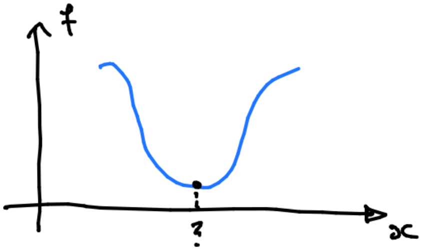
Vector and matrix derivatives

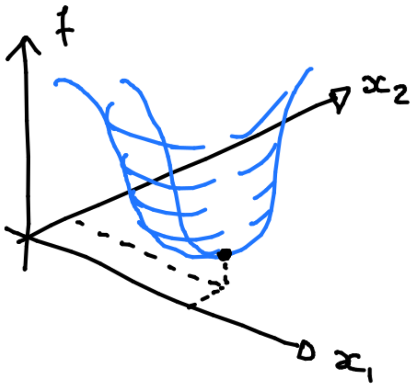
Herman Kamper

Main idea

How do we find the minimum of a scalar function?



And for a function of two variables?



Think about it:

- What if we have a function of N variables?
- Or a function with intermediate variables?
- Or a function that produces a vector as output, instead of a scalar as in the above two examples?

Main idea: Define vector and matrix derivatives to allow us to differentiate directly in a vector or matrix form. From the definitions, we obtain general rules and identities, which are very similar to those for the scalar case.

Vector and matrix derivative definitions

We simply define what vector and matrix derivatives are, and then, by following these definitions, we can obtain general identities which we can easily apply.[1]

The derivative of a scalar function $f : \mathbb{R}^N \to \mathbb{R}$ with respect to a vector $\mathbf{x} \in \mathbb{R}^N$:

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

The derivative of a vector function $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^M$ with respect to a vector $\mathbf{x} \in \mathbb{R}^N$:

$$\frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_1} \\ \frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_N} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_1} \\ \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_N} & \frac{\partial f_2(\mathbf{x})}{\partial x_N} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

where $\boldsymbol{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) & f_2(\mathbf{x}) & \cdots & f_M(\mathbf{x}) \end{bmatrix}^\top$. Here we implicitly use another definition:

$$\frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_1} \triangleq \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_1} \end{bmatrix}$$

[1] Or, more accuractely, someone on Wikipedia can do the work to get these identies for us, and as long us we understand the definitions, we can use them.

The derivative of a scalar function $f : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}$ with respect to a matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$:

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial X_{1,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{1,2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{1,N}} \\ \frac{\partial f(\mathbf{X})}{\partial X_{2,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{2,2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{2,N}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial X_{M,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{M,2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{M,N}} \end{bmatrix}$$

Using the above definitions, we can generalise the chain rule. Given $\mathbf{u} = \boldsymbol{h}(\mathbf{x})$ (i.e. $\mathbf{u}$ is a function of $\mathbf{x}$) and $\boldsymbol{g}$ is a vector function of $\mathbf{u}$, the vector-by-vector chain rule states:

$$\frac{\partial \boldsymbol{g}(\mathbf{u})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \boldsymbol{g}(\mathbf{u})}{\partial \mathbf{u}}$$

Common identities

Here's a short list of identities that I have found useful in the past. The vector $\mathbf{a}$ and matrix $\mathbf{A}$ are constants.

$$\frac{\partial(\boldsymbol{u}(\mathbf{x}) + \boldsymbol{v}(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial \boldsymbol{u}(\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \boldsymbol{v}(\mathbf{x})}{\partial \mathbf{x}}$$

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \text{ if } \mathbf{A} \text{ is symmetric}$$

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|(\mathbf{X}^{-1})^\top$$

$$\frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^\top$$

Example derivation

What is $\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}}$ with $\mathbf{a}$ a constant N-dimensional column vector?

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_{n=1}^{N} x_n a_n = a_i$$

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial x_1} \\ \frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial x_N} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \mathbf{a}$$

And so we get the identity

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}}$$

Where to find identities

I should first say something about different conventions. The definitions given above follows what is called the *denominator layout*. There is also a different layout called the *numerator layout*. This alternative layout also results in a set of easily usable identities (which are often transposed versions of the identities given here). I do not want you to worry about this too much: You should just know that people sometimes use different layouts, and that I will stick with the denominator layout in these notes. So if you look up identities on Wikipedia, make sure to look under the denominator layout column.

The following references are very useful for learning more about vector and matrix derivatives, and for finding more identities:

- http://en.wikipedia.org/wiki/Matrix_calculus
- https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf
- http://www.kamperh.com/notes/kamper_matrixcalculus13.pdf

Videos covered in this note

- Vector and matrix derivatives (13 min)