Introduction to Optimisation:

Unconstrained Nonlinear Programming

Lecture 8

Lecture notes by Dr. Julia Memar and Dr. Hanyu Gu and with an acknowledgement to Dr.FJ Hwang and Dr.Van Ha Do

Introduction

➢ Let $f(x)$ is be nonlinear function of vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ defined over the domain $D \subseteq R^n$. Consider an NLP problem

$$\min_{\boldsymbol{x} \in D} f(x) \text{ (or } \max_{\boldsymbol{x} \in D} f(x))$$

➢ If $D = R^n$, then we have an unconstrained non-linear problem (NLP)

$$\min_{\boldsymbol{x} \in D} f(x) \text{ (or } \max_{\boldsymbol{x} \in D} f(x))$$

where no constraints are placed on the decision variables x.

Introduction – some definitions

- *Global minimum:*

A point $\boldsymbol{x}^*$ is a global ***minimiser*** or a global minimum point of a function $f(\boldsymbol{x})$ if

- The value $f(\boldsymbol{x}^*)$ is a global minimum value of $f(\boldsymbol{x})$.
- A ***strict*** global minimiser or a ***strict*** global minimum point is defined as

x^* is global minimiser if
$f(x^*) \leq f(x)$ For x in Domain

x^* is strict global minimiser if
$f(x^*) < f(x)$ for any x in Domain

Introduction – some definitions

- *Local minimum:*

A point x^* is a local ***minimiser*** or a local minimum point of a function $f(x)$ if

- The value $f(x^*)$ is a local minimum value of $f(x)$.
- A ***strict*** local minimiser or a ***strict*** local minimum point is defined as

x^* is a local mimimiser, if there exists $S \subseteq D$:

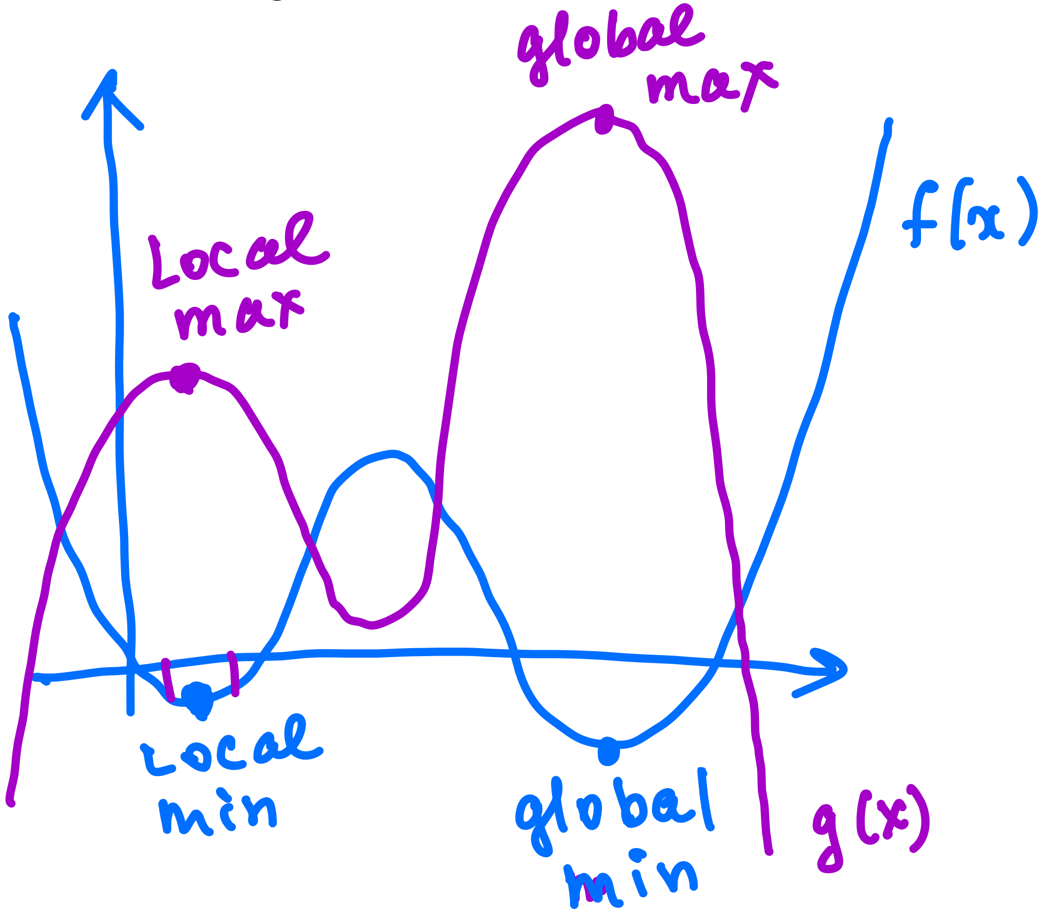
$$f(x^*) \le f(x) \text{ for any } x \in S$$

strict local mimimiser if there exists $S \in D$

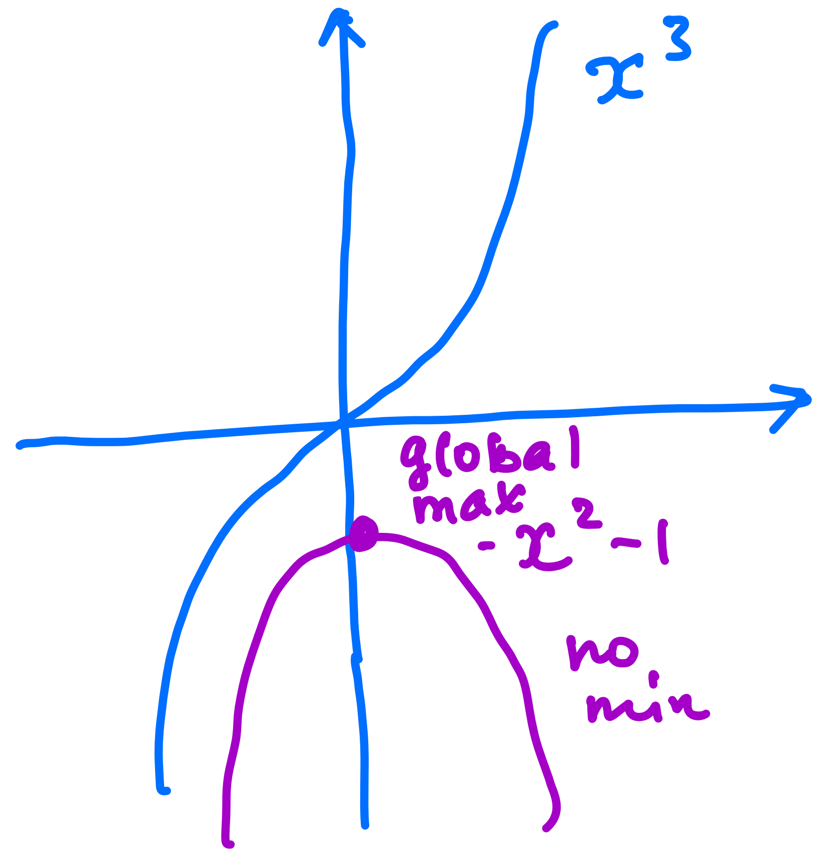
$$f(x^*) < f(x) \text{ in } x \in S$$

Introduction

- It is possible for a function to have
  - both global and local minimisers:

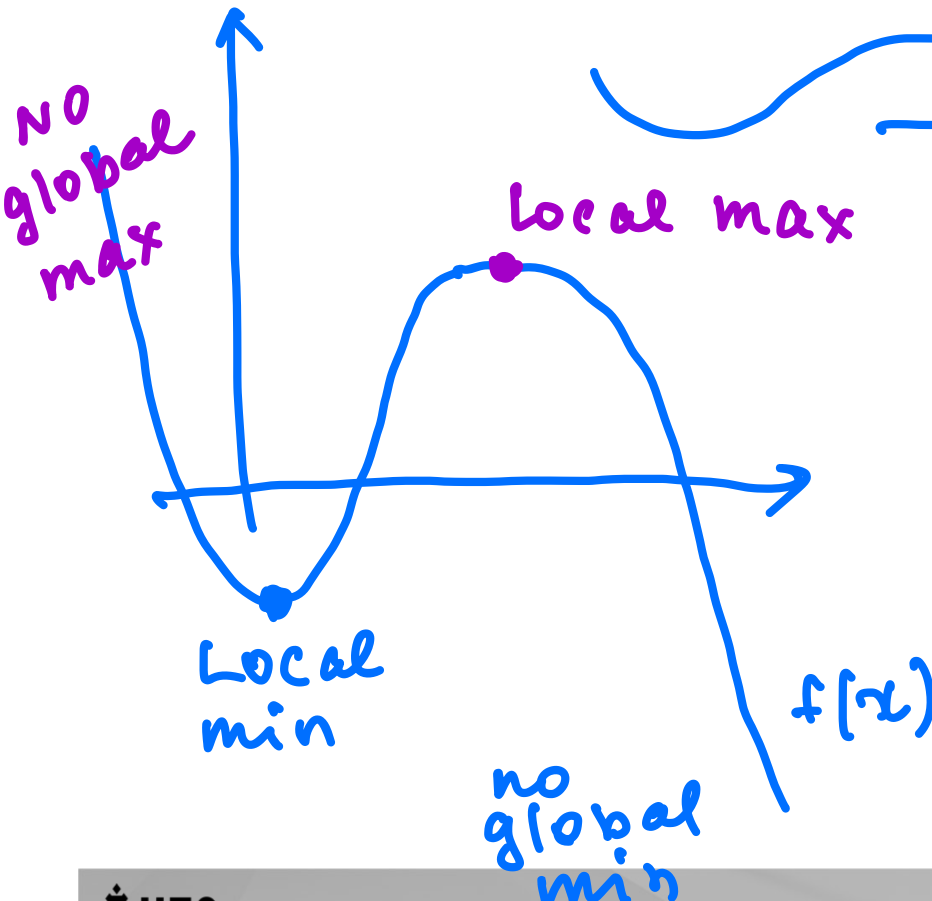


  - neither global nor local minimisers:

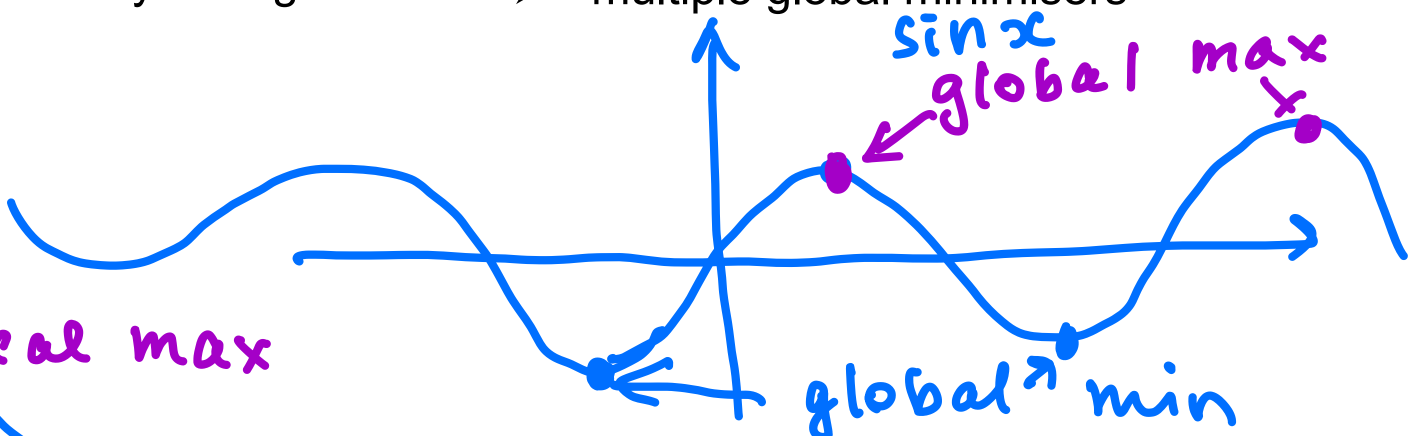


Introduction

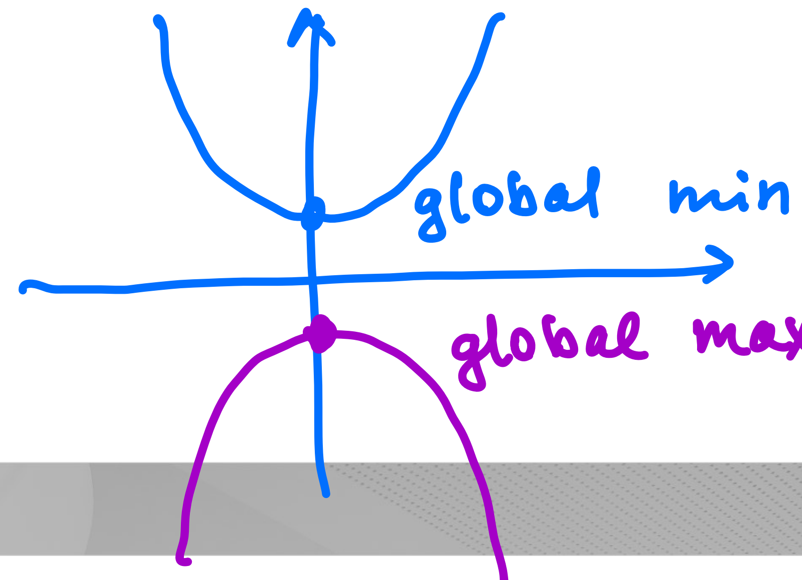
- It is possible for a function to have
  - a local minimiser and yet no global minimiser;



  - multiple global minimisers



  - unique global minimizer



Introduction

is R^n convex?

- In this course we will consider only a specific type of NLP problems – minimising a *convex* function (or maximising a *concave* function) over a *convex set*.

Definitions:

Assume that *f(x)* has continuous second-order partial derivatives. For each point $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ denote:

- *Gradient of f(x) :*

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

- Gradient of $f(x)$:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Definitions:

Assume that *f(x)* has continuous second-orde … es. For each point $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ denote:

- *Hessian matrix:*

$$H(x) = \nabla^2 f = \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array} \\ \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \end{array}$$

- $\nabla^2 f(x)$ is $n \times n$ symmetrical matrix:

Due Clairout's theorem

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Definitions:

- $\boldsymbol{i^{th}}$ ***principal minor(s) of*** $\boldsymbol{n \times n}$ ***matrix*** *is the determinant of any* $i \times i$ *matrix obtained by deleting* $(n - i)$ *row(s) and the corresponding* $(n - i)$ *column(s) of the matrix.*

Example: $A = \begin{pmatrix} -2 & -1 \\ -1 & 4 \end{pmatrix}$

$n = 2$

1st principal minor(s) : $i = 1;\ n - i = 1$

always on the main diagonal

$4, -2$

2nd principal minor(s) : $i = 2,\ n - i = 0$

$$\det\begin{pmatrix} -2 & -1 \\ -1 & 4 \end{pmatrix} = -9$$

Definitions:

$n = 3; \quad i = 1, 2, 3$

Example 1: $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 0 & 0 \end{pmatrix}$ *1st principal minor(s) :* $i = 1, \quad n - i = 2$

$1, 5, 0$

2nd principal minor(s) : $i = 2, \; n - i = 1$

$\det\begin{pmatrix} 5 & 6 \\ 0 & 0 \end{pmatrix} = 0 \qquad \det\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} = -3$

$\det\begin{pmatrix} 1 & 3 \\ 1 & 0 \end{pmatrix} = -3$

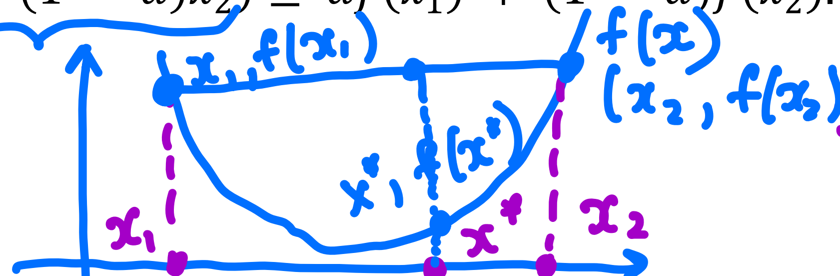
3rd principal minor(s) : $i = 3, \quad n - i = 0$

$$\det\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 0 & 0 \end{pmatrix} = 1 \times \det\begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} - 0 \times \det\begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix} + \cancel{0} \times \det\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} = -3$$

Convex function

- A function $f(x)$ is ***convex*** if for any two points (or vectors) $x_1 \in$ D and $x_2 \in$ D and for any $\alpha \in [0,1]$

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2).$$

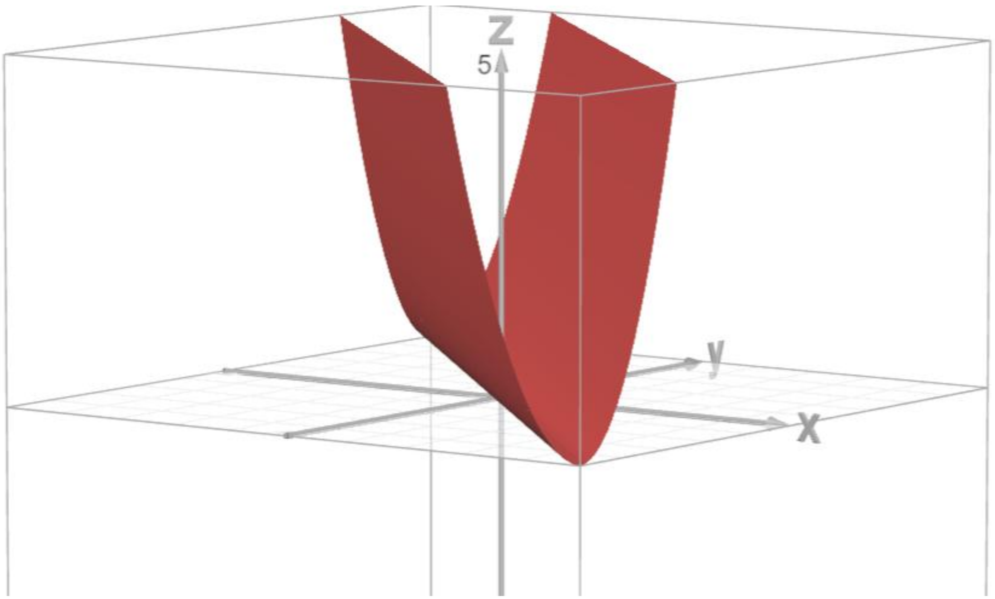


$x^* = \alpha x_1 + (1-\alpha)x_2$

- ***Theorem 1.*** Assume that $f(x)$ has continuous second-order partial derivatives for each point $x = (x_1, x_2, \ldots, x_n)$. Then $f(x)$ is convex function on D if and only if for each $x \in$ D all principal minors of its Hessian are nonnegative.

Convex function

➢ *Example 2:* $f(x) = x_1^2 + 2x_1x_2 + x_2^2$ – is it convex?



$\nabla f(x)^T = \langle 2x_1 + 2x_2,\ 2x_1 + 2x_2 \rangle$

$\nabla^2 f(x) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \rightarrow$ same for all x

1 p.m: $2, 2 > 0$

2^{nd} p.m: $\det \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = 0$

by Th 1 all p.m are ≥ 0,

Hence $f(x)$ is convex

Concave function

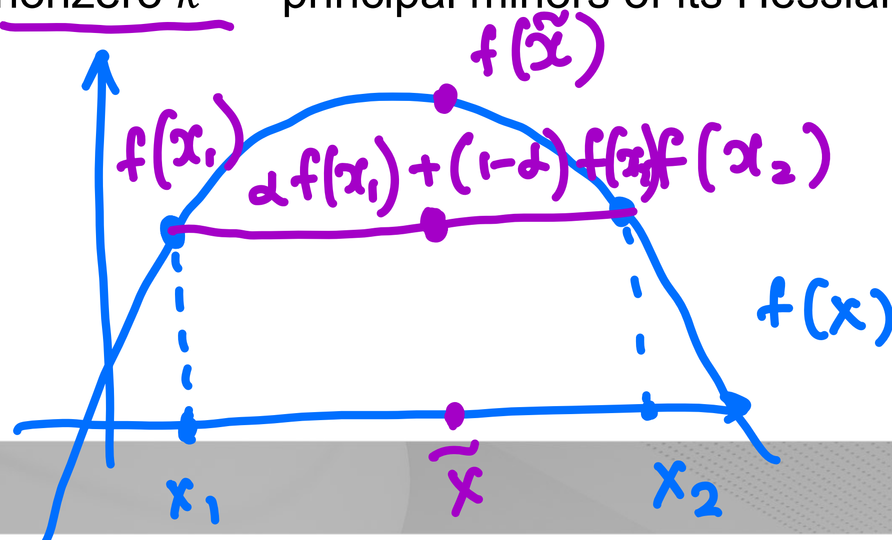
- A function $f(x)$ is ***concave*** if for any two points (or vectors) $x_1 \in$ D and $x_2 \in$ D and for any $\alpha \in [0,1]$

$\tilde{x} = \alpha x_1 + (1-\alpha) x_2$

$$f(\alpha x_1 + (1-\alpha)x_2) \geq \alpha f(x_1) + (1-\alpha) f(x_2).$$

$$f(\tilde{x}) \geq \alpha f(x_1) + (1-\alpha) f(x_2)$$

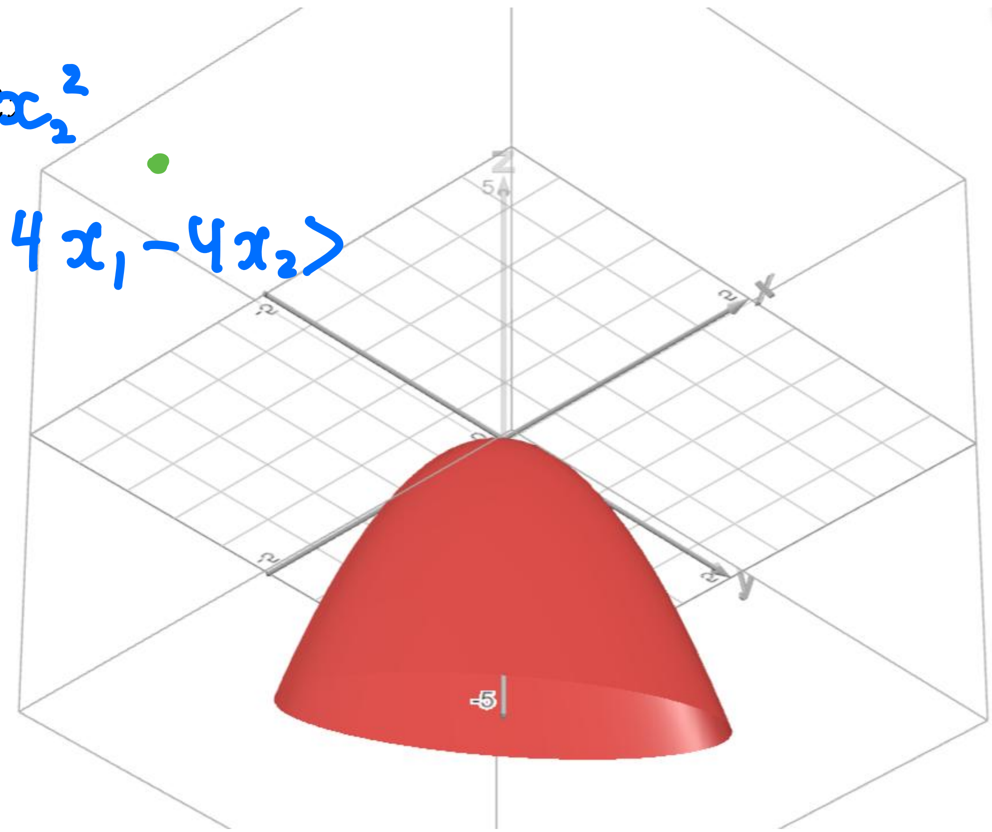
- ***Theorem 2.*** Assume that that $f(x)$ has continuous second-order partial derivatives for each point $x = (x_1, x_2, \dots, x_n)$. Then $f(x)$ is concave function on D if and only if for each $x \in$ D and $k = 1 \dots n$ all nonzero k^{th} principal minors of its Hessian matrix have the same sign as $(-1)^k$.



- $\det(-A) = (-1)^n \det A$
- if $f(x)$ is convex
  $-f(x)$ is concave

Concave function

- *Example 3:* $f(x) = -3x_1^2 + 4x_1x_2 - 2x_2^2$



$\nabla f(x)^T = \langle -6x_1 + 4x_2, \; 4x_1 - 4x_2 \rangle$

$$\nabla^2 f(x) = \begin{pmatrix} -6 & 4 \\ 4 & -4 \end{pmatrix}$$

1^{st} p.m. : $-6, -4 < 0$ as $(-1)^1$

2 p.m : $\det \begin{pmatrix} -6 & 4 \\ 4 & -4 \end{pmatrix} = 8 > 0$ as $(-1)^2$

Hence by Th 2 $f(x)$ is concave

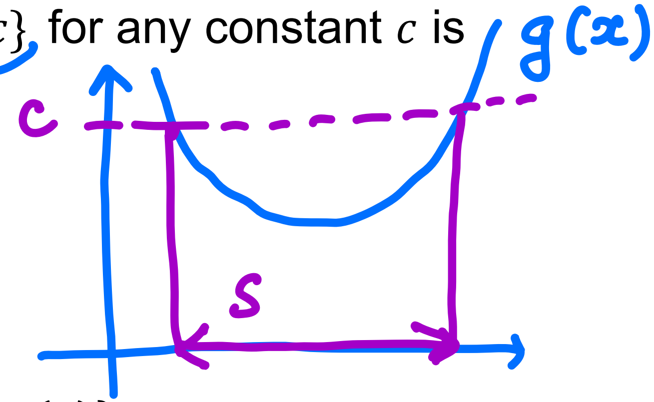
UTS

Convex set – some results

- A set S is ***convex*** if for any two points (or vectors) $x_1 \in S$ and $x_2 \in S$ and for any $\alpha \in [0,1]$ $\alpha x_1 + (1-\alpha)x_2 \in S$.

- If $g(x)$ is a convex function, then the set $S = \{x : g(x) \leq c\}$ for any constant c is convex

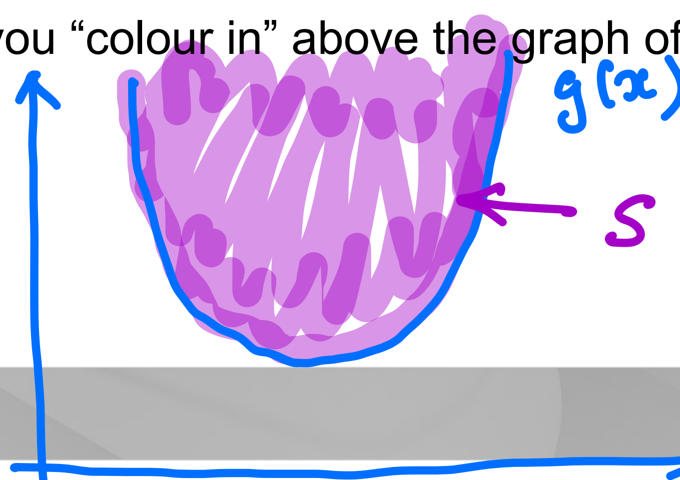
prove it as a challenge



- If $g(x)$ is a convex function, then the set

$$S = \{u = (x|y) = (x_1, x_2, \ldots, x_n, y) : y \geq g(x)\}$$

is a convex set of R^{n+1}. If you “colour in” above the graph of a convex function, then you get a convex set.



UTS

Convex set – some results

- ***Theorem 3***. If *f(x)* is a convex function and S is a convex set, then any local minimum of the minimisation NLP

$$\min_{x \in S} f(x)$$

is also a global minimum. If f(x) is a strictly convex function, then the global minimum will be unique.

f(x) is strictly convex, if for

any $x_1 \in S$ $x_2 \in S$ and

$\tilde{x} = \alpha x_1 + (1-\alpha) x_2$, $\alpha \in (0,1)$

$f(\tilde{x}) < \alpha f(x_1) + (1-\alpha) f(x_2)$

More definitions

for any $x \neq 0$

Let A be $n \times n$ symmetric matrix. Then A is

- Positive-definite if

  $x^T A x > 0$

- Positive- semidefinite if

  $x^T A x \geq 0$

- Negative- definite if

  $x^T A x < 0$

- Negative- semidefinite if

  $x^T A x \leq 0$

- Indefinite if for some $\tilde{x}$ $\tilde{x}^T A \tilde{x} > 0$
  some $\bar{\bar{x}}$ $\bar{\bar{x}}^T A \bar{\bar{x}} < 0$

More definitions

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

- Determine the type of the matrix

$A = I.$ if $x \neq 0$

$$\underbrace{x^T I x}_{x} = x^T x = x_1^2 + x_2^2 + \dots x_n^2 =$$

$$= |x|^2 > 0$$

$\downarrow$ I is pos.-def.

- $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$

Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$(x_1 \ x_2) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1 - x_2, -x_1 + x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$$

$$= x_1^2 - x_1 x_2 - x_1 x_2 + x_2^2 =$$

$$= (x_1 - x_2)^2 \geq 0 \rightarrow B \text{ is pos-semidef.}$$

$Av = \lambda v$

More definitions

to find eigenvalues

Theorem 4. A symmetric matrix *A* is positive-definite if and only if all its *eigenvalues* are positive.

solve $\det(A - \lambda I) = 0.$

- Note: we can also calculate the upper left determinants

Example: $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ → to find all eigenvalues

solve $\det \begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix} = 0.$

$(2-\lambda)^2 - 1 = 0.$

$(2-\lambda)^2 = 1$

$2-\lambda = 1$ → $\lambda = 1$

$2-\lambda = -1$ → $\lambda = 3$

by Th 4 all eigenvalues are pos-ve → $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ is positive-def.

Optimality conditions

One-dimensional case - assume that that $f(x)$ has continuous second-order partial derivatives for each x.

- Taylor series expansion of $f(x)$ centred at a : $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$; linear approximation would be

$$f(x^* + \delta) \simeq f(x^*) + \underbrace{f'(x^*)}_{=0}\,\delta$$

$$\delta = x - x^*$$

- First-order *necessary* optimality condition:

If x^* is a local minimum of $f(x)$ then $f'(x^*) = 0$. $\rightarrow$ otherwise

if $f'(x^*) > 0 \rightarrow$ choose $\delta < 0$

$$\underline{f(x^* + \delta) < f(x^*)}$$

if $f'(x^*) < 0 \rightarrow \delta > 0$

$$\underline{f(x^* + \delta) < f(x^*)}$$

Optimality conditions

One-dimensional case - assume that that $f(x)$ has continuous second-order partial derivatives for each x.

➢ Second-order ***sufficient*** optimality condition:

If $f'(x^*) = 0\ and\ \ f''(x^*) > 0$, then x^* is a local minimum.

$$\underbrace{f(x^*+\delta)} \approx \underline{\underline{f(x^*)}} + \underbrace{f'(x^*)}_{=0}\delta + \underbrace{\frac{1}{2} f''(x^*)\delta^2}_{\text{must be positive}}$$

$$\Downarrow$$

$$f(x^*) < f(x^*+\delta)$$

$$\downarrow$$

$$f''(x^*) > 0$$

Optimality conditions

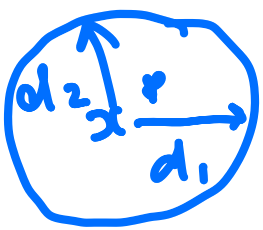
Multidimensional case - assume that that $f(x)$ has continuous second-order partial derivatives for each point $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ *in* R^n.

- By Taylor's Theorem for a small deviation $\boldsymbol{d} = (d_1, d_2, \dots, d_n)^T$:

$$f(x^* + d) \simeq f(x^*) + \underbrace{\nabla f(x^*)^T d}_{= \sum_{i=1}^{n} \frac{\partial f(x^*)}{\partial x_i} * d_i} + \frac{1}{2} d^T \nabla^2 f(x^*)\, d + o(\|d\|^2)$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(x^*)}{\partial x_i \partial x_j} d_i d_j$$

$O(\|d^2\|)$



x^* is some point

Optimality conditions

Multidimensional case - assume that that $f(x)$ has continuous second-order partial derivatives for each point $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ *in* R^n.

- First-order *necessary* optimality condition:

If x^* is a local minimum of $f(x)$ then $\nabla f(x^*) = 0$

$$f(x^* + d) \simeq f(x^*) + \underbrace{\nabla f(x^*)\, d}$$

if $\nabla f(x^*) \neq 0$,
then we always
can choose d:

$\nabla f(x^*)\, d < 0 \rightarrow$
x^* is not local min

Optimality conditions

Multidimensional case - assume that that $f(x)$ has continuous second-order partial derivatives for each point $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ *in* R^n.

➢ Second-order *sufficient* optimality condition:

If $\nabla f(x^*) = 0$ *and* $\nabla^2 f(x^*)$ is positive − definate, then x^* is a local minimum.

$$f(x^* + d) \simeq f(x^*) + \underbrace{\nabla f^T(x^*)}_{= 0} d + \underbrace{\frac{1}{2} d^T \nabla^2 f(x^*) d}_{\text{must be positive for any } d}$$

must be positive for any d

$\Downarrow$

$\nabla^2 f(x^*)$ must be pos.-def.

Summary: necessary and sufficient conditions for min problem

Theorem 5-min. (Necessary conditions) If x^* is a local minimum for an unconstrained NLP problem $\min f(x)$, then

- $\nabla f(x^*) = 0$, and
- $\nabla^2 f(x^*)$ is positive semidefinite.

Theorem 6-min. (Sufficient conditions)

- If $\nabla f(x^*) = 0$, and
- $\nabla^2 f(x^*)$ is positive-definite,

then x^* is a local minimum for the unconstrained NLP problem $\min f(x)$.

Summary: necessary and sufficient conditions for max problem

Theorem 5-max. (Necessary conditions) If x^* is a local maximum for an unconstrained NLP problem $\max f(x)$, then

- $\nabla f(x^*) = 0$, and
- $\nabla^2 f(x^*)$ is negative-semidefinite.

Theorem 6-max. (Sufficient conditions)

- If $\nabla f(x^*) = 0$, and
- $\nabla^2 f(x^*)$ is negative-definite,

then x^* is a local maximum for the unconstrained NLP problem $\max f(x)$.

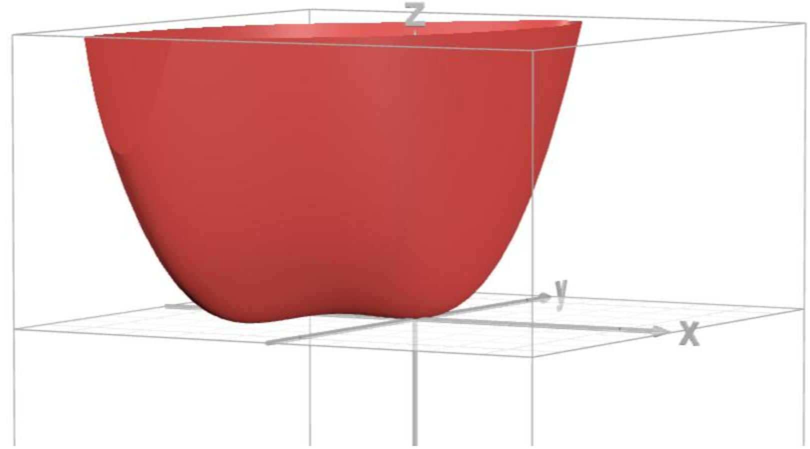
10:21

10:30

Example

- For $f(\boldsymbol{x}) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$.

a) Determine minimizers and maximizers



b) Indicate what kind of max/min are these points (local, global, strict etc)

1. Need both $\nabla f(x)$ and $\nabla^2 f(x)$

$$\nabla f(x) = \begin{pmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{array}{c} \begin{array}{cc} x_1 & \qquad\qquad x_2 \end{array} \\ \begin{pmatrix} 4 + 12x_1 + 12x_1^2 & -2 \\ -2 & 2 \end{pmatrix} \end{array}$$

2. Find x: $\nabla f(x) = 0$.

$$\downarrow$$

$$\begin{cases} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 = 0 \\ 2x_2 - 2x_1 = 0 \rightarrow x_1 = x_2 \end{cases}$$

$$\downarrow$$

$$4x_1 - 2x_1 + 6x_1^2 + 4x_1^3 = 0 \quad \div 2$$

$$x_1 + 3x_1^2 + 2x_1^3 = 0$$

$$x_1(2x_1^2 + 3x_1 + 1) = 0$$

$$x_1 = 0 \qquad x_1 = \frac{-3 \pm \sqrt{9-8}}{4} = \begin{cases} -1 \\ -\frac{1}{2} \end{cases}$$

$\nabla f(x) = 0$ when

$$x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$x = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$x = \begin{pmatrix} -1/2 \\ -1/2 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{array}{c} \begin{array}{cc} x_1 & \qquad\qquad x_2 \end{array} \\ \begin{pmatrix} 4 + 12x_1 + 12x_1^2 & -2 \\ -2 & 2 \end{pmatrix} \end{array}$$

$$\nabla^2 f(0,0) = \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix} \rightarrow \text{use Th. 4:}$$

solve

$$\det \begin{pmatrix} 4-\lambda & -2 \\ -2 & 2-\lambda \end{pmatrix} = 0.$$

$$(4-\lambda)(2-\lambda) - 4 = 0$$

$$\lambda^2 - 6\lambda + 4 = 0$$

$$\lambda = \frac{6 \pm \sqrt{36-16}}{2} = \frac{6 \pm \overset{=2\sqrt{5}}{\sqrt{20}}}{2} = 3 \pm \sqrt{5} > 0.$$

$\downarrow$

all eigenvalues of $\nabla^2 f(0,0)$ are pos-ve

$\downarrow$

by Th 4 $\nabla^2 f(0,0)$ pos-def $\rightarrow$ $x=(0,0)$ is local minimiser

$$\nabla^2 f(-1,-1) = \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix} \rightarrow$$ pos.-def $\rightarrow$ $x=(-1,-1)$ is local minimiser by Th. 6

$$\nabla^2 f(-\tfrac{1}{2}, -\tfrac{1}{2}) = \begin{pmatrix} 1 & -2 \\ -2 & 2 \end{pmatrix}$$

Try use Th. 4: $\det\begin{pmatrix} 1-\lambda & -2 \\ -2 & 2-\lambda \end{pmatrix} = 0$

$$(1-\lambda)(2-\lambda) - 4 = 0$$

$$\lambda^2 - 3\lambda - 2 = 0.$$

$$\lambda_1 = \frac{3 + \sqrt{9+8}}{2} > 0$$

$$\lambda_2 = \frac{3 - \sqrt{9+8}}{2} < 0$$

$\rightarrow$ can not use Th. 4.

use definition : $x \neq 0$

$$(x_1 \quad x_2)\begin{pmatrix}1 & -2\\ -2 & 2\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} =$$

$$= (x_1 - 2x_2 \;,\; -2x_1 + 2x_2)\begin{pmatrix}x_1\\ x_2\end{pmatrix} = x_1^2 - 2x_1x_2 - 2x_1x_2 + 2x_2^2 =$$

$$= x_1^2 - 4x_1x_2 + 4x_2^2 - 2x_2^2 = (x_1 - 2x_2)^2 - (\sqrt{2}\, x_2)^2 = (*)$$

$x_1 = 2x_2$; $x_2 \neq 0 \rightarrow (*) < 0$

$x_1 = 4x_2$; $x_2 \neq 0 \rightarrow (*) > 0$

$\rightarrow \nabla^2 f\begin{pmatrix}-1/2\\ -1/2\end{pmatrix}$ -

Indefinite

so, $(0,0)$ and $(-1,-1)$ are local min ?

$$\left.\begin{array}{l} f(0,0) = 0 \\ f(-1,-1) = 0 \end{array}\right\}$$ not unique local min

r $f(\boldsymbol{x}) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$ =

$$= x_1^2 - 2x_1x_2 + x_2^2 + x_1^2(1 + 2x_1 + x_1^2) =$$

$$= (x_1 - x_2)^2 + x_1^2(1 + x_1)^2 \geq 0 \text{ for any } x$$

$\downarrow$

$(0,0)$ and $(-1,-1)$ are global minimisers.

Necessary and sufficient conditions

Theorem 7. Consider a function $f(x)$ defined in a convex domain. Then

- *Necessary condition for convexity: if* $f(x)$ is convex , then $\nabla^2 f(x)$ is positive-semidefinite everywhere in its domain.

- *Sufficient condition for strict convexity:* Function $f(x)$ is strictly convex if its Hessian matrix $\nabla^2 f(x)$ is positive- definite for all x in its domain.

- *Example:* $f(\boldsymbol{x}) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$

$: f(\boldsymbol{x}) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$ → Find min

$$\nabla f(x) = \begin{pmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - 3 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$ → Test using Th 4:

$$\det(\nabla^2 f - \lambda I) = 0 \to (2-\lambda)^2 = 1$$

see above

$\lambda_1 = 1 \quad \lambda_2 = 3$

↓

$\nabla^2 f(x)$ is positive-def for all x

↓

- $f(x)$ strictly convex (by Th. 7).

Find min $f(x)$: Solve $\nabla f(x) = \emptyset$

$$\begin{cases} 2x_1 - x_2 = 0 & ① \\ -x_1 + 2x_2 - 3 = 0 & ② \times 2 \end{cases}$$

$+$

$$\cancel{2x_1} - x_2 - \cancel{2x_1} + 4x_2 - 6 = 0$$

$3x_2 = 6$; $x_2 = 2$; $x_1 = 1$

↓

- by Th 6 $\nabla f(1,2) = 0$ and $\nabla^2 f(1,2)$ - pos. def → $(1,2)$ is local min

• by Th 3 (1,2) is unique global min

Theorem 7 - example

➢ *Example:* $f(\boldsymbol{x}) = x_1^2 - x_1x_2 + x_2^2 - 3x_2$

Unconstrained non-linear optimisation – some results

- ***Theorem 1.*** Assume that $f(x)$ has continuous second-order partial derivatives for each point $x = (x_1, x_2, \dots, x_n)$. Then $f(x)$ is convex function on D if and only if for each $x \in$ D all principal minors of its Hessian are nonnegative.

- ***Theorem 2.*** Assume that that $f(x)$ has continuous second-order partial derivatives for each point $x = (x_1, x_2, \dots, x_n)$. Then $f(x)$ is concave function on D if and only if for each $x \in$ D and $k = 1 \dots n$ all nonzero k^{th} principal minors of its Hessian matrix have the same sign as $(-1)^k$.

- ***Theorem 3***. If *f(x)* is a convex function and S is a convex set, then any local minimum of the minimisation NLP

$$\min f(x)$$
$$s.t.\ \ x \in S$$

is also a global minimum. If f(x) is a strictly convex function, then the global minimum will be unique.

- ***Theorem 4.*** A symmetric matrix A is positive definite if and only if all its *eigenvalues* are positive. Note: we can also calculate the upper left determinants

- ***Theorem 5***. (Second-order necessary condition) If x^* is a local minimum for an unconstrained NLP problem $min\, f(x)$, then

∇f(x^*) = 0, and ∇^2f(x^*) is positive semidefinite.

- ***Theorem 6***. (Second-order sufficient condition)

If ∇f(x^*) = 0, and ∇^2f(x^*) is positive definite,

then x^* is a local minimum for the unconstrained NLP problem $\min f(x)$.

- ***Theorem 7.*** Consider a function $f(x)$ defined in a convex domain. Then

  Necessary condition for convexity: if $f(x)$ is convex , then $\nabla^2 f(x)$ is positive semidefinite everywhere in its domain.

  Sufficient condition for strict convexity: Function $f(x)$ is strictly convex if its Hessian matrix $\nabla^2 f(x)$ is positive definite for all x in its domain.

11:20!

Gradient methods - motivation

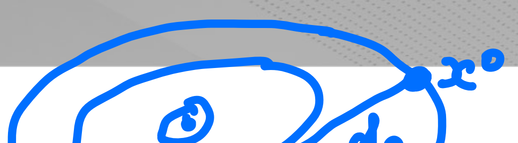
Finding stationary points is not always possible or easy

Gradient methods

is

a group of iterative procedures that approximate stationary points

by applying the optimality conditions .



Gradient methods

➢Consider the nonlinear NLP: $\min f(x)$ and let x_0 be the first approximation; $\boldsymbol{d_0}$ - initial direction; α – scalar. By Taylor's Theorem:

$$f(x_0+\alpha d) \approx f(x_0) + \nabla f(x_0)^T \alpha d + \frac{1}{2} d^T \nabla^2 f(x^0)\, d$$

if $\nabla^2 f(x^0) \neq 0 \rightarrow$ IGNORE

➢Basic idea (for min problem):

1. Start an iteration *k* with x_k and chose direction $\boldsymbol{d_k}$, so $f(x_k) > f(x_k + \alpha d_k)$
2. Find $\alpha_k > 0$ such that $f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k)$
3. Let $x_{k+1} = x_k + \alpha_k d_k$

➢ In what follows, $\boldsymbol{d_k}$ is chosen as $d_k = -D_k \nabla f(x_k)$

where D_k is pos-def:

$$f(x_k + \alpha_k d_k) \approx f(x_k) - \underbrace{\nabla f(x_k)^T D_k \nabla f(x_k)}_{> 0}\, \alpha$$

$$f(x_k + \alpha_k d_k) < f(x_k).$$

Steepest descend method

Step 0. Choose a starting point x_0, and a small positive scalar ε. Set $k = 0$.

It. k

Step 1. If $\|\nabla f(x_k)\| < \varepsilon$, then STOP: x_k is a satisfactory approximate minimum of $f(x)$.
Otherwise, set $\boldsymbol{d_k} = -\nabla f(x_k)$ $\rightarrow D_k = I$

Step 2. Choose the step size α_k by solving the one-dimensional problem

$$\min_{\alpha>0} g(\alpha) = \min_{\alpha>0} f(x_k + \alpha \boldsymbol{d_k}).$$

Set $x_{k+1} = x_k + \alpha_k \boldsymbol{d_k}$. Set $k = k + 1$ and go to Step 1.

Example

- Find min $f(\boldsymbol{x}) = x_1^2 - x_1x_2 + x_2^2 - 3x_2$

Let $x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$; $\varepsilon = 1/2$

$: f(\boldsymbol{x}) = x_1^2 - x_1x_2 + x_2^2 - 3x_2$

$$\nabla f(x) = \begin{pmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - 3 \end{pmatrix}$$

<u>It. 0</u> $x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

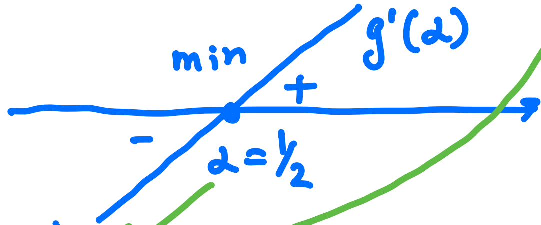
1. $\nabla f(0,0) = \begin{pmatrix} 0 \\ -3 \end{pmatrix} \rightarrow \|\nabla f(0,0)\| = 3 > \varepsilon$

2. $d_0 = -\nabla f(0,0) = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \rightarrow \underbrace{x^{(0)} + \alpha d_0}_{} = \begin{pmatrix} 0 \\ 3\alpha \end{pmatrix}$

3. $g(\alpha) = f(0, 3\alpha) = 9\alpha^2 - 9\alpha$

min $g(\alpha)$ at stat. point : $g'(\alpha) = 0$.

$$g'(\alpha) = 18\alpha - 9 = 0 \to \alpha = \tfrac{1}{2}$$



4. $x^{(1)} = \begin{pmatrix} 0 \\ 3/2 \end{pmatrix}$

<u>It. 1</u> $\quad x^{(1)} = \begin{pmatrix} 0 \\ 3/2 \end{pmatrix}$; $\nabla f(0, 3/2)^T = \langle -3/2, 0 \rangle$

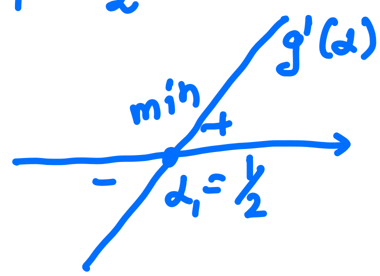
1. $\|\nabla f(0, 3/2)\| = 3/2 > \varepsilon$

2. $d_1 = -\nabla f(0, 3/2) = \begin{pmatrix} 3/2 \\ 0 \end{pmatrix}$

$$x^{(1)} + \alpha d_1 = \begin{pmatrix} 0 \\ 3/2 \end{pmatrix} + \begin{pmatrix} 3/2\,\alpha \\ 0 \end{pmatrix} = \begin{pmatrix} 3/2\,\alpha \\ 3/2 \end{pmatrix}$$

3. $g(\alpha) = f(3/2\,\alpha, 3/2) = \frac{9}{4}\alpha^2 - \frac{9}{4}\alpha + \frac{9}{4} - \frac{9}{2}$

$$g'(\alpha) = \frac{18}{4}\alpha - \frac{9}{4} = 0 \to \alpha = \tfrac{1}{2}$$



4. $x^{(2)} = x^{(1)} + \alpha_1 d_1 = \begin{pmatrix} 3/4 \\ 3/2 \end{pmatrix}$

It. 2

$$x^{(2)} = \begin{pmatrix} 3/4 \\ 3/2 \end{pmatrix} \qquad \nabla f\left(3/4,\ 3/2\right)^T = \left\langle 3/2 - 3/2,\ -3/4 + 3 - 3 \right\rangle$$
$$= \left\langle 0,\ -3/4 \right\rangle$$

1. $\left\| \nabla f\left(3/4, 3/2\right) \right\| = 3/4 > \varepsilon$

2. $d_2 = \begin{pmatrix} 0 \\ 3/4 \end{pmatrix} = -\nabla f\left(3/4,\ 3/2\right)$

$$x^{(2)} + \alpha d_2 = \begin{pmatrix} 3/4 \\ 3/2 \end{pmatrix} + \begin{pmatrix} 0 \\ 3/4\,\alpha \end{pmatrix} = \begin{pmatrix} 3/4 \\ 3/2 + 3/4\,\alpha \end{pmatrix}.$$

3. $g(\alpha) = f\left(3/4,\ 3/2 + 3/4\,\alpha\right) =$

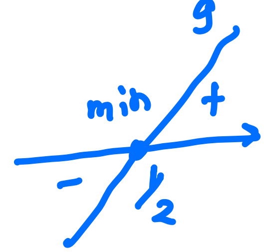
$f(\boldsymbol{x}) = x_1^2 - x_1x_2 + x_2^2 - 3x_2$

$$= \frac{9}{16} - \frac{3}{4}\left(3/2 + \frac{3}{4}\alpha\right) + \left(3/2 + 3/4\,\alpha\right)^2 - 3\left(3/2 + 3/4\,\alpha\right)$$

$$g'(\alpha) = -\frac{9}{16} + 2\left(3/2 + 3/4\,\alpha\right) \times 3/4 - \frac{9}{4} = 0.$$

$$-\frac{9}{16} + \cancel{\frac{9}{4}} + \frac{9}{8}\alpha - \cancel{\frac{9}{4}} = 0.$$

$$\frac{9}{8}\alpha - \frac{9}{16} = 0 \qquad \alpha = \frac{1}{2}$$



4. $x^{(3)} = \begin{pmatrix} 3/4 \\ 3/2 + \frac{3}{8} \end{pmatrix} = \begin{pmatrix} 3/4 \\ 15/8 \end{pmatrix}$

$\underline{\text{It}^{(3)}}$

$$\nabla f\begin{pmatrix} 3/4 \\ 15/8 \end{pmatrix} = \begin{pmatrix} 3/2 - \frac{15}{8} \\ -\frac{3}{4} + \frac{15}{4} - 3 \end{pmatrix} = \begin{pmatrix} -3/8 \\ 0 \end{pmatrix} \rightarrow$$

$\left\| \nabla f\begin{pmatrix} 3/4 \\ 15/8 \end{pmatrix} \right\| = 3/8 < 1/2 \rightarrow x = \begin{pmatrix} 3/4 \\ 15/8 \end{pmatrix}$ is sufficiently close appr-th of min

x_k - current approximation

Newton's method

1. Let $g(x) = f(x_k) + \underbrace{\nabla f(x_k)}_{\neq 0}^T (x - x_k) + \underbrace{\frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)}_{\text{pos-ve}}$

2. $min\ g(x) \Rightarrow \quad \nabla g(x) = 0$

3. $\nabla g(x) = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0$

4. Solve (3) for x:

$$\nabla^2 f(x_k)(x - x_k) = -\nabla f(x_k)$$

$$x - x_k = \left[\nabla^2 f(x_k)\right]^{-1} \times \left(-\nabla f(x_k)\right)$$

$$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \times \nabla f(x_k)$$

Newton's method

Step 0. Choose a starting point x_0, and a small positive scalar ε. Set $k = 0$.

Step 1. If $\|\nabla f(x_k)\| < \varepsilon$, then STOP: x_k is a satisfactory approximate minimum of $f(x)$. Otherwise, set

$$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \times \nabla f(x_k)$$

Step 2. Set $k = k + 1$ and go to Step 1 .

Example

- Find min : $f(\boldsymbol{x}) = x_1^2 - x_1x_2 + x_2^2 - 3x_2$

$$\nabla f(x) = \begin{pmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - 3 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad [\nabla^2 f]^{-1} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

0. $x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$; $\varepsilon = 1/2$

It. 0 $\nabla f(0,0)^T = \langle 0, -3 \rangle$; $\|\nabla f(0,0)\| = 3 > \varepsilon$.

1.

2.

$$x^{(1)} = x^{(0)} - \left[\nabla^2 f\right]^{-1} \nabla f(0,0) =$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ -3 \end{pmatrix} = -\frac{1}{3}\begin{pmatrix} -3 \\ -6 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

It. 1.

1. $\nabla f(1,2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \|\nabla f(1,2)\| = 0 < \varepsilon$

$\downarrow$

$x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is local min