Multivariate optimization

General theory and some mathematical examples

Optimization

- 2-variable Unconstrained optimization
  – Local, Global and Saddle point
  – Addressing FOC and SOC for Local optimizer
  – An example
- N-variable Unconstrained optimization
  – Generalize the result for 2-varaible case into variable case.

Local (relative) optimizers

- Definition

(x^*, y^*) is a *local* optimizer of $f(x, y)$ if and only if, for all (x', y') that are closed to (x^*, y^*),

$$f(x^*, y^*) > f(x', y') \quad \text{(local max)}$$

$$f(x^*, y^*) < f(x', y') \quad \text{(local min)}$$

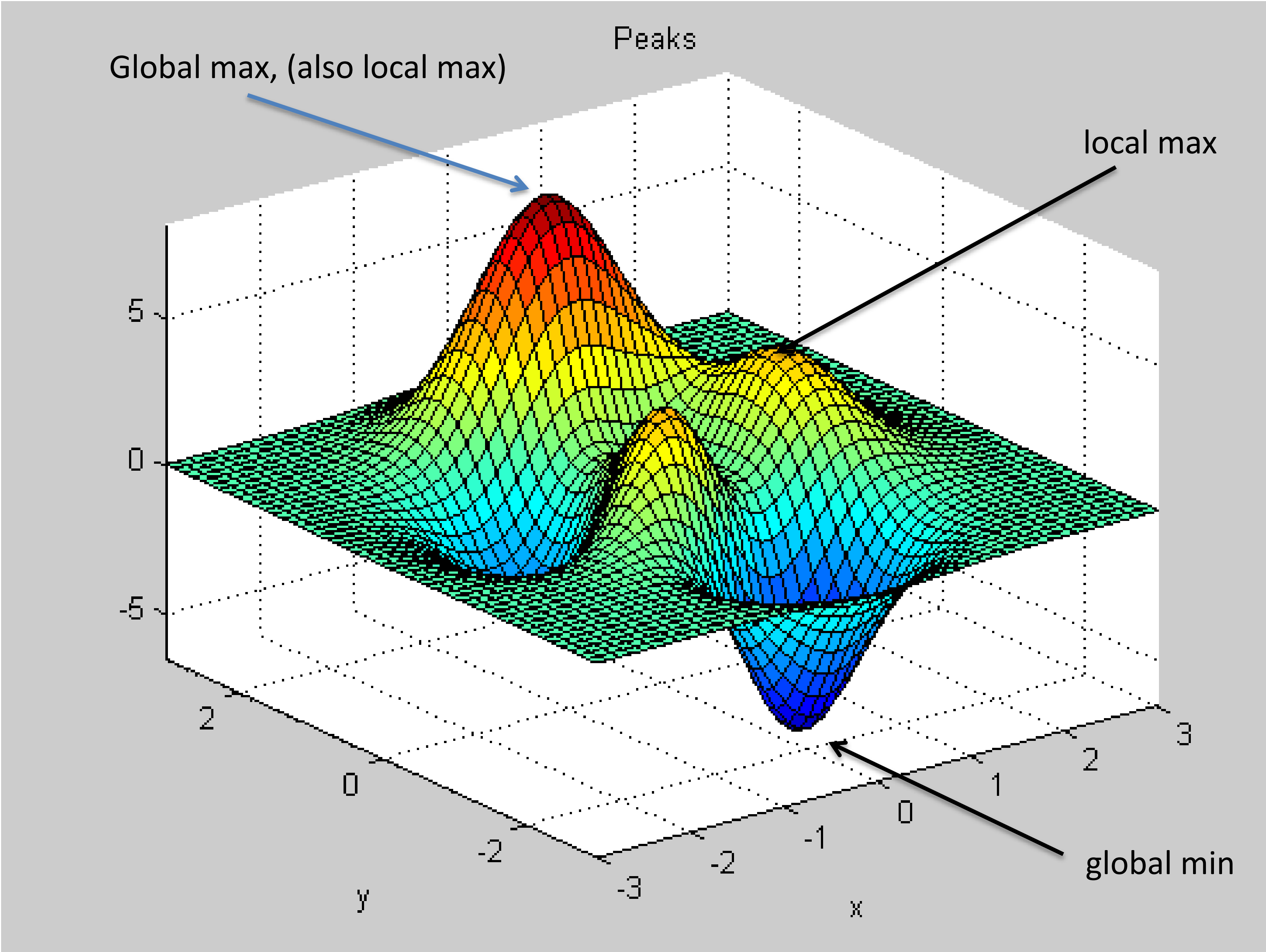
Global (absolute) optimizers

- Definition

(x^*, y^*) is a *local* optimizer of $f(x, y)$ if and only if, for all (x', y') in the entire domain set that defines "f",

$$f(x^*, y^*) > f(x', y') \quad \text{(global max)}$$

$$f(x^*, y^*) < f(x', y') \quad \text{(global min)}$$

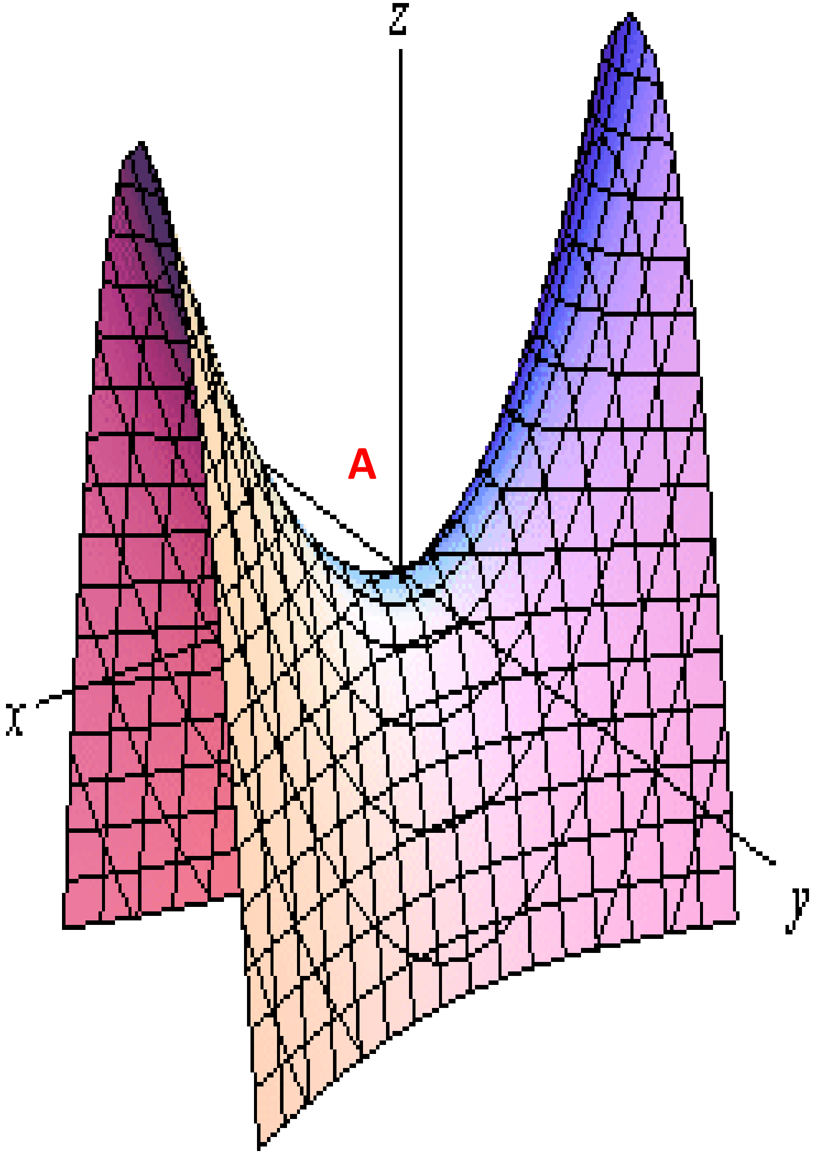

Peaks
Global max, (also local max)
local max
global min
5
0
-5
2
0
-2
y
-3
-2
-1
0
1
2
3
x

Local optimizers

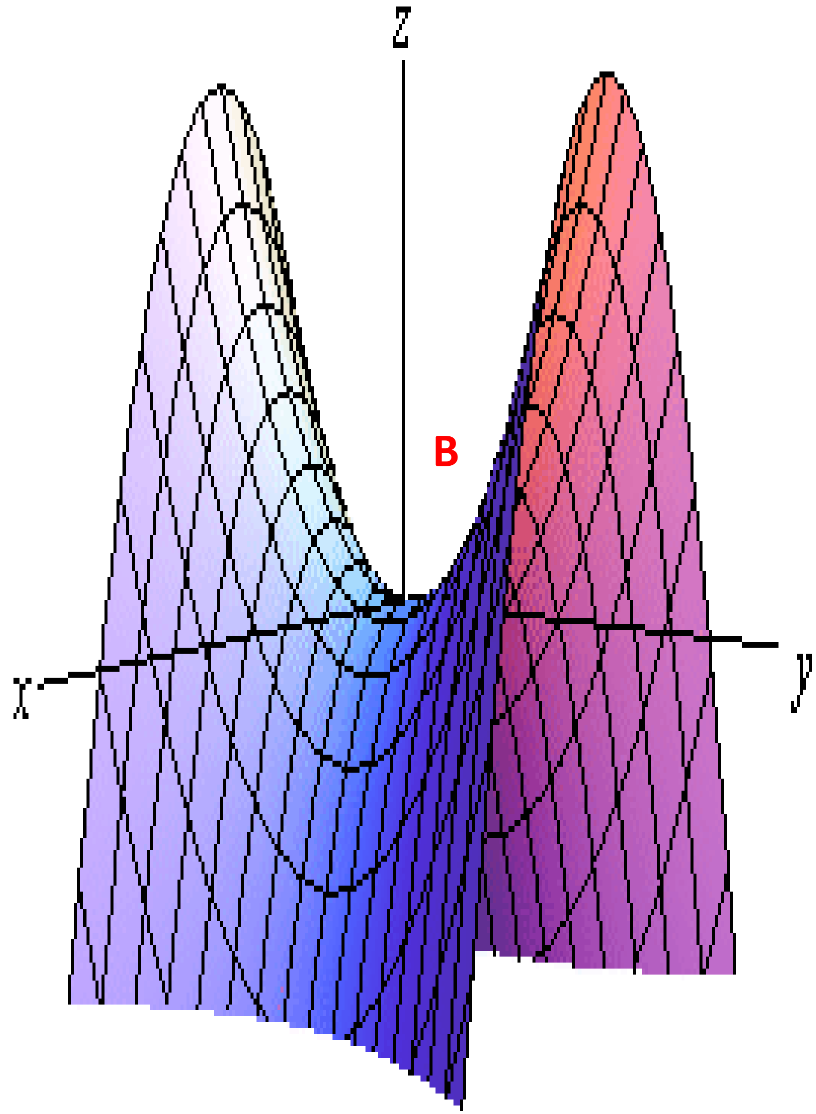
- Comment:
  - If the point is local max, it must be local max on both axis (direction).

  - If the point is local max on one axis (direction), but local min on the other axis (direction), the point is neither considered as local max nor local min. Instead, the point is called "saddle point".

Saddle point 1



Saddle point 2



Saddle point 1

- Definition

(x^*, y^*) is a *saddle point* of $f(x, y)$ if and only if, for all (x', y') that are closed to (x^*, y^*),

$$f(x^*, y^*) > f(x^*, y') \qquad \text{max in "y" axis}$$

$$f(x^*, y^*) < f(x', y^*) \qquad \text{min in "x" axis}$$

Saddle point 2

- Definition

(x^*, y^*) is a *saddle point* of $f(x, y)$ if and only if, for all (x', y') that are closed to (x^*, y^*),

$$f(x^*, y^*) < f(x^*, y')$$

min in "y" axis

$$f(x^*, y^*) > f(x', y^*)$$

max in "x" axis

Derivative method for identifying all the local optimizers

- Use the derivative method: similar to single variable case
- Recall the single variable case. We rely on the two-step derivative test.
  - First-order derivative test: *solve for all the stationary points.*
  - Second-order derivative test: *identifying maxima and minima by checking convexity/concavity of the function.*

First order condition

- Suppose that $f(x, y)$ is a differentiable function,

$$f_x(x^*, y^*) = 0 \text{ and}$$

$$f_y(x^*, y^*) = 0$$

Second-order derivative test

- Having obtained (x^*, y^*), one needs to identify/distinguish the type of local optimizer, i.e. maximizer or minimizer. (or even saddle point)

- This is done by checking the concavity/convexity at the point (x^*, y^*).

Second order condition

- Checking convexity/concavity of 2-varaible function is done by checking the value of $d^2f(x,y)$, i.e. whether it's positive or negative.

Theorem: proof omitted.

- $f(x\quad, y\quad)$ is said to be convex at the point (x^*, y^*) if and only if

$$d^2 f(x^*, y^*) > 0$$

- $f(x\quad, y\quad)$ is said to be concave at the point (x^*, y^*) if and only if

$$d^2 f(x^*, y^*) < 0$$

Deriving second-order differential

- $d^2 f(x, y) = d(d(f(x, y)))$

$$d(f(x, y)) = f_x(x, y)dx + f_y(x, y)dy$$

Deriving second-order differential

- $d\left(d\left(f(x,y)\right)\right) = \mathrm{d}\left(f_x(x,y)dx + f_y(x,y)dy\right)$

$$\mathrm{d}\left(f_x(x,y)dx\right) + \mathrm{d}\left(f_y(x,y)dy\right)$$

$$(\boldsymbol{d}\boldsymbol{f}_x(\boldsymbol{x},\boldsymbol{y}))dx \qquad (\boldsymbol{d}\boldsymbol{f}_y(\boldsymbol{x},\boldsymbol{y}))dy$$

Deriving second-order differential

$$\mathrm{d}f_x(x,y) = \mathrm{f}_{\mathrm{xx}}(\mathrm{x},\mathrm{y})\mathrm{dx} + \mathrm{f}_{\mathrm{xy}}(\mathrm{x},\mathrm{y})\mathrm{dy}$$

$$\mathrm{d}f_y(x,y) = \mathrm{f}_{\mathrm{yx}}(\mathrm{x},\mathrm{y})\mathrm{dx} + \mathrm{f}_{\mathrm{yy}}(\mathrm{x},\mathrm{y})\mathrm{dy}$$

By Young theorem: $\mathrm{f}_{\mathrm{xy}}(\mathrm{x},\mathrm{y}) = \mathrm{f}_{\mathrm{yx}}(\mathrm{x},\mathrm{y})$

Deriving second-order differential

- $d\left(d\left(f(x,y)\right)\right) =$

$$\left[\mathrm{f_{xx}(x,y)dx + f_{xy}(x,y)dy}\right]\mathrm{dx} +$$

$$\mathrm{[f_{yx}(x,y)dx + f_{yy}(x,y)dy]dy}$$

Deriving second-order differential

- $d^2 f(x, y) =$

$$\mathrm{f_{xx}(x, y)(dx)^2 + 2f_{xy}(x, y)dxd}y + \mathrm{f_{yy}(x, y)(dy)^2}$$

- By just looking at the expression, it's hard to see whether the term is positive or negative.

Deriving second-order differential

- We can rewrite the expression for $d^2 f(x, y)$ in matrix form:

$$d^2 f(x,y) = [dx \quad dy] \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix}$$

Hessian matrix

Definiteness of Hessian matrix

- If Hessian matrix is **positive definite**, then $d^2f(x^*, y^*) > 0$ for all possible combinations of dx and dy.
- If Hessian matrix is **negative definite**, then $d^2f(x^*, y^*) < 0$ for all possible combinations of dx and dy.
- How to check the definiteness of the matrix?

Testing positive/negative definiteness of a symmetric matrix

- Definition: Leading principal minors
  – Sub-matrices of "A" , indexed by A_i
  – Expand along the diagonal of the matrix.
- Example:

$$A = \begin{bmatrix} 1 & -4 & 3 \\ -4 & 5 & 0 \\ 3 & 0 & 9 \end{bmatrix} \Rightarrow A_1 = [1]; \quad A_2 = \begin{bmatrix} 1 & -4 \\ -4 & 5 \end{bmatrix}$$

$$\mathrm{A}_3 = A$$

Positive/negative definiteness matrix

- A is positive definite if the determinants of all the leading principal minors are positive.

- A is negative definite if the determinants of all the leading principal minors have alternate signs, ***with*** the negative value for the determinant of the initial leading principal minor A_1.

Example

- $A = \begin{bmatrix} -1 & 2 \\ 2 & -5 \end{bmatrix} \Rightarrow negative\ definite$

- $A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \Rightarrow positive\ definite$

Definiteness of Hessian matrix

- H is positive definite = “f” is convex
  - $\det(H_1) = f_{xx} > 0$
  - $\det(H_2) \; = f_{xx} * f_{yy} - f_{xy}^2 > 0$

- H is negative definite = “f” is concave
  - $\det(H_1) = f_{xx} < 0$
  - $\det(H_2) \; = f_{xx} * f_{yy} - f_{xy}^2 > 0$

Example

- Determine all the optimizers of
$$f(x,y) = x^2 - 2xy + 3y^2 + 2x - 2y$$

N-Variable case

$$z = f(x_1, x_2, x_3, \ldots, x_n)$$

FOC.

$$f_1 = f_2 = f_3 = \cdots = f_n = 0$$

N-Variable case

$$z = f(x_1, x_2, x_3, \ldots, x_n)$$

SOC. Check definiteness of H.

$$H = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}$$

N-Variable case

- H is negative definite ➔
  - $|\mathrm{H}_1| < 0, |\mathrm{H}_2| > 0, |\mathrm{H}_3| < 0$ ….
  - sign($|\mathrm{H}_\mathrm{i}|$) = $(-1)^i$ for each “i”.
  - Local min

- H is positive definite ➔
  - $|\mathrm{H}_1| > 0, |\mathrm{H}_2| > 0, |\mathrm{H}_3| > 0$ ….
  - sign($|\mathrm{H}_\mathrm{i}|$) > 0 for *all* “i”.
  - Local max

Optional: Saddle point

- Saddle point is a stationary point, i.e. FOC holds.
- Saddle point can be identified by checking if the following condition holds
  - $f_{xx} * f_{yy} - f_{xy}^2 < 0 \Rightarrow f_{xx} * f_{yy} < f_{xy}^2$
    - f_{xy}^2 is always positive.
    - $f_{xx} * f_{yy}$ must be less than zero.
    - Intuitively, “f” must be convex in one direction, but concave in the other direction. That’s the function produces maximum point along the direction in which function is concave, and minimum point along the direction in which function is convex.