Basic Mathematics, Fall 2020

Karen Keryan,
ASDS, YSU

November 3, 2020

Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a vector valued function. Then for a vector $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T \in \mathbb{R}^n$, the value of the function $\mathbf{f}$ at point $\mathbf{x}$ is a *vector* given as

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^m$$

Here the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are real-valued functions.

Therefore, the **partial derivative of a vector-valued function** $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ with respect to $x_i \in \mathbb{R}, i = 1, \dots, n$, is given as the vector

$$\frac{\partial \mathbf{f}}{\partial x_i} = \begin{bmatrix} \frac{\partial f_1}{\partial x_i} \\ \frac{\partial f_2}{\partial x_i} \\ \vdots \\ \frac{\partial f_m}{\partial x_i} \end{bmatrix} \in \mathbb{R}^m$$

Hence the **gradient of the vector-valued function** $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is

$$\nabla \mathbf{f} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \dots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Note that gradient of the vector-valued function can also be represented as follows

$$\nabla \mathbf{f} = \begin{bmatrix} \nabla f_1 \\ \vdots \\ \nabla f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Definition

The **Jacobian matrix** *is the matrix of all first-order partial derivatives of a vector-valued function* $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. *The Jacobian matrix* J *of* $\mathbf{f}$ *is an* $m \times n$ *matrix, usually defined and arranged as follows:*

$$\mathbf{J} = \nabla \mathbf{f} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}, \quad \mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}.$$

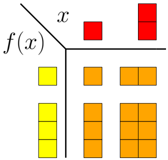


Figure: The dimension of the Jacobian $\mathbf{J_f}$

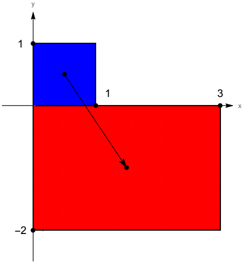
In particular, the Jacobian of a function $f : \mathbb{R}^n \to \mathbb{R}^1$, which maps a vector $\mathbf{x} \in \mathbb{R}^n$ onto a scalar, is a row vector (matrix of dimension $1 \times n$).

Example

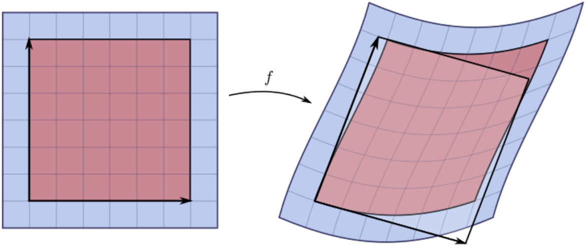
Find the Jacobian of the following functions:

a) $f : \mathbb{R}^2 \to \mathbb{R}$, *given by* $f(\mathbf{x}) = x_1 + x_2^3$

b) $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$, *given by* $f(\mathbf{x}) = \begin{bmatrix} 2x_1 \\ x_1 x_2 \\ x_1 + 3x_2 \end{bmatrix}$



Consider the vector valued function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\mathbf{f}(\mathbf{x}) = \mathbf{f}(x, y) = \begin{bmatrix} 3x \\ -2y \end{bmatrix}$. Note that the image of the $[0, 1]^2$ (the blue square) is the rectangle $[0, 3] \times [-2, 0]$ (depicted in red). The quotient of the areas of the rectangle and the square is 6 and it is equal to $|\det \mathbf{J_f}| = \left| \det \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \right|$



A nonlinear map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ sends a small square (left, in red) to a distorted parallelogram (right, in red). The Jacobian at a point gives the best linear approximation of the distorted parallelogram near that point (right, in translucent white), and the Jacobian determinant gives the ratio of the area of the approximating parallelogram to that of the original square.

Example

Find the gradient of the vector-valued function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, *given by* $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Example

Prove that the Jacobian of the vector-valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, *given by* $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, *where* $A \in \mathbb{R}^{m \times n}$ *and* $\mathbf{x} \in \mathbb{R}^n$ *is*

Example

Find the gradient of the vector-valued function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, *given by* $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Example

Prove that the Jacobian of the vector-valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, *given by* $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, *where* $A \in \mathbb{R}^{m \times n}$ *and* $\mathbf{x} \in \mathbb{R}^n$ *is the matrix* A, *i.e.*

$$\mathbf{J_f} = A.$$

Chain Rule (Matrix form)

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^k$ be differentiable functions, then

$$\mathbf{J}_{\mathbf{f}\circ\mathbf{g}}(\mathbf{a}) = \mathbf{J}_{\mathbf{f}}(\mathbf{g}(\mathbf{a}))\mathbf{J}_{\mathbf{g}}(\mathbf{a}), \quad \mathbf{a} \in \mathbb{R}^n.$$

Note that $\mathbf{J}_{\mathbf{f}\circ\mathbf{g}}(\mathbf{a}) \in \mathbb{R}^{k\times n}$, $\mathbf{J}_{\mathbf{f}}(\mathbf{g}(\mathbf{a})) \in \mathbb{R}^{k\times m}$, and $\mathbf{J}_{\mathbf{g}}(\mathbf{a}) \in \mathbb{R}^{m\times n}$.

Equivalently, if $\mathbf{z} = \mathbf{f}(\mathbf{y})$ and $\mathbf{y} = \mathbf{g}(\mathbf{x})$ then

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{z}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{x}}.$$

Example

Find the Jacobians of the functions $f \circ \mathbf{g}$ *and* $\mathbf{g} \circ f$, *where* $f : \mathbb{R}^2 \to \mathbb{R}$, *given by* $f(\mathbf{x}) = x_1 + x_2^2$ *and* $\mathbf{g} : \mathbb{R} \to \mathbb{R}^2$ *given by* $\mathbf{g}(t) = \begin{bmatrix} e^t \\ t^2 \end{bmatrix}$.

Example

Using the formula $\frac{\partial \mathbf{x}^T B \mathbf{x}}{\partial \mathbf{x}} = \mathbf{x}^T(B + B^T)$ *deduce that* $\frac{\partial \mathbf{x}^T \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}^T$.

Example

(Gradient of a Least-Squared Loss in a Linear Model)
Let $Y = X\beta + \varepsilon$, *where* $X \in \mathbb{R}^{N \times D}$ *is the input,* $\beta \in \mathbb{R}^D$ *is a parameter vector, and* ε *is the error. Define*

$$L(\varepsilon) = \|\varepsilon\|^2,$$

which is called a least-squares loss function. Using the result of previous problem find the derivative $\frac{\partial L}{\partial \beta}$.