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

Introduction – some definitions

- ***Local minimum:***

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

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

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

- 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 $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ denote:

- *Gradient of f(x) :*

Definitions:

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

- *Hessian matrix:*

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: $A = \begin{pmatrix} -2 & -1 \\ -1 & 4 \end{pmatrix}$ *1st principal minor(s) :*

2nd principal minor(s) :

Definitions:

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

2^{nd} principal minor(s) :

3^{rd} principal minor(s) :

Convex function

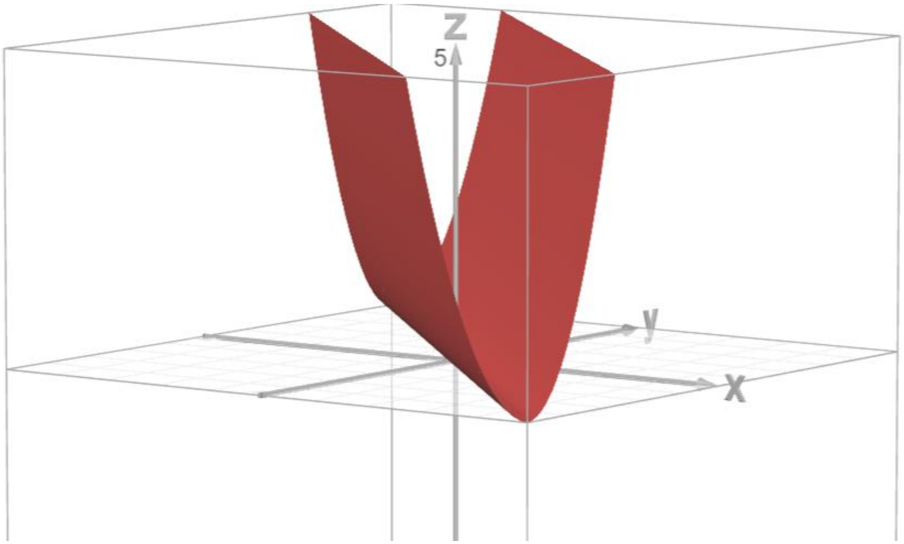
- A function $f(x)$ is ***convex*** if for any two points (or vectors) $x_1 \in$ D and $x_2 \in$ D and for any $\alpha \in [0,1]$

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

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



Concave function

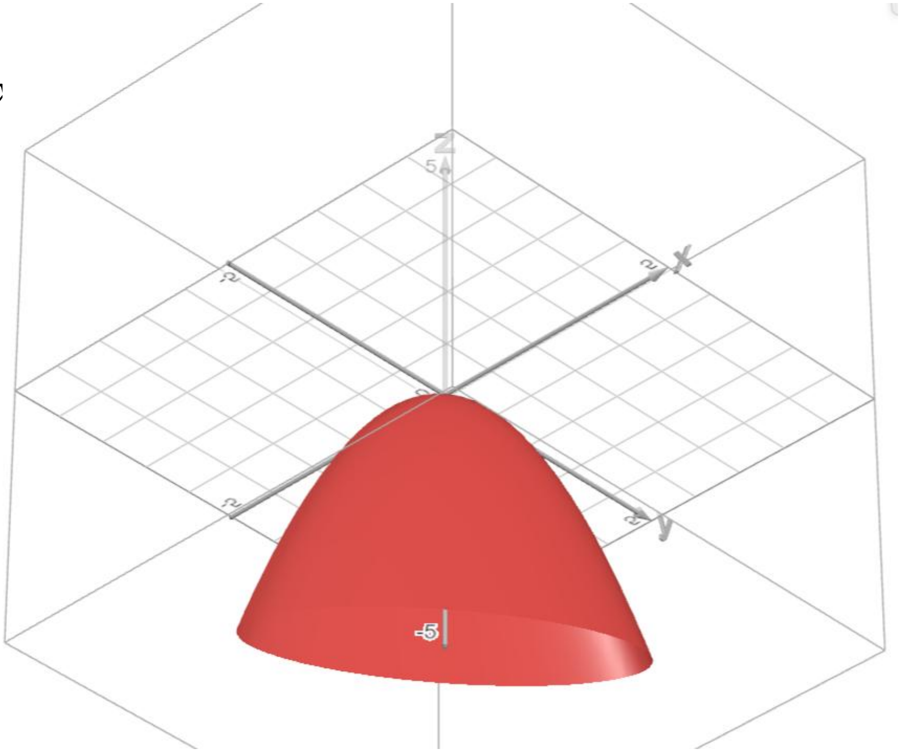
- A function $f(x)$ is ***concave*** if for any two points (or vectors) $x_1 \in$ D and $x_2 \in$ D and for any $\alpha \in [0,1]$

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

- ***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 $(-1)^k$.

Concave function

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



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

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

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

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.

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.

More definitions

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

- Positive-definite if
- Positive- semidefinite if
- Negative- definite if
- Negative- semidefinite if
- Indefinite if

More definitions

- Determine the type of the matrix

$A = I$.

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

More definitions

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

Example: $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

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

- First-order *<u>necessary</u>* optimality condition:

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

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.

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.

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

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.

➢ First-order *necessary* optimality condition:

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

Optimality conditions

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

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

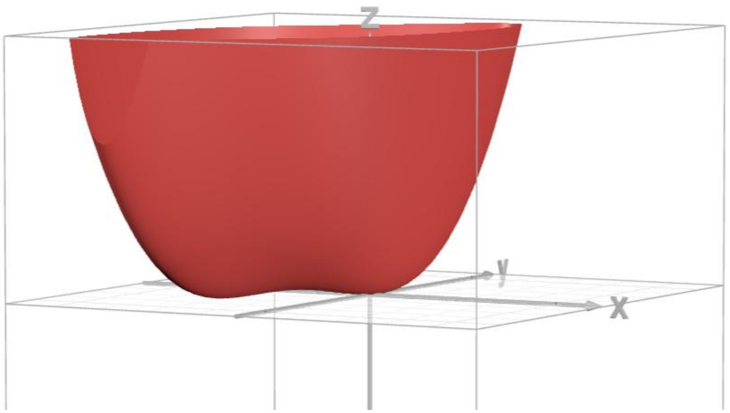
- $\nabla f(x^*) = 0$, and
- $\nabla^2 f(x^*)$ is negative-semidefinite.

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 $\max f(x)$.

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)

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

Theorem 7 - example

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

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 __________; $\boldsymbol{d_0}$ - initial direction; α – scalar. By Taylor's Theorem:

➢Basic idea (for min problem):

1. Start an iteration *k* with x_k and chose direction $\boldsymbol{d_k}$, so __________
2. Find α_k such that _________________________
3. Let $x_{k+1} =$ ___________________

➢ In what follows, $\boldsymbol{d_k}$ is chosen as __________________________

Steepest descend 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 $\boldsymbol{d_k} = -\nabla f(x_k)$

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

Newton's method

1. Let $g(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$

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:

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

$$x_{k+1} = x_k -$$

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

Example

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