B4 - Gradient, Hessian and Jacobian*

Andrea Brose

14th of February 2005

First of all note, that there is an **error** in the textbook's **example B.2** on page 650: the second to last matrix is the Jacobian, and the last matrix on the page is the Jacobian's transpose, i.e. they need to be swapped.

(A) The del Operator $\boldsymbol{\nabla}$

Think of $\boldsymbol{\nabla}$ being a vector, whose components are the partial derivatives with respect to the n variables, and which is applied to functions, i.e. an operator living on the space of 'certain' function. In particular

$$\boldsymbol{\nabla} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix}$$

(B) The Gradient and Hessian

Let

$$\begin{aligned} f\ :\ U \subset \mathbb{R}^n &\rightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto f(x_1, \dots, x_n) \end{aligned}$$

then for $\boldsymbol{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T$

(i)

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

is the gradient vector of f. Note that the operation $\boldsymbol{\nabla} f$ is like multiplying the vector $\begin{bmatrix} \frac{\partial}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} \end{bmatrix}^T$ by the scalar f. We say that the del operator $\boldsymbol{\nabla}$ operates on f.

*Please let me know if you find any errors or typos.

(ii)

$$\begin{aligned}
\boldsymbol{\nabla}^2 f(\boldsymbol{x}) &= \begin{bmatrix} \frac{\partial f^2}{\partial x_1 \partial x_1} & \cdots & \frac{\partial f^2}{\partial x_n \partial x_1} \\ & \ddots & \\ \frac{\partial f^2}{\partial x_1 \partial x_n} & \cdots & \frac{\partial f^2}{\partial x_n \partial x_n} \end{bmatrix} \\
&= \left[\boldsymbol{\nabla}^2 f(\boldsymbol{x})\right]_{i,j} \\
&= \left[\frac{\partial f^2}{\partial x_i \partial x_j}\right]_{i,j}
\end{aligned}$$

Also remember the following theorem: If $f \in C^2$, then $\boldsymbol{\nabla}^2 f(\boldsymbol{x})$ is symmetric (why?).

(C) The Jacobian

If

$$\begin{aligned}
\boldsymbol{f} : U \subset \mathbb{R}^n &\to \mathbb{R}^m \\
(x_1, \dots, x_n) &\mapsto \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix}
\end{aligned}$$

so that each $f_i : U \subset \mathbb{R}^n \to \mathbb{R}$ is as in (B), then

$$\begin{aligned}
\boldsymbol{\nabla} \boldsymbol{f}(\boldsymbol{x}) &= \begin{bmatrix} \boldsymbol{\nabla} f_1(\boldsymbol{x}) & \dots & \boldsymbol{\nabla} f_m(\boldsymbol{x}) \end{bmatrix} \\
&= \begin{bmatrix} \partial_{x_1} f_1 & \dots & \partial_{x_1} f_m \\ & \ddots & \\ \partial_{x_n} f_1 & \dots & \partial_{x_n} f_m \end{bmatrix}
\end{aligned}$$

is an $n \times m$ matrix of partial derivatives. Note that here the operation $\boldsymbol{\nabla} \boldsymbol{f}$ is like applying the del operator to each component f_i of $\boldsymbol{f}$.

$\boldsymbol{\nabla}^T \boldsymbol{f}$ is the called Jacobian of $\boldsymbol{f}$ and is often denoted by $J(\boldsymbol{f})$. It is the derivative matrix of $\boldsymbol{f}$. The i-th row of $\boldsymbol{\nabla}^T \boldsymbol{f}$ is the transpose of the gradient of f_i, i.e.:

$$\boldsymbol{\nabla}^T \boldsymbol{f} = \begin{bmatrix} \boldsymbol{\nabla}^T f_1 \\ \vdots \\ \boldsymbol{\nabla}^T f_m \end{bmatrix}$$

For examples, please see various calculus books, as well as the appendix of the textbook we use. Please familiarize yourself with the the various 'derivatives' so that we can use them without thinking too much, what they are.