WITH CONTINUOUS DERIVATIVES AND LET Ω BE CONVEX AND OPEN

WE SAY THAT f IS CONVEX IFF $\forall x \in \Omega$, THE HESSIAN IS PSD

$$\underbrace{\nabla^2 f(w) \succeq 0}, \quad \text{for all } x \in \Omega$$

$\rightarrow$ Think of it as positive 'curvature'

1D 2D 3D

WE ARE NOT GOING TO DERIVE THE CONDITIONS
→ REFER TO A TEXTBOOK FOR DETAILS

The first order condition can be formulated only if the technical 'constraint classification' is satisfied

LINEAR INDEPENDENCE OF VECTORS : DEFINITION

FULL RANK EQUIVALENCE

FIRST ORDER OPTIMALITY CONDITIONS

IF A FEASIBLE POINT $w^* \in \Omega$ DOES NOT SATISFY FIRST ORDER $\overset{\text{NECESSARY}}{\quad}$ OPTIMALITY CONDITIONS THEN w^* CANNOT BE A LOCAL MINIMIZER

$\rightsquigarrow$ IF IT DOES SATISFY FIRST-ORDER NECESSARY CONDITIONS, THEN IT IS A VALID CANDIDATE TO BE A LOCAL MINIMISER

If the problem is convex, these conditions are even sufficient to guarantee that the minimizer is actually a global one

$\rightarrow$ MOST ALGORITHMS FOR NONLINEAR OPTIMIZATION SEARCH FOR SUCH POINTS

FIRST ORDER NECESSARY OPT. CONDITIONS CAN BE FORMULATED IF THE GRADIENTS $\nabla g_i(w^*)$ AND $\nabla h_i(w^*)$ ARE LINEARLY INDEPENDENT $(i \in A(w^*))$

Def(.) THE LINEAR INDEPENDENCE CONSTRAINT $\overset{\text{QUALIFICATION (LICQ)}}{\quad}$ VALIDITY HOLDS AT $w^* \in \Omega$ IFF ALL VECTORS $\nabla g_i(w^*)$ FOR $i \in \{1, 2, \dots, N_g\}$ AND $\nabla h_i(w^*)$ FOR $i \in A(w^*)$ ARE LINEARLY INDEPENDENT (ONLY THOSE THAT ARE ACTIVE MATTER)

Let us combine all active inequalities with all equalities, by stacking all functions in a column vector

$$\tilde{g}(w) = \left[\begin{array}{c} g(w) \\ \hline h_i(w),\ i \in A(w^*) \end{array}\right] \rightsquigarrow \text{then } \frac{\partial \tilde{g}(w)}{\partial w}(w^*) \ \overset{\text{ROW}}{\text{FULL RANK}}$$

$$\hookrightarrow \begin{bmatrix} \nabla g(w)^T \\ \vdots \\ \nabla g(w)^T \\ \nabla h(w)^T \end{bmatrix}$$

$\rightarrow$ LICQ IS DEFINED ONLY AT FEASIBLE POINTS

∇ GRADIENTS

∇ GRADIENTS

(λ, μ) ARE CALLED LAGRANGIAN MULTIPLIERS
(OR DUAL VARIABLES)

$\lambda \in \mathbb{R}^{N_g}$

$\mu \in \mathbb{R}^{N_h}$

---

* The first condition:
  - $\nabla f(w^*) + \nabla g(w^*)\lambda^* + \nabla h(w^*)\mu^* = 0$
  - The weighted sum of the gradients, with weights the respective multipliers must be equal to zero

* The second condition:
  - $\nabla g(w^*) = 0$
  - The equality constraints must be satisfied (w^* is a feasible sol.)

* The third condition:
  - $\nabla h_i(w^*) = 0$ for all $i = 1, \dots N_h$
  - The inequality constraints must be satisfied (w^* is a feasible sol.)

THE KARUSH KUHN TUCKER OPTIMALITY CONDITION

Th (KKT Conditions) IF w^* IS A LOCAL MINIMIZER OF THE NLP AND THE LINEAR INDEPENDENCE CONSTRAINT QUALIFICATION CONDITION IS SATISFIED AT w^* THEN THERE EXIST MULTIPLIER VECTORS $\lambda^* \in \mathbb{R}^{N_g}$ AND $\mu^* \in \mathbb{R}^{N_h}$ SUCH THAT

KKT CONDITIONS

$$\left\{\begin{array}{rl}\nabla f(w^*) + \nabla g(w^*)\lambda^* + \nabla h(w^*)\mu^* &= 0 \\ g(w^*) &= 0 \\ h_i(w^*) &\leq 0 \\ \mu_i^* &\geq 0 \\ \underbrace{\mu_i^* h_i(w^*)}, \; i = 1,2,\dots,N_h &= 0\end{array}\right.$$

THE WEIGHTED SUM OF THE GRADIENTS ADDS TO ZERO

THEY ARE SATISFIED IF FEASIBLE POINT w^*

$i = 1, \dots, N_h$

↳ CONDITIONS FOR THESE BE ZERO

↳ FIRST ORDER NECESSARY CONDITIONS (FONC)

They become also sufficient if the problem is convex and $w^* \in \Omega$ is a global minimiser

$\exists \lambda^*, \mu^*$ such that KKT hold

Th. ( ) CONSIDER A CONVEX NONLINEAR PROGRAM AND A POINT w^* AT WHICH LICQ HOLD, THEN,

$\rightsquigarrow$ w^* IS A GLOBAL MINIMISER $\iff$ $\exists \lambda^*, \mu^*$ such that the KKT conditions hold

The Lagrangian function depends on the <u>primal variables</u> (the decision variables, w) and on the <u>dual variables</u> (the multipliers, λ and μ)

→ IT IS A MIXTURE OF THE OBJECTIVE FUNCTION AND THE COSTRAINTS (EQUALITY AND INEQUALITY)

The KKT conditions are FIRST-ORDER NECESSARY CONDITIONS FOR OPTIMALITY for constrained optimization problems

→ They correspond to $\nabla f(w^*) = 0$ for unconstrained optimisation problems

→ (THIS IS A SUBCASE)

$$f(w) = f(w_1, w_2, \_\_\_\_, w_N)$$

$$\nabla f(w) = \begin{bmatrix} \partial f / \partial w_1 \\ \partial f / \partial w_2 \\ \vdots \\ \partial f / \partial w_N \end{bmatrix}$$

(Gradient vector)

$$g(w) = \begin{bmatrix} g_1(w_1, w_2, \_\_\_\_, w_N) \\ g_2(w_1, w_2, \_\_\_\_, w_N) \\ \vdots \\ g_d(w_1, w_2, \_\_\_\_, w_N) \end{bmatrix}$$

$$\nabla g(w) = \begin{bmatrix} \partial g_1/\partial w_1 & \partial g_1/\partial w_2 & \cdots & \partial g_1/\partial w_N \\ \partial g_2/\partial w_1 & \partial g_2/\partial w_2 & \cdots & \partial g_2/\partial w_N \\ \vdots & & \ddots & \vdots \\ \partial g_d/\partial w_1 & \partial g_d/\partial w_2 & \cdots & \partial g_d/\partial w_N \end{bmatrix}^T$$

(Transpose of the Jacobian)

$$h(w) = \begin{bmatrix} h_1(w_1, w_2, \_\_\_\_, w_N) \\ h_2(w_1, w_2, \_\_\_\_, w_N) \\ \vdots \\ h_d(w_1, w_2, \_\_\_\_, w_N) \end{bmatrix}$$

DIMENSIONS !!

$$\nabla h(w) = ?$$

No 6a

The Lagrangian function plays an important role in general nonlinear optimisation, including convex problems

THE LAGRANGIAN FUNCTION (IN OPTIMISATION)

$$\underbrace{\mathcal{L}(w,\lambda,\mu)} = f(w) + \lambda^T g(w) + \mu^T h(w)$$

LINEAR COMBINATION OF THE EQUALITY CONSTRAINTS

LINEAR COMBINATION OF INEQUALITY CONSTRAINTS

→ IT CAN BE SEEN AS A SHORT-HAND FORMULATION OF KKT CONDITIONS

⇝ $\nabla_w \mathcal{L}(w^*, \lambda^*, \mu^*) = 0$ (1st KKT CONDITION)

IN THE SPECIAL CASE THAT THERE ARE NO INEQUALITY CONSTRAINTS $h(w)$ WE HAVE THAT $\underbrace{\mathcal{L}(w,\lambda)=0,\ g(w)=0}$ AND WE CAN WRITE

- $\nabla_w \mathcal{L}(w^*, \lambda^*) = 0$
- $g(w^*) = 0$

AS w^* IS A FEASIBLE POINT

ALL OF THE KKT CONDITIONS (for equality constraint optimiz.)

→ $N_g + N$ CONDITIONS (Eqs)

→ $N_g + N$ VARIABLES (w, λ)

SQUARE NON LINEAR SYSTEM

⇝ ROOT FINDING PROBLEM

ONE KKT CONDITION STATES THAT $\mu_i \geq 0$ WHILE THE SIGN OF λ_i IS ARBITRARY ... HOW CAN THIS BE MOTIVATED?

POSITIVE FOR EQUALITY

ARBITRARY FOR INEQUALITY

FOR CONVEX PROBLEMS

In the Lagrangian $\mu^T h(w)$ is a linear comb combination of convex functions

$g(w)$ are affine/convex

THE LAGRANGIAN IS ALSO CONVEX

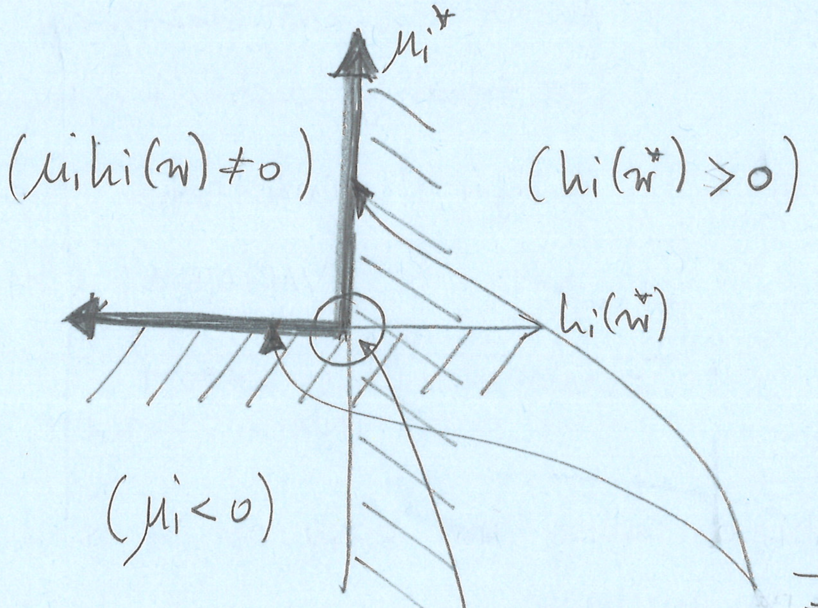
THE LAST THREE KKT CONDITIONS → <u>COMPLEMENTARITY CONDITIONS</u>

$$h_i(w^*) \le 0$$

$$\mu_i^* \ge 0$$

$$\mu_i^* h_i(w^*) = 0$$

They are often denoted as <u>COMPLEMENTARITY CONDITIONS</u>



THIS IS VALID FOR EACH INDEX i

THE TWO LINES ARE THE ONLY ALLOWED CONFIGURATION

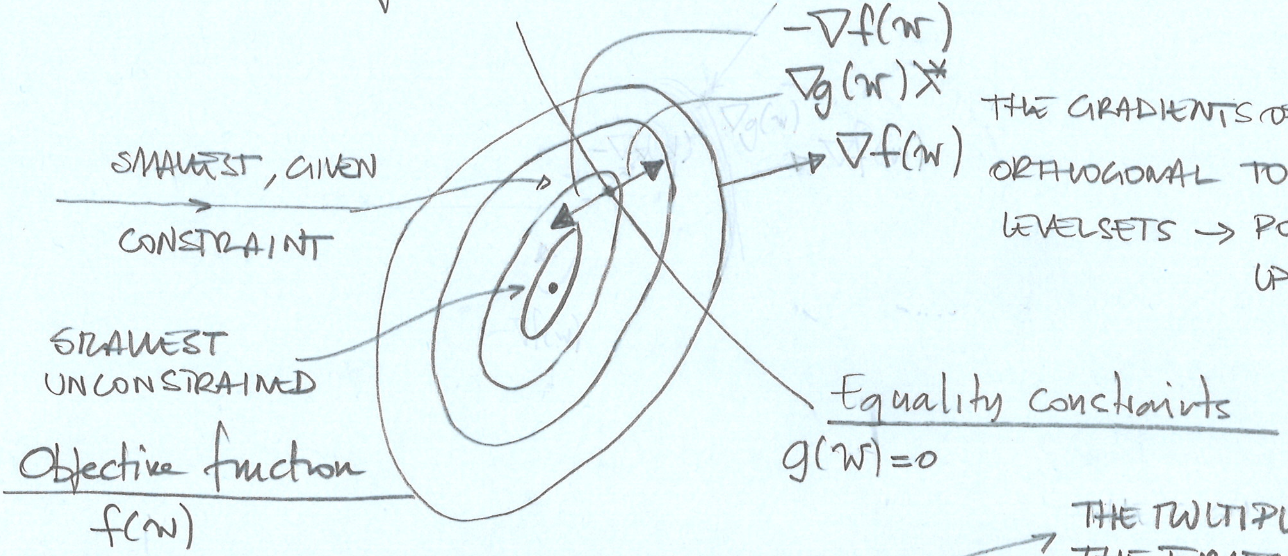
THE SET IS NOT DIFFERENTIABLE

KINKY POINT

<u>Def</u> (Strict complementarity)

HOW TO READ THE KKT CONDITIONS

1/ CONSIDER THE OBJECTIVE FUNCTION $f(w)$, ILLUSTRATED VIA ITS LEVELSETS



THE GRADIENTS OF f ARE ORTHOGONAL TO THE LEVELSETS → POINTING UPWARDS

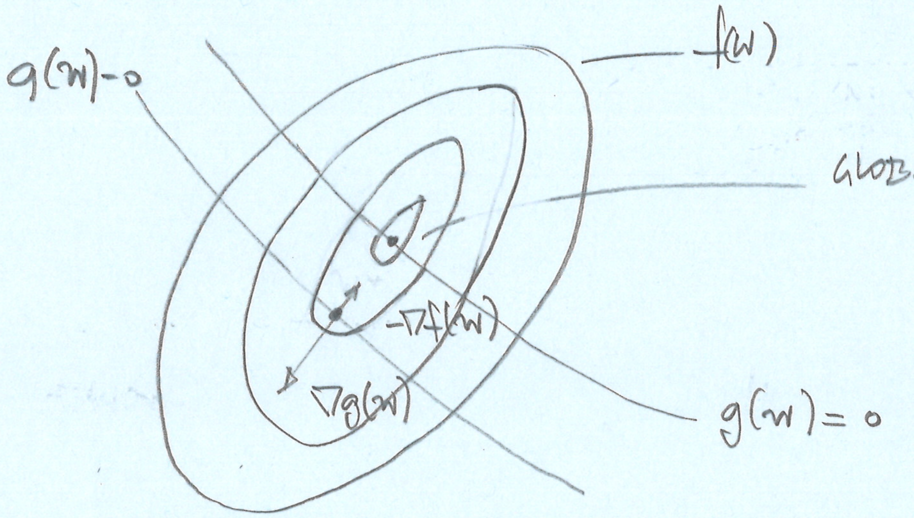
THE MULTIPLIER MAKES THE FORCES COMPARABLE (THE SIGN RETAINS)

AT w^* WE HAVE THAT $-\nabla f(w^*) = \lambda^* g(w^*)$

→ the two gradients are equal, up to some factor (multiplier)

→ THE TWO VECTOR BALANCE, AT w^*

2+3 / CONSIDER THE OBJECTIVE FUNCTION $f(w)$, AGAIN VIA LEVELSETS



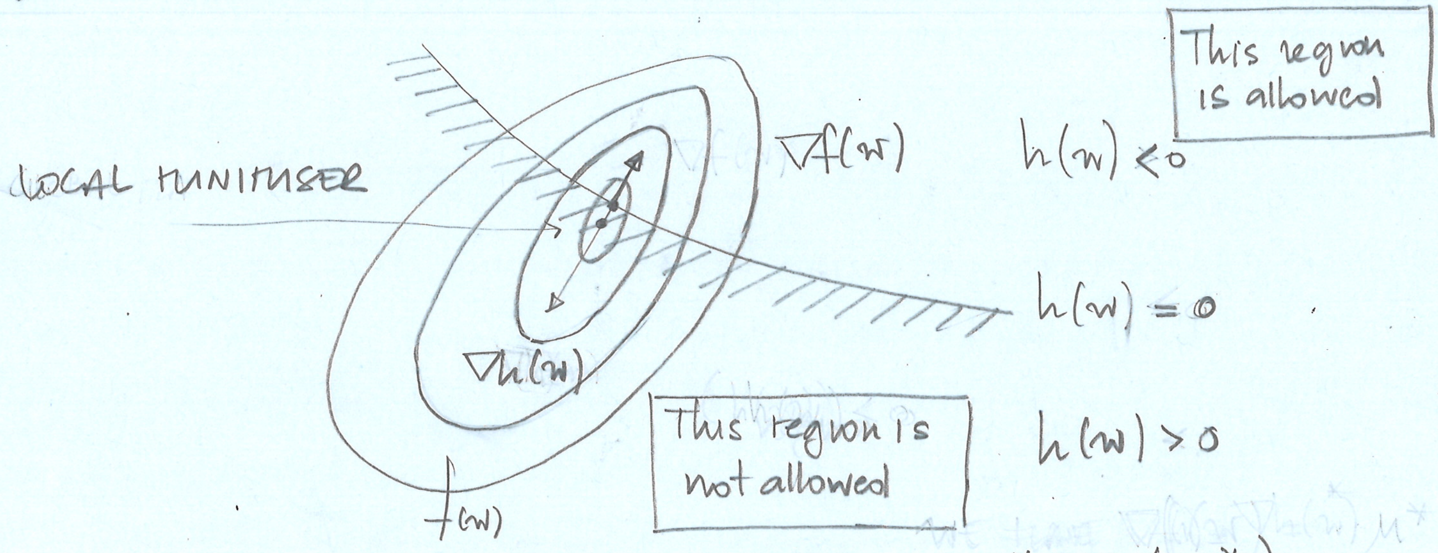
Equality contraint at w^*

THERE IS NO NEED TO BALANCE THE GRADIENTS

$$\underbrace{-\nabla f(w^*)}_{=0} = \lambda^* g(w^*)$$

$=0$ AS THIS QUANTITY IS ZERO, ALSO λ^* IS ZERO

4/ CONSIDER THE OBJECTIVE FUNCTION $f(w)$, VIA ITS LEVEL SETS



WE HAVE AGAIN THAT AT w^*, $\nabla f(w^*) + \lambda^* \nabla h(w^*) = 0$

→ THE VECTOR EVEN OUT

5/

In this case, the multiplier is equal to zero