Note on Gradients and Jacobians

Sergey

December 13, 2016

1 Introduction

The gradient is a generalization of derivative for scalar valued functions of several variables. I.e. consider $g(x) : \mathbb{R}^n \to \mathbb{R}$. The vector gradient $\nabla g(x) : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the following limit definition:

$$\lim_{d \to 0} \frac{\|g(x+d) - g(x) - \nabla g(x) \cdot d\|}{\|d\|} = 0$$

Another way to write this is to say that the *gradient* of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$, if exists, is the unique vector v that satisfies

$$g(x+d) = g(x) + v^T d + o(\|d\|), \quad \forall\, d \in \mathbb{R}^n.$$

As an example, consider for example, the function $g(x) = x^T M x$ where M is a symmetric matrix. It follows that:

$$g(x+d) = (x+d)^T M(x+d) = x^T Mx + 2x^T Md + d^T Md = g(x) + 2x^T Md + d^T Md$$

From the term $2x^T Md$ we deduce that $v^T = 2x^T M \implies v = 2M^T x = 2Mx$ is the gradient of g. Notice that $\frac{d^T Md}{\|d\|} \to 0$ as $d \to 0$ (see HW 1 note for details).

Next, for vector functions which map $\mathbb{R}^n$ to spaces $\mathbb{R}^m$, the generalization of the derivative is the Jacobian. The simplest example is $f(x) = Ax$ for a general $m \times n$ matrix A. In this case, $f : \mathbb{R}^n \to \mathbb{R}^m$. The Jacobian of f is defined as:

$$J[f(x)] = J = \left[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ & \vdots & \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} [\nabla f_1(x)]^T \\ \vdots \\ [\nabla f_m(x)]^T \end{bmatrix}$$

Notice that for our example, $f_i = (Ax)_i = \sum_{k=1}^n A_{ik} x_k$. It follows that $\frac{\partial f_i}{\partial x_j} = A_{ij}$ and thus, $J[f(x)] = A$.

One connection between the gradient and Jacobian occurs in the so called chain rule. Recall that for single variable functions mapping to $\mathbb{R}$, we have that $\frac{d}{dx} f(g(x)) = g'(x) f'(g(x))$. Notice that when we have the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^l \to \mathbb{R}^n$, then the analogue of the derivative for g is the Jacobian, while the analogue of the derivative for f is the gradient (since it maps to $\mathbb{R}$). Thus one may write:

$$\nabla f(g(x)) = J[g(x)]^T \left[\nabla f(y)\right]|_{y=g(x)}$$

This is of course just an analogue to the single variable case without proof. Suppose $f : \mathbb{R}^m \to \mathbb{R} : y \mapsto y_1^2 + \cdots + y_m^2$ and $g : \mathbb{R}^n \to \mathbb{R}^m : x \mapsto Ax - b$ for $A \in \mathbb{R}^{m \times n}$. Then $f(g(x)) = \|Ax - b\|_2^2$. Now, $\nabla f(y) = 2y$, since each partial derivative is $2y_k$. Also, the Jacobian of $Ax - b$ is A, following the logic above. Thus, we get:

$$\nabla f(g(x)) = J[g(x)]^T \left[\nabla f(y)\right]|_{y=g(x)} = 2A^T(Ax - b).$$

Note that the same result may be derived from the limit definition above, using a Taylor series expansion.