10.018 Modelling Space and Systems

Lecture 2:
Minima & Maxima

Term 2, 2021

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Before we start....

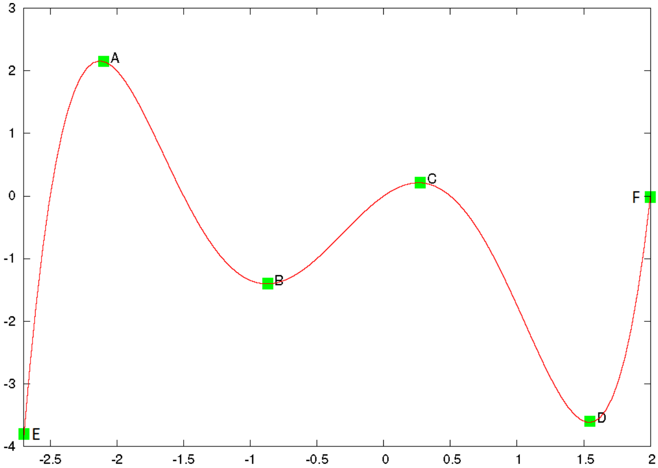
To get the most out of this lecture, you should already be familiar with

1. Critical points and extrema (Math 1)
2. Second derivative test (Math 1)
3. Logic (Math 1)
4. Surfaces in 3D space (Math 2)

as we will be going through

1. Maxima, minima, saddle points: The 2D analogue of maxima / minima / point of inflexion
2. Second derivative test: The 2D analogue of the second derivative test

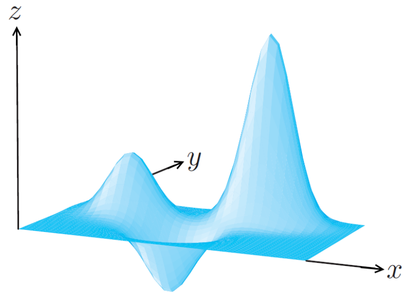
Activity 1 (5 minutes)



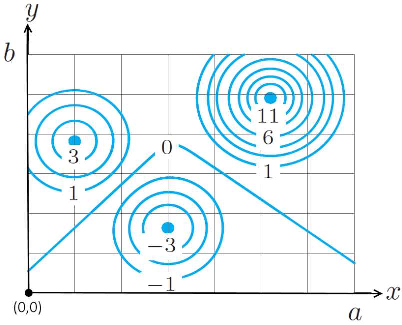
State the local minima, local maxima, global minimum, global maximum for the function shown.

Optimization

Similarly, functions of several variables can have *local* and *global* extrema.



Graph of function $f(x,y)$.



Contour map of function $f(x,y)$.

Global vs. local

Suppose we have the unconstrained optimization problem:

$$\min\{f(x) : \vec{x} \in \mathbb{R}^n\} \qquad \text{or} \qquad \max\{f(x) : \vec{x} \in \mathbb{R}^n\} \qquad \text{(P)}.$$

Definition (Local min/max)

The point $\vec{x}^*$ is a local minimum of (P) if there exists $\varepsilon > 0$ (possibly very small) such that $f(\vec{x}^*) \leq f(\vec{x})$ for all points $\vec{x} \in \mathbb{R}^n$ such that $\|\vec{x}^* - \vec{x}\| \leq \varepsilon$.

The point $\vec{x}^*$ is a local maximum of (P) if there exists $\varepsilon > 0$ (possibly very small) such that $f(\vec{x}^*) \geq f(\vec{x})$ for all points $\vec{x} \in \mathbb{R}^n$ such that $\|\vec{x}^* - \vec{x}\| \leq \varepsilon$.

Here $\|\vec{x}^* - \vec{x}\| \leq \varepsilon$ just means that "points $\vec{x}$ are *near* the point $\vec{x}^*$".

Global vs. local

Suppose we have the unconstrained optimization problem:

$$\min\{f(x) : \vec{x} \in \mathbb{R}^n\} \qquad \text{or} \qquad \max\{f(x) : \vec{x} \in \mathbb{R}^n\} \qquad \text{(P)}.$$

Definition (Global min/max)

The point $\vec{x}^*$ is a global minimum of (P) if $f(\vec{x}^*) \leq f(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$.

The point $\vec{x}^*$ is a global maximum of (P) if $f(\vec{x}^*) \geq f(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$.

Max and min for single variable functions

For a twice-differentiable, **single variable** function $f(x)$, a critical point is a point x_0 where $f'(x_0) = 0$ or undefined. The critical point is a:

- Local *maximum* if $f''(x_0) < 0$,
- Local *minimum* if $f''(x_0) > 0$,
- If $f''(x_0) = 0$ or $f'(x) = 0$ is undefined then the second derivative test is inconclusive. Need to investigate further. For example, if f'' changes sign near x_0 then x_0 is a *point of inflection*.

We will learn that the first and second order derivatives of a two-variable function also give us information about the maxima and minima of the function.

A blast from the past: Math 1

Conditional propositions, continued

Equivalent ways to say $p \Rightarrow q$:

- If p, then q.
- p implies q.
- q, if p.
- p, only if q.
- p is a sufficient condition for q.
- q is a necessary condition for p.

Try to understand the sentences above using p : I am a father, q : I am a man.

Negation: The negation of $p \Rightarrow q$ is the case when the deduction is false, which is p AND (NOT q).

A note on necessary versus sufficient!

It is necessary to have clouds, when it is raining.

But having clouds is NOT sufficient to have rain.

Rain $\Rightarrow$ Clouds. But Clouds $\not\Rightarrow$ Rain.

A note on necessary versus sufficient!

Based on what we discussed just now write the following statements for the function of one variable:

- What's the necessary condition for $f(x)$ to have a minimum at x_0?

  Answer:______________________________

- What's the sufficient condition for $f(x)$ to have a minimum at x_0?

  Answer:______________________________

- What's the necessary condition for $f(x)$ to have a maximum at x_0?

  Answer:______________________________

- What's the sufficient condition for $f(x)$ to have a maximum at x_0?

  Answer:______________________________

The Gradient Vector

Suppose we have $f(\vec{x}), x \in \mathbb{R}^n$. We define

$$\nabla f(\vec{x}) = \operatorname{grad} f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

You can think of this as a vector of partial derivatives.
In our (x, y) notation, we can write

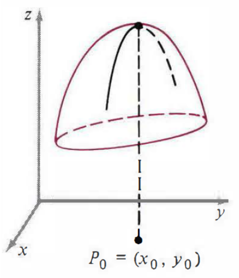
$$\nabla f(x, y) = \operatorname{grad} f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$$

The vector $\nabla f(x, y)$ is called a gradient vector.

Max and min for functions of two variables

Suppose a function $f(x, y)$ has a local maximum at a point (x_0, y_0) where it is *differentiable*. What must happen to the partial derivatives at that point?

Because (x_0, y_0) is a local maximum of $f(x, y)$, x_0 must be a local maximum of the *single variable* function $f(x, y_0)$.



Therefore

$$\left.\frac{\mathrm{d}f(x, y_0)}{\mathrm{d}x}\right|_{x=x_0} = f_x(x_0, y_0) = 0.$$

Similarly, $f_y(x_0, y_0) = 0$.

The same analysis applies if the point is a local minimum.

Max, min and the gradient

Necessary condition for min/max

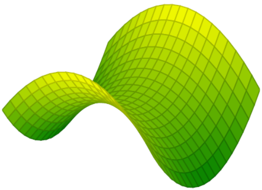
Therefore, if a function $f(x, y)$ has a local maximum or local minimum at a point (x_0, y_0) where it is *differentiable*, then

$$f_x(x_0, y_0) = f_y(x_0, y_0) = 0, \quad \text{equivalently,} \quad \nabla f(x_0, y_0) = \vec{0}.$$

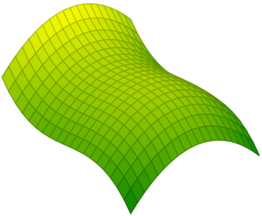
Note that this is *a necessary* condition: if $\nabla f(x_0, y_0) = \vec{0}$, then it is *not* always true that (x_0, y_0) is a local maximum or minimum. Just like (critical) points of inflection, the function $f(x, y)$ can also have a critical point (that is, where either the gradient is zero or undefined) that is not a local extremum. A non-extremal critical point where the function is differentiable is known as a saddle point.

Saddle points

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



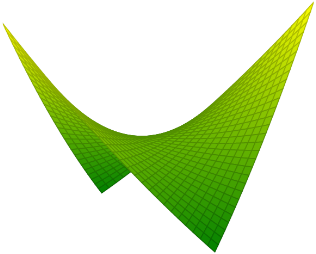
$$f(x,y) = y^3 - 2x^2$$



The surface on the left exhibits a *saddle point*: the origin is a local minimum in one direction and a local maximum in another direction.

The surface on the right also has a saddle point: the point is a local maximum in one direction, and a point of inflection in another direction.

Saddle point, example



$$f(x, y) = x^2 + 10xy + y^2$$
$$\nabla f(x, y) = [2x + 10y, 10x + 2y]$$

Solving the *linear system* $\nabla f(x, y) = \vec{0}$, we find that $x = y = 0$.

$$f_x(0, 0) = f_y(0, 0) = 0, \quad \text{also}$$
$$f_{xx}(0, 0) = f_{yy}(0, 0) = 2,$$
$$f_{xy}(0, 0) = f_{yx}(0, 0) = 10.$$

So at $(0, 0)$, all the first order derivatives are 0, and all the second order derivatives are positive, but unlike in the single variable case, it is not a local minimum (it is in fact a saddle point).
We need a way to identify saddle points and local extrema.

Second derivative test (Sufficient condition)

Let (x_0, y_0) be a critical point of f when $\nabla f(x_0, y_0) = \vec{0}$. Define the quantity $D = D(x_0, y_0)$ by

$$D(x_0, y_0) = f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - f_{xy}(x_0, y_0) f_{yx}(x_0, y_0).$$

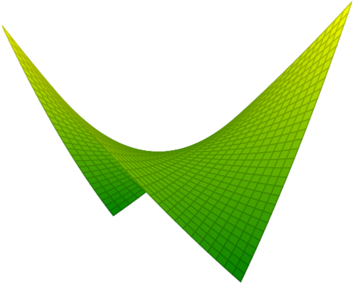
Second derivative test

When $\nabla f(x_0, y_0) = \vec{0}$, if

- $D < 0$, then (x_0, y_0) is a saddle point.
- $D > 0$ and $f_{xx} > 0$, then (x_0, y_0) is a local minimum.
- $D > 0$ and $f_{xx} < 0$, then (x_0, y_0) is a local maximum.
- $D = 0$, then this test is inconclusive and another method is required to determine the nature of the critical point.

For the derivation see the end of the slides after Summary.

Saddle point, revisited



$$f(x, y) = x^2 + 10xy + y^2$$

$$f_x(0,0) = f_y(0,0) = 0,$$
$$f_{xx}(0,0) = f_{yy}(0,0) = 2,$$
$$f_{xy}(0,0) = f_{yx}(0,0) = 10.$$

At the critical point $(0, 0)$,

$$f_{xx}(0,0) = 2, f_{xy}(0,0) = f_{y,x}(0,0) = 10, f_{y,y}(0,0) = 2$$

$D = -96 < 0$, so $(0, 0)$ is a saddle point.

The Big Question

Some of you may wonder: *For practical problems, most of the time (stochastic) gradient descent (or generic Newton Raphson) is used to compute maxima or minima, and all we need to know is the* **computation of the first and second partial derivatives**.

Moreover, if the function is convex / concave (from Math I), I know that any optimum I find is the global optimum. If not, I will run stochastic gradient descent with many random starting points to find the global optimum.

Doesn't this mean I just need to know how to compute the partial derivatives and use an appropriate software package?

Summary

We have covered:

- Critical points, local maxima, local minima and saddle points.
- Finding critical points using the gradient.
- The second derivative test for determining the nature of a critical point.

Textbook: read Section 19.7, then try some of Exercises 19.7.1–19.7.31. You may discuss them on Piazza.

[OPTIONAL] 2nd derivative test (Sufficient condition) – proof

Given a critical point (x, y) and a unit vector $\vec{u} = [u_1, u_2]$, the idea is to compute the *second order directional derivative*. Recall that $D_{\vec{u}} f = \nabla f \cdot \vec{u}$.

$$\begin{aligned} D_{\vec{u}}(D_{\vec{u}} f) &= D_{\vec{u}}(f_x u_1 + f_y u_2) \\ &= (f_x u_1 + f_y u_2)_x u_1 + (f_x u_1 + f_y u_2)_y u_2 \\ &= (f_{xx} u_1 + f_{yx} u_2) u_1 + (f_{xy} u_1 + f_{yy} u_2) u_2 \\ &= f_{xx} u_1^2 + 2 f_{xy} u_1 u_2 + f_{yy} u_2^2. \end{aligned}$$

Completing the square (as long as $f_{xx} \neq 0$):

$$\begin{aligned} D_{\vec{u}}(D_{\vec{u}} f) &= f_{xx} \left[u_1^2 + 2 \frac{f_{xy}}{f_{xx}} u_1 u_2 + \left(\frac{f_{xy}}{f_{xx}} \right)^2 u_2^2 \right] - \frac{f_{xy}^2}{f_{xx}} u_2^2 + f_{yy} u_2^2 \\ &= f_{xx} \left[u_1 + \frac{f_{xy}}{f_{xx}} u_2 \right]^2 + \frac{u_2^2}{f_{xx}} \left(f_{xx} f_{yy} - f_{xy}^2 \right). \end{aligned}$$

Second derivative test (Sufficient condition) – proof

$$D_{\vec{u}}(D_{\vec{u}} f) = f_{xx}\left[u_1 + \frac{f_{xy}}{f_{xx}} u_2\right]^2 + \frac{u_2^2}{f_{xx}}\left(f_{xx} f_{yy} - f_{xy}^2\right).$$

If $f_{xx} > 0$ and $D = f_{xx} f_{yy} - f_{xy}^2 > 0$, then $D_{\vec{u}}(D_{\vec{u}} f) > 0$. I.e. the function has a local minimum in any direction $\vec{u}$, therefore (x, y) is a local minimum.

Similarly, if $f_{xx} < 0$ and $D > 0$, then $D_{\vec{u}}(D_{\vec{u}} f) < 0$, so the function has a local maximum in any direction, therefore (x, y) is a local maximum.

Second derivative test (Sufficient condition) – proof

$$D_{\vec{u}}(D_{\vec{u}}f) = f_{xx}\left[u_1 + \frac{f_{xy}}{f_{xx}}u_2\right]^2 + \frac{u_2^2}{f_{xx}}\left(f_{xx}f_{yy} - f_{xy}^2\right).$$

If $D = 0$, then $D_{\vec{u}}(D_{\vec{u}}f) \geq 0$ or ≤ 0. The lack of strict inequality means the test is inconclusive.

If $D < 0$, then we can pick different values of u_1 and u_2 such that $D_{\vec{u}}(D_{\vec{u}}f)$ can be positive or negative, thus (x, y) is a saddle.

For instance, if $u_1 = 1, u_2 = 0$, then $D_{\vec{u}}(D_{\vec{u}}f) = f_{xx}$, but if $u_1/u_2 = -f_{xy}/f_{xx}$, then $D_{\vec{u}}(D_{\vec{u}}f) = \frac{u_2^2}{f_{xx}}D$. These two derivatives have opposite signs.