In class exercises for CW 43

The idea of matrix calculus is to write down the partial derivatives of multivariate/multivalued functions into a matrix grid. This notation gives rise to a powerful chain rule which simplifies many calculations. However there is a caveat: two competing systems are in wide spread use. For a vector to vector function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ some people prefer to work the Jacobian, others with the gradient.

Numerator layout	$\left(\frac{\partial f}{\partial x}\right)_{ij} = \frac{\partial f_i}{\partial x_j}$	"Jacobian way"
Denominator layout	$\left(\frac{\partial f}{\partial x}\right)_{ij} = \frac{\partial f_j}{\partial x_i}$	"Gradient way"

For the purpose of this exercise we will stick to the denominator layout

The base definition can be extended to the case of matrix-valued functions of one variable and matrix dependent scalar functions:

Definition 1. Consider functions $f\colon \mathbb{R}^n \to \mathbb{R}^m, x \mapsto f(x)$ (vector to vector), $Y\colon \mathbb{R} \to \mathbb{R}^{m\times n}, t \mapsto Y(t)$ (scalar to matrix) and $\phi\colon \mathbb{R}^{m\times n} \to \mathbb{R}, X \mapsto \phi(X)$ (matrix to scalar). The denominator convention consists of the following convections:

$$\boxed{\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}} \quad \boxed{\frac{\partial Y}{\partial t} = \begin{pmatrix} \frac{\partial Y_{11}}{\partial t} & \cdots & \frac{\partial Y_{1n}}{\partial t} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_{m1}}{\partial t} & \cdots & \frac{\partial Y_{mn}}{\partial t} \end{pmatrix}} \quad \boxed{\frac{\partial \phi}{\partial X} = \begin{pmatrix} \frac{\partial \phi}{\partial X_{11}} & \cdots & \frac{\partial \phi}{\partial X_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi}{\partial X_{m1}} & \cdots & \frac{\partial \phi}{\partial X_{mn}} \end{pmatrix}}$$

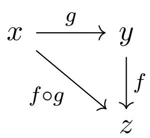
Remark 2. Using matrix calculus we can write down the following derivatives in concise form:

$x \setminus f$	scalar	vector	matrix
scalar	✓	✓	✓
vector	✓	✓	×
matrix	✓	×	×

The missing ones would require to write the entries not in a matrix, but in higher dimensional analogues (tensors).

Theorem 3 (Chain Rule for denominator layout)**.**

$$\boxed{\frac{\partial (f\circ g)}{\partial x} = \frac{\partial g}{\partial x}\cdot\frac{\partial (f\circ g)}{\partial g}} \tag{1}$$

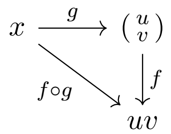


Note: In the case 'scalar by matrix' the order is swapped: $\frac{\partial(f\circ g)}{\partial X} = \frac{\partial(f\circ g)}{\partial g}\cdot\frac{\partial g}{\partial X}$. *If numerator layout is used instead, the order is* ***always*** $\frac{\partial(f\circ g)}{\partial x} = \frac{\partial(f\circ g)}{\partial g}\cdot\frac{\partial g}{\partial x}$. [1]

Exercise 4. Assume $g\colon \mathbb{R}^n \to \mathbb{R}^k$ and $f\colon \mathbb{R}^k \to \mathbb{R}^m$ what are dims. of $\frac{\partial(f\circ g)}{\partial x}$, $\frac{\partial g}{\partial x}$ and $\frac{\partial f\circ g}{\partial g}$?

The chain rule is in some sense the most important differentiation rule, since all other differentiation rules can be derived from it.

Example 5. The one dimensional product rule is $\frac{\partial}{\partial x}(uv) = u'v + v'u$. Proof:



$$\frac{\partial(f\circ g)}{\partial x} = \frac{\partial g}{\partial x}\cdot\frac{\partial(f\circ g)}{\partial g} = \begin{pmatrix}u' & v'\end{pmatrix}\cdot\begin{pmatrix}v\\ u\end{pmatrix} = u'v + v'u$$

Exercise 6 (scalar by vector)**.** Let $a, x \in \mathbb{R}^n$ and $u, v : \mathbb{R}^n \to \mathbb{R}^m$ be functions of x

- $\frac{\partial}{\partial x}(a^\intercal x)$ and $\frac{\partial}{\partial x}(x^\intercal a)$ (sol: a and a)
- Show that $\frac{\partial}{\partial x}(u^\intercal v) = \frac{\partial u}{\partial x}v + \frac{\partial v}{\partial x}u$ (scalar product rule)
- Compute $\frac{\partial}{\partial x}\|x\|_2^2$ (sol: $2x$)

Exercise 7 (vector by vector)**.**

- $\frac{\partial}{\partial x}(Ax)$ and $\frac{\partial}{\partial y}(y^\intercal A)$ (sol: $A^\intercal$ and sol: A)
- $\frac{\partial}{\partial x}\frac{1}{2}\|y - Ax\|_2^2$ (sol: $A^\intercal Ax - A^\intercal y$)

Exercise 8 (scalar by matrix)**.** Let $x \in \mathbb{R}^n, y \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n}$ and $\phi\colon \mathbb{R}^m \to \mathbb{R}$

- Compute $\frac{\partial}{\partial A}y^\intercal Ax$ (sol: $yx^\intercal$)
- Compute $\frac{\partial}{\partial A}\phi(Ax)$ (sol: $\nabla\phi[Ax]\cdot x^\intercal$) (Hint: This exercise goes beyond what matrix calculus is able to handle. After applying the chain rule the derivative $\frac{\partial Ax}{\partial A}$ arises which is of type vector-by-matrix. Fundamentally it would need to represented by a 3-dimensional tensor. We can bypass this issue by writing $Ax = \sum_i (e_i^\intercal Ax)e_i$ and form the derivative component wise.)

[1] This can be seen as a fundamental flaw of the denominator convention. To fix it one would have to redefine $(\frac{\partial f}{\partial X})_{ij} = \frac{\partial f}{\partial X_{ji}}$, which is counter-intuitive.

Literature Recommendation

Basic multivariate calculus can be found in almost any appropriate undergraduate text. Two on-line sources that summarize vectorized multivariate calculus and can be used to look up formulas are:

- Wikipedia page on Matrix Calculus [2]
- Matrix Cookbook [3]

[2] https://en.wikipedia.org/wiki/Matrix_calculus
[3] https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf