

NUMERICAL OPTIMIZATION

CONSIDER THE STANDARD NONLINEAR PROGRAM (NLP)

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

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

OBJECTIVE FUNCTION OF w
(SCALAR)

2b

$$\left\{ \begin{array}{l} f: \mathbb{R}^n \rightarrow \mathbb{R} \\ g: \mathbb{R}^n \rightarrow \mathbb{R}^m \\ h: \mathbb{R}^n \rightarrow \mathbb{R}^n \end{array} \right.$$

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

Def. (Feasible set) THE FEASIBLE SET Ω IS THE SET

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

SATISFY BOTH TYPE
OF CONSTRAINTS

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

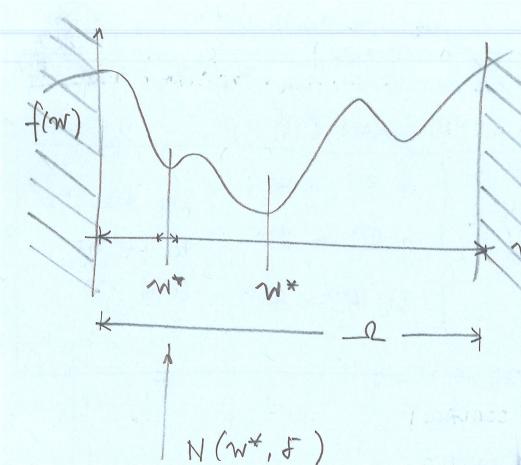
Def. (Global minimum) THE POINT $w^* \in \Omega$ IS A GLOBAL
MINIMIZER IFF $w^* \in \Omega$ AND $\forall w \in \Omega$ WE HAVE THAT
 $f(w) \geq f(w^*)$

The value $f(w^*)$ is the global minimum

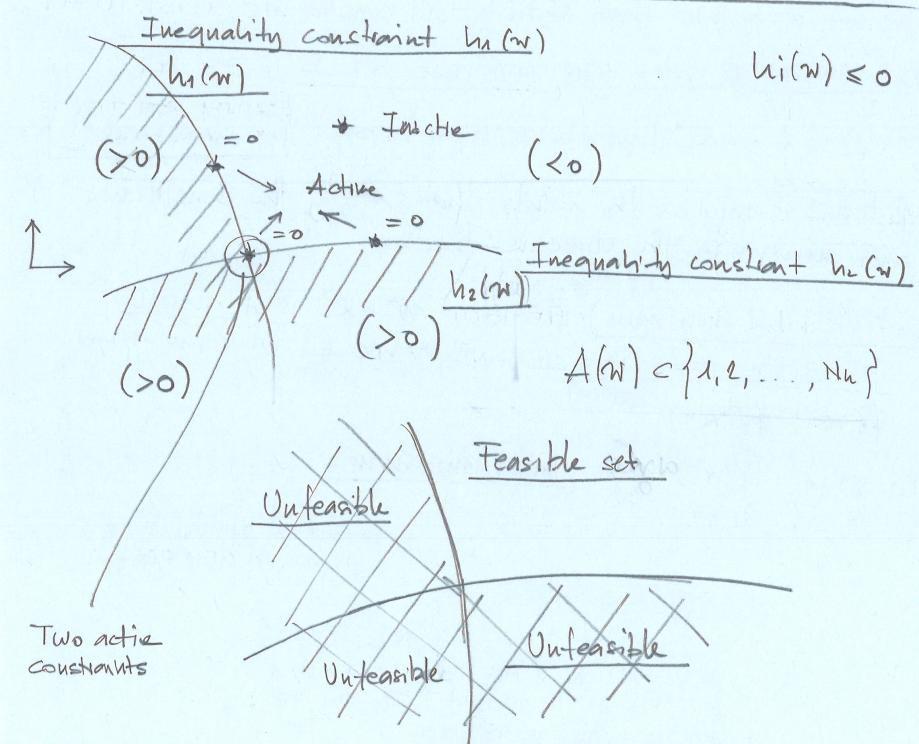
THIS IS DIFFICULT TO
FIND, IN GENERAL

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

NO 1



GLOBAL
vs
LOCAL



Def. (local minimum) THE POINT $w^* \in \mathbb{R}^n$ IS A LOCAL MINIMUM IFF $w^* \in \Omega$ AND THERE EXIST AN OPEN BALL $N(w^*, \varepsilon)$ CENTERED IN w^* AND WITH RADIUS ε SUCH THAT $\forall w \in \Omega \cap N, w$ 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 $h(w) = 0 \Leftrightarrow \begin{cases} h_1(w) \leq 0 \\ h_2(w) \leq 0 \\ \vdots \\ h_i(w) \leq 0 \\ \vdots \\ h_n(w) \leq 0 \end{cases}$

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

NO 2

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

↓
linear + constant
it does not change
the solution

LINEAR OPTIMISATION (LINEAR PROGRAMMING)

minimize $f(w)$

wrt

subject to $g(w) = 0$

$h(w) \leq 0$

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

→ minimize $C^T w$ (+ constraints)

wrt $w \in \mathbb{R}^n$

subject to $Aw - b = 0$

$Cw - d \leq 0$

constant wrt w

with $C \in \mathbb{R}^{n \times n}$
 $A \in \mathbb{R}^{N_1 \times n}, b \in \mathbb{R}^{N_1}$
 $C \in \mathbb{R}^{N_2 \times n}, d \in \mathbb{R}^{N_2}$

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

(STABILITY)

QUADRATIC OPTIMISATION

minimize $f(w)$

$w \in \mathbb{R}^n$

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

with $\begin{cases} f \text{ quadratic-linear} \\ g, h \text{ affine (as in LP)} \end{cases}$

→ minimize $\frac{1}{2} w^T B w + c^T w$ (+ constant)

subject to $Aw - b = 0$

$Cw - d \leq 0$

with $\begin{cases} c \in \mathbb{R}^n \\ A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \\ C \in \mathbb{R}^{k \times n}, d \in \mathbb{R}^k \end{cases}$

MATRIX $B \in \mathbb{R}^{n \times n}$ IS KNOWN 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)

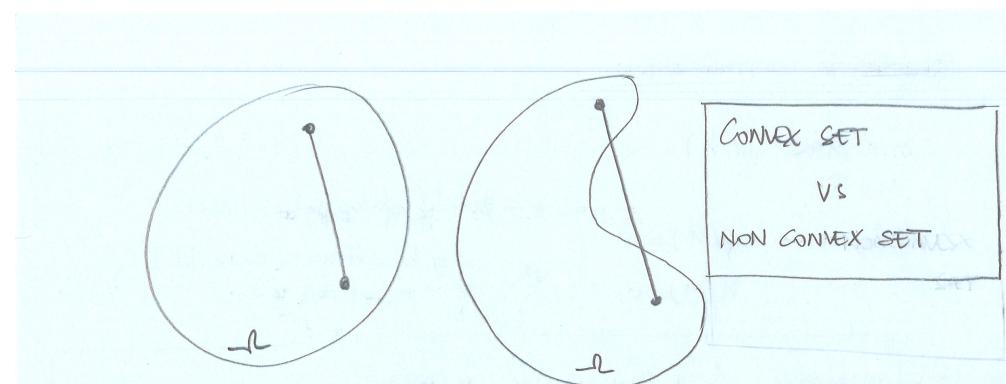
• $B \geq 0$ (CONVEX)

• $B > 0$ (STRICTLY CONVEX) → Unique minimizers

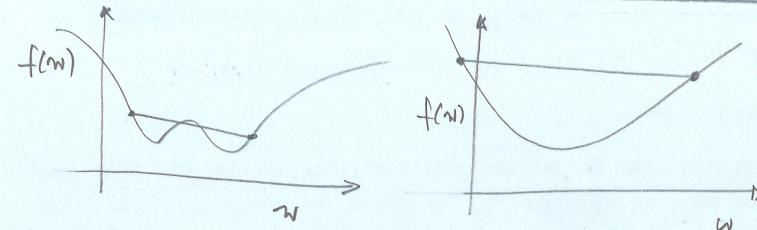
→ It is the second derivative of the cost function at any point (!)

→ WHEN $B=0$, WE GET A LINEAR PROGRAM

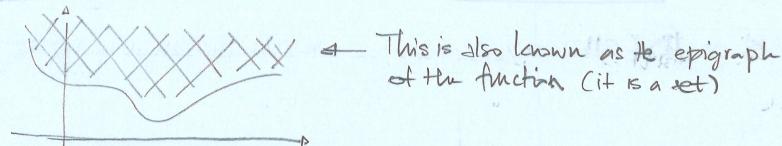
Always symmetric for twice differentiable continuous $f(w)$



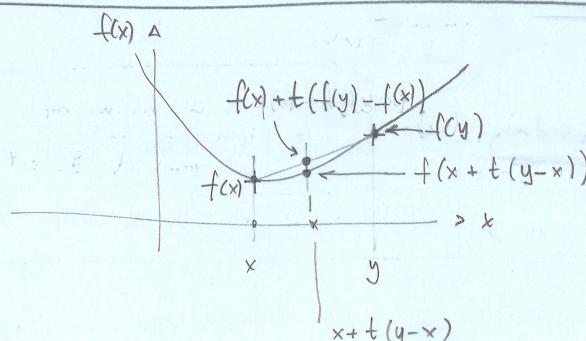
The all line is/ is not contained in the set



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

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

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

Def. (Convex function) A FUNCTION $f: \mathcal{L} \rightarrow \mathbb{R}$ IS CONVEX IF

\mathcal{L} IS CONVEX AND IF $\forall (x, y) \in \mathcal{L}, t \in [0, 1]$ WE HAVE

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

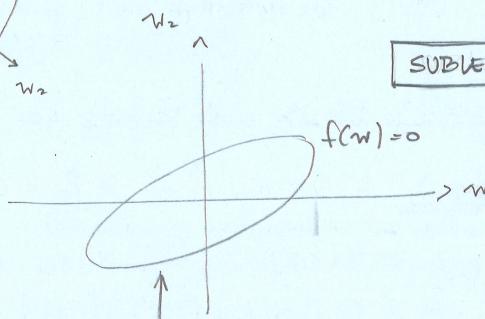
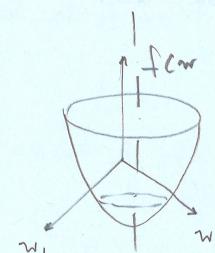
Def. (Concave function) A FUNCTION $f: \mathcal{L} \rightarrow \mathbb{R}$ IS CONCAVE IF

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

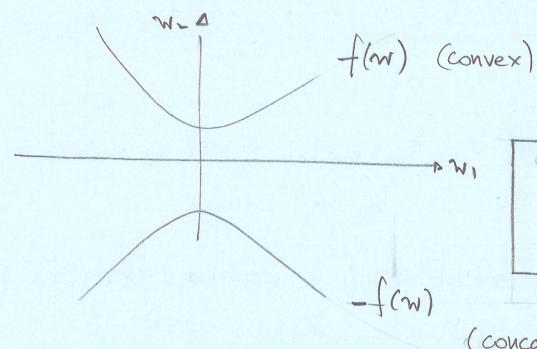
Th. (Convexity of Sublevel sets) A SUBLEVEL SET $\{x \in \mathcal{L} \mid h(x) \leq 0\}$ OF A CONVEX FUNCTION $h: \mathcal{L} \rightarrow \mathbb{R}$ IS CONVEX

Def (Convex optimization) AN OPTIMIZATION PROBLEM WITH CONVEX FEASIBLE SET \mathcal{L} AND CONVEX OBJECTIVE FUNCTION $f: \mathcal{L} \rightarrow \mathbb{R}$ IS CALLED A CONVEX OPTIMIZATION PROBLEM



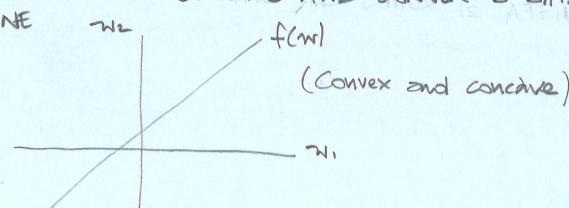
SUBLEVEL SET

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



(Convex and concave)

CERTAIN LINEAR TRANSFORMATIONS PRESERVE THE CONNEXITY OF FUNCTIONS AND SETS (SEE BENTAL/NETIPOVSKI 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 SEMI-DEFINITE, THAT IS
 $\mathbf{z}^T B \mathbf{z} \geq 0$ FOR ALL $\mathbf{z} \in \mathbb{R}^n$

EQUivalently ALL REAL EIGENVALUES OF B (symmetric)
ARE NON-NEGATIVE

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

↑
NON-NEGATIVE

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

$B > 0$ IFF B IS POSITIVE DEFINITE, $\mathbf{z}^T B \mathbf{z} > 0$ FOR ALL $\mathbf{z} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$

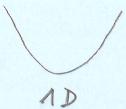
$$B > 0 \iff \min \text{eig}(B) > 0$$

↑
STRICTLY POSITIVE

Th (Convexity of C^2 functions) Let $f: \mathbb{R} \rightarrow \mathbb{R}$ 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
 $\nabla^2 f(x) \geq 0$, for all $x \in \Omega$

think of it as positive 'curvature'



...

WE ARE NOT GOING TO DERIVE THE CONDITIONS
AND 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 $w^* \in \mathcal{L}$ DOES NOT SATISFY FIRST ORDER OPTIMALITY CONDITIONS THEN w^* CANNOT BE A LOCAL MINIMIZER

AND 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

→ 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 $(i \in A(w^*))$

QUALIFICATION (LICQ)

Def(•) THE LINEAR INDEPENDENCE CONSTRAINT VALIDITY HOLDS AT $w^* \in \mathcal{L}$ 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, try stacking all functions in a column vector

$$\tilde{g}(w) = \begin{bmatrix} g(w) \\ h_i(w), i \in A(w^*) \end{bmatrix} \text{ then } \frac{\partial \tilde{g}(w)}{\partial w} (w^*) \text{ FULL RANK}$$

LICQ IS DEFINED ONLY AT FEASIBLE POINTS

V GRADIENTS

D GRADIENTS

NOS

(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 :

$$-\nabla h_i(w^*) = 0 \text{ for all } i=1, \dots, N_h$$

- 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 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

$$\left. \begin{array}{l} \nabla f(w^*) + \nabla g(w^*) \lambda^* + \nabla h(w^*) \mu^* = 0 \\ g(w^*) = 0 \\ h_i(w^*) \leq 0 \\ \mu_i^* \geq 0 \\ \sum \mu_i^* h_i(w^*) = 0 \end{array} \right\} \begin{array}{l} \text{THE WEIGHTED SUM OF THE GRADIENTS ADDS TO ZERO} \\ \text{THEY ARE SATISFIED IF FEASIBLE POINT } w^* \\ i=1, \dots, n_h \\ \text{CONDITIONS FOR THESE BELOW} \end{array}$$

↳ FIRST ORDER NECESSARY CONDITIONS (FONC)

$\exists \lambda^*, \mu^*$ such that KKT hold
They become also sufficient if the problem is convex and $w^* \in L$ is a global minimiser

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

$\Rightarrow w^*$ is a global minimiser $\Leftrightarrow \exists \lambda^*, \mu^*$ such that the KKT conditions hold

The Lagrangian function depends on the primal variables (the decision variables, w) ~~or the dual variables~~ and on the dual variables (the multipliers, λ and μ)

→ IT IS A MIXTURE OF THE OBJECTIVE FUNCTION AND THE CONSTRAINTS (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)

$$f(\mathbf{w}) = f(w_1, w_2, \dots, w_n)$$

$$\nabla f(\mathbf{w}) = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{bmatrix}$$

(Gradient vector)

$$g(\mathbf{w}) = \begin{bmatrix} g_1(w_1, w_2, \dots, w_n) \\ g_2(w_1, w_2, \dots, w_n) \\ \vdots \\ g_d(w_1, w_2, \dots, w_n) \end{bmatrix}$$

$$\nabla g(\mathbf{w}) = \begin{bmatrix} \frac{\partial g_1}{\partial w_1}, \frac{\partial g_1}{\partial w_2}, \dots, \frac{\partial g_1}{\partial w_n} \\ \frac{\partial g_2}{\partial w_1}, \frac{\partial g_2}{\partial w_2}, \dots, \frac{\partial g_2}{\partial w_n} \\ \vdots \\ \frac{\partial g_d}{\partial w_1}, \frac{\partial g_d}{\partial w_2}, \dots, \frac{\partial g_d}{\partial w_n} \end{bmatrix}^T$$

(Transpose of
the Jacobian)

$$h(\mathbf{w}) = \begin{bmatrix} h_1(w_1, w_2, \dots, w_n) \\ h_2(w_1, w_2, \dots, w_n) \\ \vdots \\ h_d(w_1, w_2, \dots, w_n) \end{bmatrix}$$

DIMENSIONS !!

$$\nabla h(\mathbf{w}) = ?$$

No 6a

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

THE LAGRANGIAN FUNCTION (IN OPTIMISATION)

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

LINESAR
LINEAR COMBINATION OF THE EQUALITY CONSTRAINTS

LINEAR COMBINATION OF INEQUALITY CONSTRAINTS

SHORT HAND

IT CAN BE SEEN AS A FORMULATION OF KKT CONDITIONS

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

IN THE SPECIAL CASE THAT THERE ARE NO INEQUALITY CONSTRAINTS $h(w)$ WE HAVE THAT $\mathcal{L}(w, \lambda) = 0$, $g(w) = 0$, 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.)
 $\rightarrow N_g + N$ CONDITIONS (Eqs)
 $\rightarrow N_g + N$ VARIABLES (w, λ)

SQUARE NON LINEAR SYSTEM
 \rightarrow ROOT FINDING PROBLEM

THAT

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

POSITIVE FOR EQUALITY

FOR CONVEX PROBLEMS

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

$g(w)$ is a convex function

ARBITRARY FOR INEQUALITY

NOT

THE LAGRANGIAN IS ALSO CONVEX

THE LAST THREE KKT CONDITIONS \rightarrow COMPLEMENTARITY CONDITIONS

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

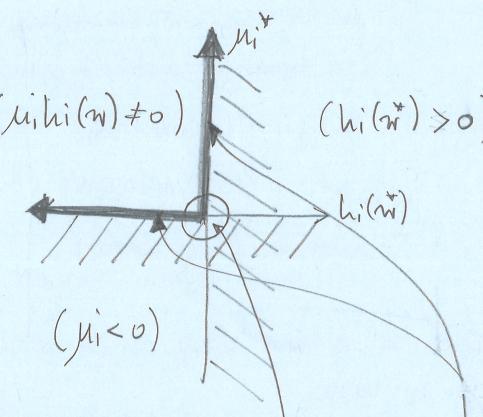
$$\lambda_i^* \geq 0$$

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

They are often denoted as

COMPLEMENTARITY CONDITIONS

THIS IS VALID FOR EACH INDEX i



THE TWO LINES ARE THE ONLY ALLOWED CONFIGURATION

THE SET IS NOT DIFFERENTIABLE

KINKY POINT

Def (strict complementarity)

HOW TO READ THE KKT CONDITIONS

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

SMALLEST, GIVEN
CONSTRAINT
SMALLEST UNCONSTRAINED
Objective function $f(w)$

$-\nabla f(w)$
 $\nabla g(w)^*$
 $\nabla f(w)$

THE GRADIENTS OF f ARE
ORTHOGONAL TO THE
LEVELSETS \rightarrow POINTING
UPWARDS

Equality constraints
 $g(w) = 0$

AT w^* WE HAVE THAT $-\nabla f(w^*) = \lambda^* \nabla g(w^*)$ \rightarrow THE MULTIPLIER MAKES
THE FORCES COMPARABLE
(THE SIGN RETAINS)
→ the two gradients are equal, up to some factor (multiplier)
→ THE TWO VECTORS BALANCE, AT w^*

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

$g(w) = 0$
 $f(w)$

$-\nabla f(w)$
 $\nabla g(w)$

GLOBAL / LOCAL MINIMUM

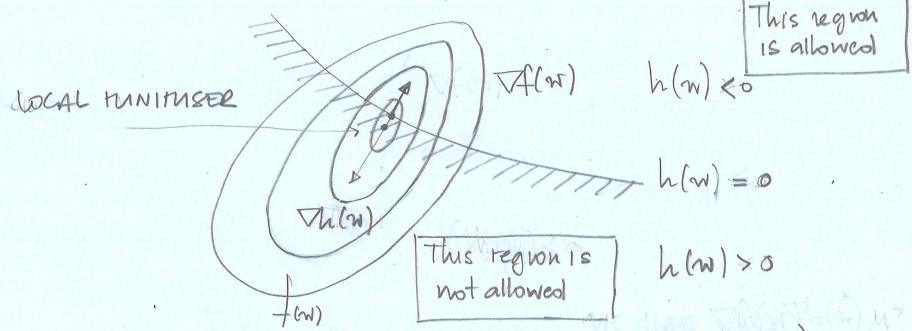
$g(w) = 0$ Equality
constraint at w^*

THERE IS NO NEED TO BALANCE THE GRADIENTS

$$-\nabla f(w^*) = \lambda^* \nabla g(w^*)$$

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

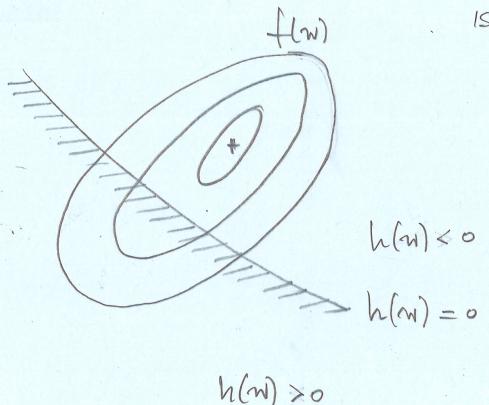
4/ CONSIDER THE OBJECTIVE FUNCTION $f(w)$, VIA ITS LEVEL SETS



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



Nod