ECON4140 Mathematics 3 2019:

Practical information

Books: English OR Norwegian.

Minor curriculum change 2018: the Leibniz rule out (moved to Math2), more on differential equation systems in $\mathbb{R}^n$.

Some changes to the schedule to be discussed with you.

About this document

For 2018 I hand-wrote notes intended for "preview":

- Intended for "overview" before the lecture.
- Not for "lecture slideshow". Lectures were chalk-and-board.

I will hopefully return to that format soon.

Vector and matrix notation – on the board

This course will use vector and matrix notation.

- ...also in *transfomations*, i.e., "functions that can output something else than numbers". Main focus: $\vec{f}$ outputs a column m-vector. Such an $\vec{f}$ is just m "Math 2 functions" f_i stacked up in a column vector $\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(x_1,\dots,x_n) \\ \vdots \\ f_m(x_1,\dots,x_n) \end{pmatrix}$

Question: Handwriting type for the board? Vectors: Overarrow $\vec{x}$? Overbar? Undercurl? Blackboard bold? Merely $x \in \mathbb{R}^n$? Matrices: caps? Suggestion:

- Overarrow $\vec{x}$ for vectors: on the board, the above is $\vec{f}(\vec{x})$
- Overbar capitals $\overline{A}$ for matrices. Why not arrow?
  - Book sometimes uses $\mathbf{F}$ for vector of functions.
  - Sometimes I will refer to stats/probability applications. $X =$ random variable. Capital $\vec{X}$ for random vector, not matrix.
  - Overarrow notation from physics; rare for matrices.
  - OTOH: my 2018 notes use arrowheads. (You are free to add arrowheads if you like!)

Vector and matrix notation – on the board

Need to distinguish x_2 (coord. element no. 2 of $\vec{x}$) from "vector number two". Possible: $\mathbf{x}^{(2)}$ for the latter. $\vec{x}^{(2)}$ on the board. But:

- Although – see next slide – should I deviate in "linear algebra only" applications and use $\vec{a}_j$ and $\vec{r}_i^\top$?
- Will switch when we get to difference equations, where the book uses e.g. $\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t)$, wherein $\mathbf{x}_t$ is state at time t.
- Book often uses $\mathbf{x}^*$ for optimal point or candidate for optimal. I might use that even more often.

When to use the symbol $\mathbf{y}$? Often we think $y = f(\mathbf{x})$, but the book sometimes uses e.g. $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})$, and ... suggestion:

- I'll often use $\mathbf{u}$ and $\mathbf{v}$ for "generic vector input" ($\vec{u}$ and $\vec{v}$ on board). When book uses $\mathbf{y}$, I shall try to avoid using y for output. E.g., $z_\lambda = f(\lambda\vec{u} + (1-\lambda)\vec{v})$ avoids this.

Vectors default to columns! (But ...)

The following convention is not uncommon, and I will stick to it (caveats to be given)

- Vectors are *columns* unless otherwise stated or clear from context.
- To get a row: *transpose!* I will write $\vec{x}^\top$ for row vector.
  - Why not the book's prime symbol? Because we shall now use derivatives ...
  - There is a symbol: $\nabla f(\vec{x}) =$ the *row* vector of partial first derivatives. (Write $\vec{\nabla}$ if you like ...)
- Example: a matrix $\overline{A} = \left(\vec{a}^{(1)} \mid \ldots \mid \vec{a}^{(n)}\right)$ has columns $\{\vec{a}^{(j)}\}$.

  To write out its rows: $\overline{A} = \begin{pmatrix} \vec{r}^{(1)\,\top} \\ \vdots \\ \vec{r}^{(m)\,\top} \end{pmatrix}$ has rows $\{\vec{r}^{(i)\,\top}\}$
  - Would it be better to write $\vec{a}_j$ and $\vec{r}_i^\top$ here, despite ... ?
- In mixing linear algebra and calculus, we often write "det$(\overline{A})$" or "det $\overline{A}$" for determinant to keep them from absolute values.

Vectors default to columns! (But ...)

A “cheat” in the convention already: $f(x_1, \ldots, x_n)$ “looks like” $\vec{x}$ is a row (the variables are on the same line).

- Imagine that we really mean $f(\vec{x}^\top)$ but are too lazy to write it.
- Consistent with ∇f being row: variables counted left-to-right. Later: if $\vec{f}$ (column) is a transformation, we shall use one row for each ∇f_i. Inputs horizontally, outputs vertically.
- And, often the $^\top$ is omitted from first-order conditions $\nabla f(\vec{x}^*) = \vec{0}^\top$. I will try to remember it!

Using vector and matrix notation, we can formulate a Lagrange problem like $\max f(\vec{x})$ subject to $\vec{g}(\vec{x}) = \vec{b}$. Later today we shall see that the conditions can be written $\nabla f(\vec{x}^*) = \vec{\lambda}^\top \frac{\partial \vec{g}}{\partial \vec{x}}(\vec{x}^*)$ and the constrains.

Some simple (?) linear algebra not stressed in Mathematics 2

- The matrix product $\overline{A}\vec{x}$ can be written as $x_1\vec{a}^{(1)} + \ldots x_n\vec{a}^{(n)}$. And $\vec{u}^\top\overline{A} = u_1\vec{r}^{(1)\top} + \cdots + u_m\vec{r}^{(m)\top}$.
- $\vec{x}$ itself equals $x_1\vec{e}^{(1)} + \ldots x_n\vec{e}^{(n)}$, where $\vec{e}^{(j)}$ is the jth *standard unit vector:* 1 at element j, zeroes elsewhere.
  - IOW, $\vec{e}^{(j)}$ is the jth column of the identity matrix $\bar{I}$.
  - The collection of $\vec{e}^{(j)}$ vectors are called the "standard basis" for $\mathbb{R}^n$. We are not going to fiddle around with bases, but we shall sometimes use properties that a so-called *linear combination* $c_1\vec{v}^{(1)} + \ldots c_n\vec{v}^{(n)}$ can represent any point in $\mathbb{R}^n$.
- Q: You have a matrix $\overline{A}$. Pick a single element a_{ij}; is it possible to write that element in linear algebra terms?
  A: Yes. It is the form $\vec{e}^{(k)\top}\overline{A}\vec{e}^{(\ell)}$ ... but what is k and ℓ?
  Exercise: which is j and which is i?
- Quadratic functions $f(\vec{x}) = \vec{x}^\top\overline{A}\vec{x} + \vec{b}^\top\vec{x} + c$ [later]
- Matrices as linear transformations. [next slide]

Matrices as linear transformations

Mathematics 2 focuses on matrices as objects in their own right.
But let $\vec{f}(\vec{x}) = \overline{A}\mathbf{x}$ where $\overline{A}$ is $m \times n$; i.e., $\vec{f}$ outputs m-vectors.
This $\vec{f}$ is *linear:* it has the property that
$\vec{f}(\alpha\vec{u} + \beta\vec{v}) = \alpha\vec{f}(\vec{u}) + \beta\vec{f}(\vec{v})$. [you know "linear" vs "affine"?]

- Fact: each linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is of the form $\overline{A}\mathbf{x}$, some $\overline{A}$.
- Example: If $f(\vec{x})$ turns the n-vector upside down and returns $(x_n, x_{n-1}, \ldots, x_1)^\top$, then $\overline{A} = \begin{pmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & 1 & 0 \\ & & \vdots & \\ 0 & 1 & 0 & \ldots \\ 1 & 0 & \ldots & 0 \end{pmatrix}$

We want to differentiate transformations. And just like the function $g(t) = at$ has derivative $g'(t) = a$, the transformation $\overline{A}\vec{x}$ should have a derivative of $\overline{A}$.

- In particular, we want $\frac{\partial \vec{x}}{\partial \vec{x}} = \ldots$ what?

Derivatives

Already have: if f is a function (real function outputting real numbers, not vectors ...) then we can gather its first derivatives in a row vector $\nabla f(\vec{x})$.
Called the *gradient* of f. Differential: $\nabla f \, d\vec{x}$, i.e.:
First-order approximation: $f(\vec{x}) \approx f(\vec{x}^*) + \nabla f(\vec{x}^*)\ (\vec{x} - \vec{x}^*)$.

Questions:

- Transformations ...?
- Differentiation rules?
- Second derivatives ...?

Last first: If f is a real function, then the *Hessian matrix* $\overline{H} = \overline{H}(\vec{x})$ of f has elements $h_{ij} = f''_{ij}(\vec{x})$. [Second-order approx. $= \ldots$?]
For transformations, we have no such second-derivative *matrix.*

(Not to say that second derivatives do not exist – just that it needs more than matrices to formulate compactly. Not this course!)

Derivatives: The Jacobian of a transformation

Let $\vec{f} = (f_1, \ldots, f_m)^\top$ be a transformation. It so-called *Jacobian* is the matrix of first derivatives, so that row i is the gradient of f_i.

- Matrix of first derivatives – not the Hessian!
- I will use the notation $\frac{\partial \vec{f}}{\partial \vec{x}}$ to avoid the prime symbol.
  Or sometimes write "Jacobi[f]".
  You might find $\vec{f}'(\vec{x})$ in the literature, though I will avoid it.
- Differential: $\frac{\partial \vec{f}}{\partial \vec{x}}\, d\vec{x}$.
  First-order approximation: $\vec{f}(\vec{x}) \approx \vec{f}(\vec{x}^*) + \frac{\partial \vec{f}(\vec{x}^*)}{\partial \vec{x}}(\vec{x} - \vec{x}^*)s$.
- Confused over "element (3,2)" vs "element (2,3)"?
  - It is input horizontally and output vertically.
  - Easier to keep track of if $m \neq n$: the orders must match to make $\frac{\partial \vec{f}}{\partial \vec{x}} d\vec{x}$ meaningful.
- Examples: $\frac{\partial \vec{x}}{\partial \vec{x}} = \bar{I}$. And $\begin{pmatrix} x_1^2 + x_2^2 \\ x_1 x_2 + x_3 \end{pmatrix}$ has Jacobian $\begin{pmatrix} 2x_1 & 2x_2 & 0 \\ x_2 & x_1 & 1 \end{pmatrix}$.
- Example: The Jacobian of $\vec{f} := (\nabla g(\vec{x}))^\top$ is the Hessian of g.

Derivatives: Rules: linearity and the chain rule

Linearity: Just as $\frac{d}{dt}(\alpha f(t) + \beta g(t)) = \alpha\dot{f} + \beta\dot{g}$ (constant α, β):

- $\alpha\,\vec{f}(\vec{x}) + \beta\,\vec{g}(\vec{x})$ has Jacobian $= \alpha\,\frac{\partial\vec{f}}{\partial\vec{x}} + \beta\,\frac{\partial\vec{g}}{\partial\vec{x}}$
- $\overline{A}\,\vec{f}(\vec{x}) + \overline{B}\,\vec{g}(\vec{x})$ has Jacobian $= \overline{A}\,\frac{\partial\vec{f}}{\partial\vec{x}} + \overline{B}\,\frac{\partial\vec{g}}{\partial\vec{x}}$
  (for constant matrices $\overline{A}$, $\overline{B}$ – incl. row vectors $\vec{a}^\top$, $\vec{b}^\top$.)
- First-order approximation: $\overline{A}\,\vec{f}(\vec{x}) + \overline{B}\,\vec{g}(\vec{x})$
  $$\approx \overline{A}\,\vec{f}(\vec{x}^*) + \overline{B}\,\vec{g}(\vec{x}^*) + \left(\overline{A}\,\frac{\partial\vec{f}(\vec{x}^*)}{\partial\vec{x}} + \overline{B}\,\frac{\partial\vec{g}(\vec{x}^*)}{\partial\vec{x}}\right)(\vec{x} - \vec{x}^*)$$

The chain rule (nice!): $\vec{z}(\vec{x}) = \vec{f}(\vec{g}(\vec{x}))$ has Jacobian

$$\frac{\partial\vec{z}}{\partial\vec{x}} = \frac{\partial\vec{f}}{\partial\vec{y}}\,\frac{\partial\vec{g}}{\partial\vec{x}}\Big|_{\vec{y}=\vec{g}(\vec{x})}$$

(note: $k \times m$ by $m \times n$ if $\vec{y} \in \mathbb{R}^m$)

- If $\vec{f}$ is linear, it is multiplication by some matrix, cf. above.
- Example: If $\vec{g}$ is linear, $\vec{z} = \vec{f}(\overline{A}\vec{x})$: $\quad \frac{\partial(\vec{f}(\overline{A}\vec{x}))}{\partial\vec{x}} = \frac{\partial\vec{f}}{\partial\vec{y}}\Big|_{\vec{y}=\overline{A}\vec{x}}\overline{A}$
  - If furthermore $k = 1$ so $\vec{f} = f$, we can use this to get a Hessian. $\nabla z = \nabla f|_{\vec{y}=\overline{A}\vec{x}}\overline{A}$ and $\text{Jacobi}[(\nabla z)^\top]^\top = (\text{Hesse}[f]\big|_{\vec{y}=\overline{A}\vec{x}})^\top\overline{A}$, so $\text{Hesse}[z] = \overline{A}^\top\text{Hesse}[f]\big|_{\vec{y}=\overline{A}\vec{x}}\overline{A}$ (as Hessians are symmetric).

Derivatives: product rules

Product rules could require some thinking:

- The real function $f(\vec{x})g(\vec{x})$ has gradient $g\,\nabla f + f\,\nabla g$.
- Thus if $\vec{f}$, $\vec{g}$ are transformations, the real function $\vec{f}^\top \vec{g} = \sum_i f_i g_i$ has gradient $\sum_i [g_i\,\nabla f_i + f_i\,\nabla g_i]$, which equals (the row vector) $\quad \vec{f}^\top \frac{\partial \vec{g}}{\partial \vec{x}} + \vec{g}^\top \frac{\partial \vec{f}}{\partial \vec{x}}$.
  - Example (also, previous slide): $\vec{a}^\top \vec{x}$ has gradient $\vec{a}^\top$.
  - Example: $\vec{x}^\top \overline{A}\vec{x}$. Let $\vec{f} = \vec{x}$ and $\vec{g} = \overline{A}\vec{x}$ (Jacobians $\bar{I}$ and $\overline{A}$. We get gradient $\vec{x}^\top \overline{A} + (\overline{A}\vec{x})^\top \bar{I} = \vec{x}^\top (\overline{A} + \overline{A}^\top)$.

    Note that $\vec{x}^\top \overline{A}\vec{x} = \vec{x}^\top \overline{A}^\top \vec{x} = \frac{1}{2}\vec{x}^\top (\overline{A} + \overline{A}^\top)\vec{x}$. More later!
- If f is a real function and $\vec{g}$ is a transformation, then the transformation $f\vec{g}$ has a Jacobian whose row i is $g_i\,\nabla f + f\,\nabla g_i$. That is, $\frac{\partial [f\vec{g}]}{\partial x} = \vec{g}\,\nabla f + f\frac{\partial \vec{g}}{\partial x}$.
  Note the orders of $\vec{g}\,\nabla f$: it is $(m \times 1)$ by $(1 \times n)$.

"Combining rules" example: $\nabla ||\vec{x}|| = \nabla \sqrt{\vec{x}^\top \vec{x}} = \frac{1}{2\vec{x}^\top \vec{x}} \vec{x}(\bar{I} + \bar{I}^\top) = \ldots$

... so $||\vec{x}||$ has Hessian = Jacobi$[\frac{1}{||\vec{x}||} x]$ = [not trivial exercise!]

Derivatives: cases too ugly for ordinary linear algebra

Some things are not so nice. Just to mention:

- There is no third derivative matrix for a function of more than 1 variable. It would have to be a 3D cubic array.
- If f is a real function, then the Hessian of $f(\vec{g}(\vec{x}))$ cannot be written as a matrix product except special cases: the second derivative of $\vec{g}$ would be a 3D cubic array.
- The transformation $\overline{F}(x)\vec{g}(x)$, has a Jacobian whose row i equals the ("Math 3 calculatable") gradient of $\vec{r}^{(i)\,T}\vec{g}$. (That's row i of $\overline{F}$.) We cannot outright write this as a matrix product. But the differential can be written $(d\overline{F})g + \overline{F}d\vec{g}$. Of course, if $\overline{F}$ constant, then $d\overline{F} = \overline{0}$ and we are back to linearity.

But on the other hand, if there is only one variable: a matrix $\overline{M}(t)$ is differentiated element-wise as far as each element is differentiable. $\dot{\overline{M}}(t)$, $\ddot{\overline{M}}(t)$ or even $\dddot{\overline{M}}(t)$? No prob!

Continuity and differentiability – some technicalities that will be glossed over, and some issues we will cover

Continuity of in more than one variable, is quite a delicate matter.

- Let $f(x, y) = 1$ when $x^2 < y < 2x^2$, and 0 elsewhere. f is certainly not continuous at $(0, 0)$, but both $f'_x(0, 0)$ and $f'_y(0, 0)$ exist and are 0. So: existence of partial derivatives does not even grant continuity of the function!
- However, if the partial derivatives are continuous *on some neighbourhood* – not merely at the point – we are in business.
- By multivariate "differentiability" we actually mean that the first-order approximation "is good". I'll skip the details.

More non-differentiabilities in this course than in Mathematics 2!

- Shall cover: non-differentiable concave/convex functions
- Will encounter optimal control problems where you jump straight from "max saving, no consumption" to other corner.

Curriculum! Implicit derivatives; the implicit function theorem

Recall differentiation in equation systems from Mathematics 2. Let $\vec{f} = \vec{f}(\vec{x}, \vec{u})$ take as input $n + m$ variables $\vec{x} \in \mathbb{R}^n$ and $\vec{u} \in \mathbb{R}^m$, and output m-vectors.

- We have written f as function of two vectors. Notation: let $\dfrac{\partial \vec{f}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ & \vdots & & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$ and similar for $\dfrac{\partial \vec{f}}{\partial \vec{u}}$.
  - I.e.: $\frac{\partial \vec{f}}{\partial \vec{x}}$ "as if $\vec{u}$ were constant" and vice versa.
- Formal differentiation yields $\vec{0} = \frac{\partial \vec{f}}{\partial \vec{x}}\, d\vec{x} + \frac{\partial \vec{f}}{\partial \vec{u}}\, d\vec{u}$ and so $\frac{\partial \vec{u}}{\partial \vec{x}} = -\left(\frac{\partial \vec{f}}{\partial \vec{u}}\right)^{-1} \frac{\partial \vec{f}}{\partial \vec{x}}$ for the transformation $\vec{u} = \vec{u}(\vec{x})$.
- Valid? Fact: Pick a point $P:\ (\vec{x}^*, \vec{u}^*)$ with $\vec{f}(\vec{x}^*, \vec{u}^*) = \vec{c}$. Assume that on some neighbourhood of P, we have $\vec{f}$ continuously differentiable (i.e., in each coordinate) with $\frac{\partial f}{\partial u}$ invertible at P. Then there exists indeed a transformation $\vec{u} = \vec{u}(\vec{x})$ with that Jacobian.