10.018 Modelling Space and Systems

Cohort 2.1

The Gradient Vector and Chain Rule

Term 2, 2021

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Before we start....

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

1. Chain rule (Math I)
2. Tangent plane approximation (Math II)
3. Directional derivatives (Math II)

as we will be going through

1. Properties of the gradient vector
2. Chain rule in Math II: The 2D analogue of the chain rule

Directional derivatives – formula (seen last week)

There is an easier way of finding the directional derivative without taking the limit. Consider

$$D_{\vec{u}} f(a,b) = \lim_{h\to 0} \frac{f(a + h\,u_1,\, b + h\,u_2) - f(a,b)}{h} = \lim_{h\to 0} \frac{\Delta f}{h}$$

From the tangent plane approximation

$$\Delta f \approx f_x\,\Delta x + f_y\,\Delta y = f_x\,hu_1 + f_y\,hu_2$$

Plugging this into the definition:

$$\frac{\Delta f}{h} \approx \frac{f_x(a,b)\,hu_1 + f_y(a,b)\,hu_2}{h} = f_x(a,b)\,u_1 + f_y(a,b)\,u_2$$

This approximation becomes exact as $h \to 0$ and thus we have

Directional derivative formula

For a *unit vector* $\vec{u} = [u_1, u_2]$, the directional derivative of f in the direction of $\vec{u}$ is given by

$$D_{\vec{u}} f(a,b) = f_x(a,b)\,u_1 + f_y(a,b)\,u_2.$$

The gradient vector

Note that the expression for $D_{\vec{u}}f(a,b)$ could be written as a dot product of $\vec{u}$ and *another vector*:

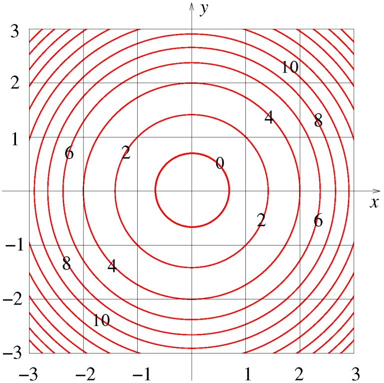
$$\begin{aligned} D_{\vec{u}}f(a,b) &= f_x(a,b)\,u_1 + f_y(a,b)\,u_2 \\ &= [f_x(a,b), f_y(a,b)] \cdot [u_1, u_2] \\ &= \nabla f(a,b) \cdot \vec{u} \end{aligned}$$

where as mentioned in the lecture, the vector $\nabla f(a,b)$ is the gradient vector.

Activity 1 (10 minutes)

(1) Using the contour map below, determine whether $D_{\vec{u}} f(\vec{p})$ is positive, negative, or zero:

$\vec{p}$	$\vec{u}$
$(1.5, 0)$	$[1, 0]$
$(1, 1)$	$[-1, 1]$
$(2, -2)$	$[-2, 1]$



(2) Which directional derivative has the larger value: at $\vec{p} = (1, 0)$ in direction $\vec{u} = [1, 0]$, or at $\vec{p} = (0, -2)$ in direction $\vec{u} = [0, -1]$?

Activity 1 (solution)

(1)

$\vec{p}$	$\vec{u}$	sign
$(1.5, 0)$	$[1, 0]$	$+$
$(1, 1)$	$[-1, 1]$	0
$(2, -2)$	$[-2, 1]$	$-$

(2) The directional derivative of $\vec{p} = (0, -2)$ in direction $\vec{u} = [0, -1]$ has the larger value, because the level curves are closer to each other.

What Does the Gradient Tell Us? Activity 2 (10 minutes)

Recall that a directional derivative of $f(x, y)$ at a point (a, b) in the direction $\vec{u}$ is given by $D_{\vec{u}} f(a, b) = \nabla f(a, b) \cdot \vec{u}$.

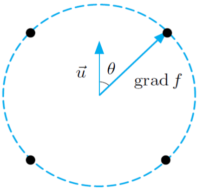


Figure 1

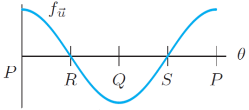


Figure 2

Fig 2 is a graph of the directional derivative $D_{\vec{u}} f$ at the point (a, b) versus θ, the angle in Fig 1.

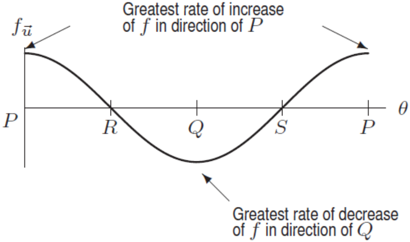
1. Which points on the graph in Fig 2 correspond to the greatest rate of increase of f? The greatest rate of decrease?
2. Mark points on the circle in Fig 1 corresponding to the points P, Q, R, S.
3. What is the amplitude of the function graphed in Fig 2? What is the exact formula for that function?
4. How are the level sets of f related to points R, S?

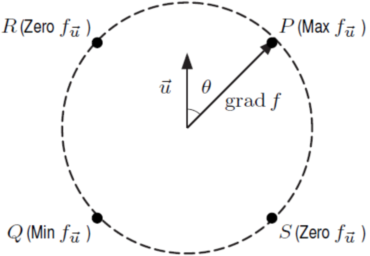
Activity 2 (solutions)

1. P corresponds to greatest rate of increase of f and Q corresponds to greatest rate of decrease of f.
2. The points are marked on the figure.
3. Amplitude is $\|\nabla f\|$. The equation is

$$f_{\vec{u}} = \|\nabla f\| \cos\theta.$$

4. Directions $\vec{u}$ corresponding to R and S are tangents to the level sets of f.





Geometric Properties of the Gradient Vector of $f(x,y)$

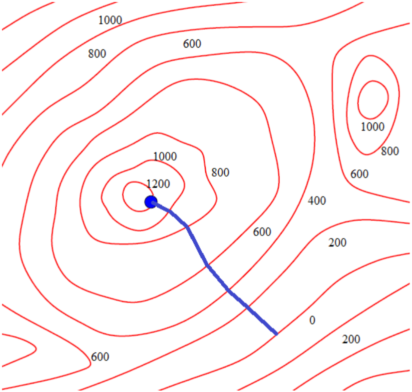
From the previous activities we can summarize the following properties of the gradient vector.

If f is a differentiable function at the point (a,b) and $\nabla f(a,b) \neq \vec{0}$, then:

- The direction of $\nabla f(a,b)$ is:
  - Perpendicular* to the contour of f through (a,b);
  - In the direction of the maximum rate of increase of f.
- The magnitude of the gradient vector, $\|\nabla f\|$, is:
  - The maximum rate of change of f at that point;
  - Large when the contours are close together and small when they are far apart.

* This assumes that the same scale is used on both axes.

Gradient Descent



This is a basis for an extremely important method of finding a minimum of a function: gradient descent. It is used extensively in engineering fields, statistics, machine learning etc.
For more information check wiki `https://en.wikipedia.org/wiki/Gradient_descent`.

Break

5 min break

Don't be late.

Motivation

Recall from Math 1: for a function of one variable, $y = f(x)$ where $x = g(t)$ (that is, $y = f(g(t))$), the **chain rule** states that

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}.$$

The chain rule is used very frequently. It is helpful in both differentiation and integration.

It allows us to treat $\frac{\mathrm{d}y}{\mathrm{d}x}$ and $\frac{\mathrm{d}x}{\mathrm{d}t}$ like fractions, even though they are *not* (it's just a mnemonic rule).

We now generalize the chain rule to multivariable functions.

Multivariable chain rule

Let $z = f(x, y)$, where $x = g(t)$ and $y = h(t)$.

Then $z = f(g(t), h(t))$ is a function of t.

Recall from Cohort 1.2 the tangent plane approximation, written with slightly different notation here:

$$\Delta z = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y + \text{error terms.}$$

Divide both sides by Δt:

$$\frac{\Delta z}{\Delta t} = \frac{\partial z}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial z}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\text{error terms}}{\Delta t}.$$

Fact: if the partial derivatives are continuous, then (error terms)$/\Delta t \to 0$ as $\Delta t \to 0$. Upon taking the limit $\Delta t \to 0$:

Multivariable chain rule

Chain rule for a function of two variables

For $z = f(x, y)$, where $x = g(t)$ and $y = h(t)$,

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}.$$

Note: the formula contains both ordinary and partial derivatives!

Example

Let $z = x^2 y + y^2 x$, where $x = t^2$ and $y = t^3$. Compute $\frac{\mathrm{d}z}{\mathrm{d}t}$.

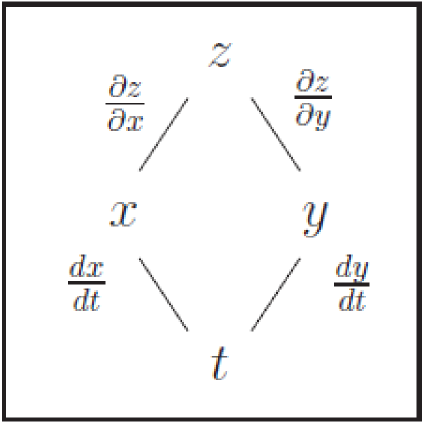
$$\begin{aligned}\frac{\mathrm{d}z}{\mathrm{d}t} &= \left(2xy + y^2\right)(2t) + \left(x^2 + 2yx\right)\left(3t^2\right)\\ &= \left(2t^5 + t^6\right)(2t) + \left(t^4 + 2t^5\right)\left(3t^2\right) = 7t^6 + 8t^7.\end{aligned}$$

We can check the answer directly, since by substitution $z = t^7 + t^8$.

Chain Rule Diagram

There is a nice diagram you can use to remember the formula

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}.$$



Activity 3 (5 minutes)

Let $z = xy$, where $x = \sin(t)$ and $y = \cos(t)$. Compute $\frac{\mathrm{d}z}{\mathrm{d}t}$, using

(1) direct substitution, and

(2) the chain rule.

Check that your answers agree.

Activity 3 (solution)

(1) $z = \sin(t)\cos(t) = \dfrac{1}{2}\sin(2t)$, so $\dfrac{\mathrm{d}z}{\mathrm{d}t} = \cos(2t)$.

Alternatively,

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\big(\sin(t)\cos(t)\big) = \cos(t)\cos(t) + \sin(t)(-\sin(t)) = \cos(2t),$$

where we have used the product rule.

Do not forget your trig identities!

(2) By the chain rule,

$$\begin{aligned}\frac{\mathrm{d}z}{\mathrm{d}t} &= y\,\cos(t) + x\,(-\sin(t)) \\ &= \cos^2 t - \sin^2 t \\ &= \cos(2t).\end{aligned}$$

Multivariable chain rule – generalizations

The chain rule extends in an obvious way to functions of more variables. For example, if $w = f(x, y, z)$ where x, y, and z are each functions of t, then

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \frac{\partial w}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial w}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial w}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t}.$$

The chain rule extends to cases where x, y themselves are multivariable functions. For instance, if $z = f(x, y)$, where $x = g(u, v)$ and $y = h(u, v)$, then

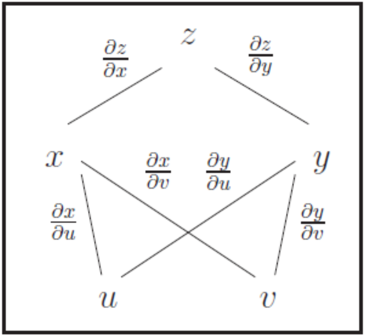
$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u},$$
$$\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v}.$$

Note: all the d's have changed to ∂'s. For multivariable functions, the derivatives no longer behave like fractions!

Chain Rule Diagram

There is a nice diagram you can use to remember the formula

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u},$$
$$\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v}.$$



Activity 4 (10 minutes)

Let $z = f(x,y)$, where

$$x = x(r,\theta) = r\cos(\theta), \quad y = y(r,\theta) = r\sin(\theta).$$

(So we are expressing z in polar coordinates.)

(1) Find the general expressions for $\frac{\partial z}{\partial r}$ and $\frac{\partial z}{\partial \theta}$ in terms of $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$, r and θ.

(2) Find $\frac{\partial z}{\partial r}$ and $\frac{\partial z}{\partial \theta}$ if $z = xy$. Check your answers.

(Note: this is different from Activity 3, since x and y are now functions of both r and θ.)

Activity 4 (solution)

(1) Using the bottom two formulas on Slide 18, we have

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial z}{\partial x}\cos(\theta) + \frac{\partial z}{\partial y}\sin(\theta),$$
$$\frac{\partial z}{\partial \theta} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \theta} = -\frac{\partial z}{\partial x}r\sin(\theta) + \frac{\partial z}{\partial y}r\cos(\theta).$$

(2) If $z = xy$ then $\frac{\partial z}{\partial x} = y$ and $\frac{\partial z}{\partial y} = x$. From part (1), we get:

$$\begin{aligned}\frac{\partial z}{\partial r} &= y\cos(\theta) + x\sin(\theta) = r\sin(\theta)\cos(\theta) + r\cos(\theta)\sin(\theta)\\ &= r\sin(2\theta),\\ \frac{\partial z}{\partial \theta} &= -yr\sin(\theta) + xr\cos(\theta) = -r\sin(\theta)r\sin(\theta) + r\cos(\theta)r\cos(\theta)\\ &= r^2\cos(2\theta).\end{aligned}$$

We can verify these calculations by writing $z = r^2\cos(\theta)\sin(\theta)$, and differentiating with respect to r and θ directly.

Math Modeling

- Defining the Problem Statement.
- Making Assumptions. Defining Variables.
- **Getting a Solution.**
- Analysis and Model Assessment.
- Reporting the Results.

Getting a Solution

Now that you have your problem definition, your list of assumptions and your variables, it is time to **get a solution!**

Resources: internet, your math knowledge from JC/poly/uni, or some software packages (Excel, Wolframalpha, Matlab, Python), etc.

No restriction and no requirement on the sophistication of math!

Often the simplest math modeling solutions are the most beautiful.

Getting a Solution: Getting Started

- Have I seen this type of problem before?
- If so, how did I solve it? If not, how is this problem different?
- Do I have a single unknown, or is this a multivariable problem with many interdependent variables?
- Is the model linear or nonlinear?
- Am I solving a system of equations simultaneously, or can I solve sequentially?
- What software or computational tools can I use in this case?
- Would a graph or visualization provide some insight?
- Could I approximate my complicated model with a simpler one?
- Can I hold some values constant and allow others to vary to see what is going on?

Getting a Solution

- How you build a solution may depend on what mathematical tools are available to you.
- There is often more than one way to tackle a problem, so just start and see what happens.
- If you don't immediately know how to solve the problem at hand, ask yourself the above set of questions to get you started.
- Different solution methods can lead to different solutions. This is OK!
- If the problem still looks difficult you can **make more assumptions** to make it more manageable! Think about the recycling math modeling VS math problem in cohort 1.1. Math problem was super easy to do. Come up with *justified* assumptions to make your **math modeling problem** look more like a **math problem**.

Getting a Solution: Example

Let's consider approaching a general problem by building different models that may lead to different results.

We want to determine which recycling method is best for Singapore.

We already know from our previous activities that this open-ended statement needs to be refined into a concise problem statement. The new problem re-statement is

Determine the cost of recycling plastic using 3 methods (a centralized drop-off location, a single stream and dual stream kerbside recycling) for four HDB estates: Tampines, Bedok, Clementi and Punggol.

Approach 1

Ranking.

Rank each of the 3 methods of recycling for each location from least costly to most costly (with justifications, e.g. from an internet research), possibly taking into account population density, income, demographics. For each method sum the ranks. The lowest sum denotes the least costly method.

This is a qualitative approach but it incorporates at least some quantitative analysis, since it requires us to roughly estimate the costs for each method.

Approach 2

Equation-Based Solution

We can compute net costs of each method by estimating the operational costs and subtracting the income generated by selling the recycled plastic. The cost for each method will consist of costs of labour, trucks (gasoline and maintenance), salaries, setting up and maintaining facilities etc.

We can obtain net costs for each HDB estate, as well as average cost for each method. This requires algebraic modeling and manipulation. It also requires extra assumptions on what are the main contributions to the cost (labour, maintenance of facilities, salaries of employees, revenue from selling recycled plastic etc), and what are the specific values.

Approach 3

Qualitative Comparison

We can decide if any of the three methods is the most expensive for all (or majority) of HDB estates, by taking into account population, the area of estate, demographics. For example, if one HDB estate has a lot of young families (we assume the younger generation recycles more than the older generation), with good transport accessibility, it may be less costly to set up a single drop-off centre. We eliminate the most expensive method, and then we use the same idea to eliminate the most expensive among the other two methods, eventually arriving at the cheapest one.

For you to submit

Start building your solution now with your teammates! Spend 20-30mins on it today. And another 30-40mins over the weekend.

You will submit your solution, Part 3 (together with Part 4 from cohort 2.2) on Piazza by 6pm, Monday, Feb 8th.

We do not require any particular math to be used.

It is confusing and not clear how to start, but give it your best try! Do not spend more than 1-2 hours working together on a problem. Keep a copy of your solution. You will need it later.

Remember, we are not grading the product, but the process! You will get full marks for a reasonable submission. We will review the submissions and provide the feedback.

This will train you for math modeling that you will be doing in 1D and 2D projects.

Summary

We have covered:

- Properties of the gradient vector
- General version of the chain rule for multivariable functions.
- Math Modeling, Part 3: Getting a Solution.

Textbook: read Sections 19.5 and 19.6, then try some of Exercises 19.5.1–19.5.10 and Exercises 19.6.1–19.6.19. You may discuss them on Piazza.