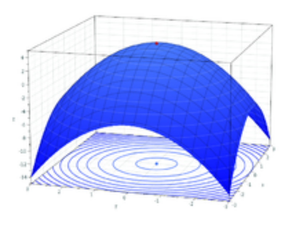
Lecture 12 - Optimization by Calculus

COMP1046 - Maths for Computer Scientists

Dr. Tony Bellotti



Learning outcomes

Prerequisites:

- Limits, Derivatives and Derivative Rules

By the end of this lecture we will have learned:

- Maxima and Minima
- Optimization
- Gradient Descent

Based on Sections 1.1 to 1.5, 2.1 to 2.4, 2.7 and 3.7 of the textbook (Smith and Minton 2002).

Maxima and Minima

Maxima and Minima

We can use differential calculus to find the maximum or minimum values of differentiable functions. This is very useful for problems that require optimization. For example,

- Maximizing profit;
- Minimizing costs;
- Optimizing resource use.

Extrema

Definition

For some function $f(x)$,

(i) $f(c)$ is a *local maximum* of f if $f(c) \geq f(x)$ for all x in some open interval containing c.

(ii) $f(c)$ is a *local minimum* of f if $f(c) \leq f(x)$ for all x in some open interval containing c.

In either case, we call $f(c)$ a *local extremum* of f.

Fermat's Theorem

Definition

Let $f'(c)$ denote the derivative of f at value c.
A number c in the domain of f is called a *critical number* of f if $f'(c) = 0$ or $f'(c)$ is undefined.

Theorem (Fermat's Theorem)

Suppose that $f(c)$ is a local extremum. Then c must be a critical number of f.

Fermat's Theorem

Proof.

Suppose f is differentiable at $x = c$. (If not, c is a critical number and we are done.) Suppose further that $f'(c) \neq 0$.

If $f'(c) > 0$,

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} > 0.$$

- For $h > 0$, $f(c+h) - f(c) > 0 \Rightarrow f(c+h) > f(c)$;
- For $h < 0$, $f(c+h) - f(c) < 0 \Rightarrow f(c+h) < f(c)$.

Thus, $f(c)$ is not a local maximum or minimum.

...continued...

Fermat's Theorem

Proof.

Since we had assumed that $f(c)$ is a local extremum, this is a contradiction. Hence $f'(c) \leq 0$.

Similiarly, if $f'(c) < 0$, we obtain a similar contradiction, through a similar argument.

Therefore the only remaining possibility is $f'(c) = 0$. $\square$

Example

Find the critical numbers and local extrema of
$f(x) = 2x^3 - 3x^2 - 12x + 5$.

$$\begin{aligned} f'(x) &= 6x^2 - 6x - 12 \\ &= 6(x^2 - x - 2) \\ &= 6(x-2)(x+1) \end{aligned}$$

Critical numbers are $x = 2$ and $x = -1$ since either of these make $f'(x) = 0$. Corresponding local extrema are

$$f(2) = 16 - 3 \times 4 - 24 + 5 = -15$$

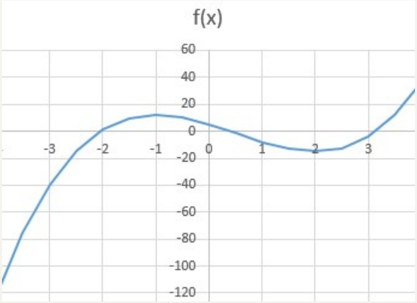
and

$$f(-1) = -2 - 3 + 12 + 5 = 12.$$

Fermat's Theorem

Example

$$f(x) = 2x^3 - 3x^2 - 12x + 5$$



Exercise 4: Fermat's Theorem

Find the critical numbers and local extrema of
$f(x) = x^3 - 3x + 3$.

Exercise 4: *Solution*

To be completed.

Optimization problems

Optimization problems

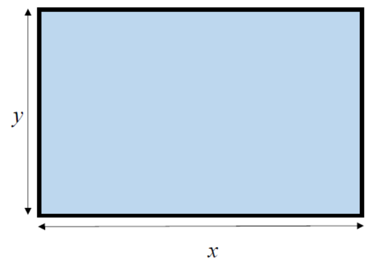
We can use differential calculus and Fermat's Theorem to solve many optimization problems.

In this section we give some examples.

The Fenced Garden

Example

You have 40 (linear) meters of fencing with which to enclose a rectangular space for a garden. Find the *largest* area that can be enclosed with this much fencing and the dimensions of the corresponding garden.



Rectangular garden plot with lengths x and y on either side.

The Fenced Garden

Example

- We can write a formula for the garden perimeter which must be 40 meters:

$$2x + 2y = 40 \quad \Rightarrow \quad y = 20 - x.$$

- We can also write a formula for the area

$$A = xy = x(20 - x) = 20x - x^2.$$

- We want to maximum the area, so find the critical point of A as a function of x:

$$\frac{d}{dx}(20x - x^2) = 20 - 2x.$$

The Fenced Garden

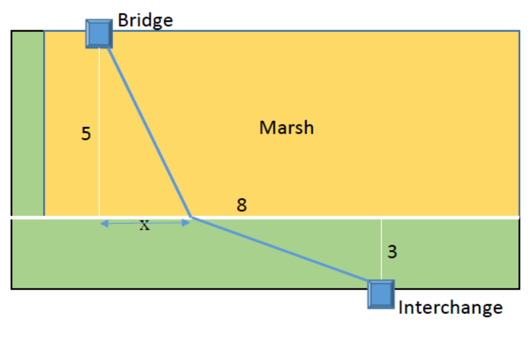
Example

- This has a critical point ($20 - 2x = 0$) when $x = 10$.
- Hence the solution is $x = y = 10$.
- that is, a square garden, 10 meters each side and area 100 meters square.

Building a Highway

Example

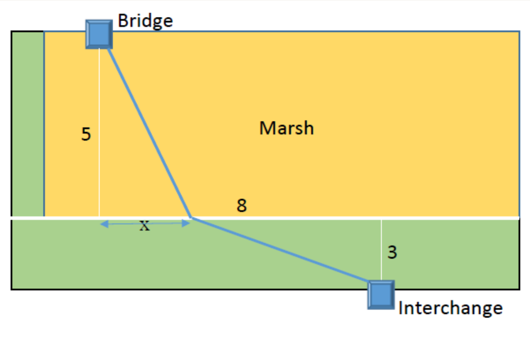
The state wants to build a new stretch of highway to link an existing bridge with a turnpike interchange, located 8 miles to the east and 8 miles to the south of the bridge. There is a 5-mile-wide stretch of marsh land adjacent to the bridge which must be crossed.



Building a Highway

Example

Given that the highway costs \$10 million per mile to build over marsh and only \$7 million to build over dry land, how far east of the bridge should the highway be when it crosses out of the marsh?



Building a Highway

Example

Total cost is

$10 \times (\text{distance across marsh}) + 7 \times (\text{distance across dry land})$.

Use Pythagoras' Theorem for distances:-

- distance across marsh $= \sqrt{5^2 + x^2}$;
- distance across dry land $= \sqrt{(8-x)^2 + 3^2}$

so total cost as a function of x is

$$C(x) = 10\sqrt{5^2 + x^2} + 7\sqrt{(8-x)^2 + 3^2}.$$

Building a Highway

Example

$$C(x) = 10\sqrt{5^2 + x^2} + 7\sqrt{(8-x)^2 + 3^2}.$$

Now differentiate to find extrema:

$$\begin{aligned} C'(x) &= 10\tfrac{1}{2\sqrt{25+x^2}}(2x) + 7\tfrac{1}{2\sqrt{(8-x)^2+9}} \times 2(8-x)(-1) \\ &= \tfrac{10x}{2\sqrt{25+x^2}} - \tfrac{7(8-x)}{2\sqrt{(8-x)^2+9}} \end{aligned}$$

Critical values need to be computed for $C'(x)$. The only way is to use a numerical method. For example, use bisection on the interval $[0, 8]$. If this is done the critical value is

$$c \approx 3.56.$$

Building a Highway

Example

Then the cost for $c \approx 3.56$. is $C(c) \approx \$98.9$ million.

This is a saving of more than \$10 million compared to planning the highway around the marsh; i.e. $C(0) \approx \$109.8$ million.

Exercise 5: Optimization

You manage a factory manufacturing headphones. You need to decide how to price your product and estimate how many headphones you will need to make over the next month. Suppose the factory manufactures r headphones and sell them at p RMB each.

- The total cost to manufacture these headphones is given by the formula $5r + 10000$ RMB.
- Since people will resist paying a high price, you project that the number of headphones that can be sold is related to price by the formula $10^6/p^2$.

Based on maximizing gross profit (i.e. without considering tax or other costs), determine the optimal price p and number of headphones r to manufacture. How much gross profit is estimated?

Exercise 5: *Solution*

To be completed.

Gradient Descent

Gradient Descent

As seen in the last example, it may be that a critical point cannot be found *analytically*. That is, $f'(x) = 0$ cannot be rearranged in terms of x.

For these problems, a numeric method is required.

A popular approach is to use *gradient descent*.

Gradient descent is the basic optimization technique for neural networks, e.g., which involve complex minimization problems.

Gradient Descent

The idea of gradient descent is to start at some point on the curve f and move in the downward direction of the slope in small steps until a local minima is reached. The direction at a point a is given by $f'(a)$:

- if $f'(a) > 0$ then the slope is upwards so move to some new point $a' < a$ to go down.
- if $f'(a) < 0$ then the slope is downwards so move to some new point $a' > a$ to go down.
- if $f'(a) = 0$ then a is already an extremum.

Gradient Descent

Gradient descent operates over iterations $i = 0, 1, 2, \cdots$, so that

- At iteration 1, $a^{[0]}$ = starting point;
- At iteration $i + 1$,

$$a^{[i+1]} = a^{[i]} - \eta f'(a^{[i]})$$

where $\eta > 0$ is a constant that controls the size of each step.

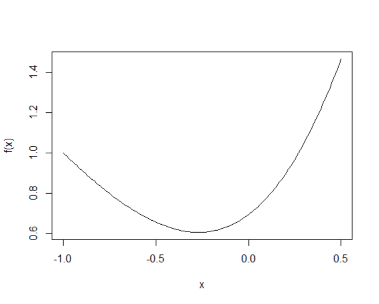
This algorithm describes gradient descent for minimization. It can easily be used for maximization, however, by changing the sign: $a^{[i+1]} = a^{[i]} + \eta f'(a^{[i]})$.

Gradient Descent

Example

Consider finding the minimum of

$$f(x) = x^2 + (x+1)\ln(x^2+2):$$



The curve shows a clear minima beween -0.5 and 0.

Example

$f(x) = x^2 + (x+1)\ln(x^2+2) \Rightarrow$

$$f'(x) = 2x + \ln(x^2+2) + \frac{x+1}{x^2+2}2x^2.$$

It is not possible to rearrange this to compute x analytically from $f'(x) = 0$, so instead we can use gradient descent. See what happens with each iteration of gradient descent, starting at $a = 0.5$ and with $\eta = 0.2$ on the next slide.

Gradient Descent

Example

i	a	$f'(a)$	$f(a)$
0	0.5	2.4776	1.4664
1	0.0045	0.7066	0.6963
2	-0.1368	0.3118	0.6251
3	-0.1992	0.158	0.6105
4	-0.2308	0.0849	0.6067
5	-0.2478	0.047	0.6055
6	-0.2572	0.0264	0.6052
7	-0.2625	0.015	0.6051
8	-0.2655	0.0085	0.605
9	-0.2672	0.0049	0.605
10	-0.2681	0.0028	0.605

Multivariate Optimization

Optimization using maximization/minimization techniques can be extended to *multivariate* frameworks.

That is, to functions of two or more variables. This is useful if we want to optimize with respect to more than one parameter (as we often do).

This is done using *partial* derivatives. That is,

$$\frac{\partial f(x, y)}{\partial x}$$

is the derivative of f with respect to x whilst keeping y fixed.

Multivariate Gradient Descent

In particular, gradient descent very easily generalizes to the multivariate case.

Consider minimizing a differentiable function $f(x_1, \cdots, x_n)$ over n variables.

Gradient descent uses the gradient defined as

$$\nabla f(x_1, \cdots, x_n) = \left(\frac{\partial f(x_1, \cdots, x_n)}{\partial x_1}, \cdots, \frac{\partial f(x_1, \cdots, x_n)}{\partial x_n} \right)$$

and operates over iterations $i = 0, 1, 2, \cdots$ to compute the optimal *vector* of n values, so that

- At iteration 1, $\mathbf{a}^{[0]}$ = starting vector of n values;
- At iteration $i + 1$,

$$\mathbf{a}^{[i+1]} = \mathbf{a}^{[i]} - \eta \nabla f(\mathbf{a}^{[i]})$$

where $\eta > 0$ is a constant that controls the size of each step.

Multivariate Optimization

We do not cover the topic of multivariate calculus or optimization in any further detail in this course.

Summary

We have covered the following topics in this lecture:

- Maxima and Minima
- Optimization
- Gradient Descent