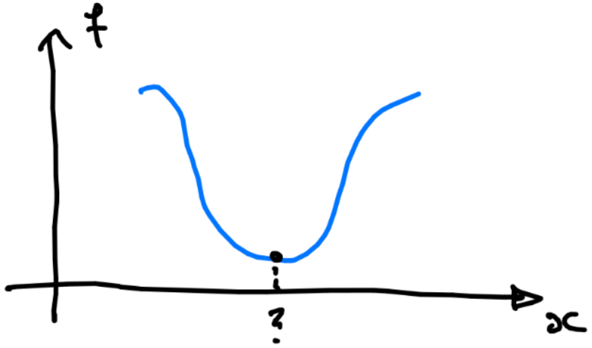
Vector and matrix derivatives

Herman Kamper

http://www.kamperh.com/

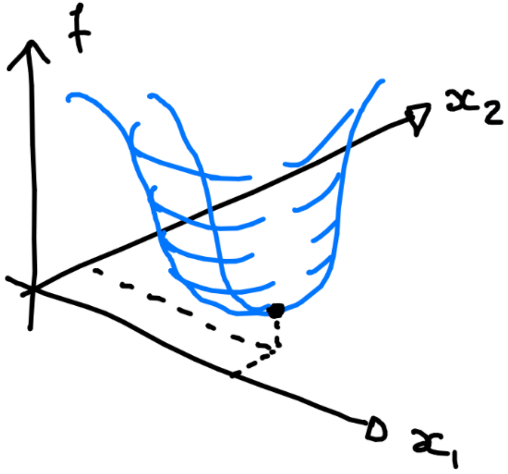
Main idea

How do we find minimum of a scalar function?



Set $\dfrac{df}{dx} = 0$.

And for a function of two variables?



Set $\dfrac{\partial f}{\partial x_1} = 0$

set $\dfrac{\partial f}{\partial x_2} = 0$

$\dfrac{\partial f}{\partial \underline{x}} = ?$

- What if we have a function with N variables?
- Functions with intermediate variables?
- Functions producing a vector as output instead of a scalar?

Main idea:

Define vector and matrix derivatives to allow us to differentiate directly in vector/matrix form.

Definitions

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

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

$$\frac{\partial \underline{f}(\underline{x})}{\partial x_1} = \left[\frac{\partial f_1(\underline{x})}{\partial x_1} \cdots \frac{\partial f_M(\underline{x})}{\partial x_1} \right]$$

$$\underline{f}(\underline{x}) = \left[f_1(\underline{x}) \; f_2(\underline{x}) \cdots f_M(\underline{x}) \right]^T$$

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

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

Definitions

- Derivative of a scalar function $f : \mathbb{R}^{M\times N} \to \mathbb{R}$ with respect to 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:

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

Order matters!

Common identities:

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

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

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

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

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

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

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

Example derivation:

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

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

$$\frac{\partial \underline{x}^T \underline{a}}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial \underline{x}^T \underline{a}}{\partial x_1} \\ \vdots \\ \frac{\partial \underline{x}^T \underline{a}}{\partial x_N} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \underline{a}$$

$$\therefore \boxed{\frac{\partial \underline{x}^T \underline{a}}{\partial \underline{x}} = \underline{a}}$$

Where to find identities

Denominator layout

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