Linear Approximations to Vector Functions; the Jacobian Matrix

Considerations similar to those developed previously for scalar functions $\phi(X)$ apply also to vector functions $F : R^n \rightarrow R^m$, typically written in the form

$$Y = F(X), \; X \in R^n, \; Y \in R^m.$$

Recalling that what we actually have here is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, ..., x_n) \\ f_2(x_1, x_2, ..., x_n) \\ \vdots \\ f_m(x_1, x_2, ..., x_n) \end{pmatrix},$$

and assuming $F(X)$, and hence each of the $f_i(X)$, is continuously differentiable in some region $\mathcal{D}$ including a base point $X_0 = (x_{01}, x_{02}, ..., x_{0n})$, then for $i = 1, 2, ..., m$ we have, for X near X_0,

$$f_i(X) = f_i(X_0) + \nabla f_i(X_0)(X - X_0) + o(\|X - X_0\|).$$

Each of the gradient (row) vectors $\nabla f_i(X_0)$ can be made the i-th row of an $m \times n$ matrix

$$\nabla F(X_0) \equiv \begin{pmatrix} \nabla f_1(X_0) \\ \nabla f_2(X_0) \\ \vdots \\ \nabla f_m(X_0) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(X_0) & \frac{\partial f_1}{\partial x_2}(X_0) & \cdots & \frac{\partial f_1}{\partial x_n}(X_0) \\ \frac{\partial f_2}{\partial x_1}(X_0) & \frac{\partial f_2}{\partial x_2}(X_0) & \cdots & \frac{\partial f_2}{\partial x_n}(X_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(X_0) & \frac{\partial f_m}{\partial x_2}(X_0) & \cdots & \frac{\partial f_m}{\partial x_n}(X_0) \end{pmatrix}.$$

Combining the equations above involving the linear approximation $f_i(X_0) + \nabla f_i(X_0)(X - X_0)$ to each f_i at X_0, we obtain, again for X near X_0,

$$F(X) = F(X_0) + \nabla F(X_0)(X - X_0) + o(\|X - X_0\|),$$

where the indicated product

$$\underset{m \times n}{\nabla F(X_0)} \quad \underset{n \times 1}{(X - X_0)}$$

is formed according to the standard rules of matrix/vector multiplication.

Example 1 If

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f(x,y,z) \\ g(x,y,z) \end{pmatrix} = \begin{pmatrix} \cos(\sqrt{x^2 + y^2 + z^2 - 1}) \\ \sin(\sqrt{x^2 + y^2 + z^2 - 1}) \end{pmatrix}$$

and we take $(x_0, y_0, z_0) = (1, 2, -1)$, then we have, for general X, and with $r^2 \equiv x^2 + y^2 + z^2$, $\nabla F(X_0) =$

$$\begin{pmatrix} -\sin(\sqrt{r^2 - 1})\frac{x}{\sqrt{r^2-1}} & -\sin(\sqrt{r^2 - 1})\frac{y}{\sqrt{r^2-1}} & -\sin(\sqrt{r^2 - 1})\frac{z}{\sqrt{r^2-1}} \\ \cos(\sqrt{r^2 - 1})\frac{x}{\sqrt{r^2-1}} & \cos(\sqrt{r^2 - 1})\frac{y}{\sqrt{r^2-1}} & \cos(\sqrt{r^2 - 1})\frac{z}{\sqrt{r^2-1}} \end{pmatrix}$$

which at $X_0 = (x_0, y_0, z_0) = (1, 2, -1)$ reduces to

$$\nabla F(X_0) = \begin{pmatrix} -\sin(\sqrt{5})\frac{1}{\sqrt{5}} & -\sin(\sqrt{5})\frac{2}{\sqrt{5}} & -\sin(\sqrt{5})\frac{-1}{\sqrt{5}} \\ \cos(\sqrt{5})\frac{1}{\sqrt{5}} & \cos(\sqrt{5})\frac{2}{\sqrt{5}} & \cos(\sqrt{5})\frac{-1}{\sqrt{5}} \end{pmatrix}.$$

Here we have

$$F(X_0) = \begin{pmatrix} \cos(\sqrt{5}) \\ \sin(\sqrt{5}) \end{pmatrix}$$

so the linear approximation to $F(X)$ at $X_0 = (x_0, y_0, z_0) = (1, 2, -1)$ in this case is the linear vector expression

$$\begin{pmatrix} \cos(\sqrt{5}) \\ \sin(\sqrt{5}) \end{pmatrix} + \begin{pmatrix} -\sin(\sqrt{5})\frac{1}{\sqrt{5}} & -\sin(\sqrt{5})\frac{2}{\sqrt{5}} & -\sin(\sqrt{5})\frac{-1}{\sqrt{5}} \\ \cos(\sqrt{5})\frac{1}{\sqrt{5}} & \cos(\sqrt{5})\frac{2}{\sqrt{5}} & \cos(\sqrt{5})\frac{-1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} x - 1 \\ y - 2 \\ z + 1 \end{pmatrix} =$$

$$\begin{pmatrix} \cos(\sqrt{5}) - \sin(\sqrt{5})\frac{1}{\sqrt{5}}(x-1) - \sin(\sqrt{5})\frac{2}{\sqrt{5}}(y-2) - \sin(\sqrt{5})\frac{-1}{\sqrt{5}}(z+1) \\ \\ \sin(\sqrt{5}) + \cos(\sqrt{5})\frac{1}{\sqrt{5}}(x-1) + \cos(\sqrt{5})\frac{2}{\sqrt{5}}(y-2) + \cos(\sqrt{5})\frac{-1}{\sqrt{5}}(z+1) \end{pmatrix}.$$

The matrix $\nabla F(X_0)$ is called the **Jacobian matrix** of the vector function $F(X)$ at the point X_0. When $m = n$, so that $F(X)$ is a vector field, the Jacobian matrix is an $n \times n$ square matrix and thus has a determinant, $\det(\nabla F(X_0))$. This determinant is called the **Jacobian (determinant)** of $F(x)$ at the point X_0. An important example involves the vector relationship which defines polar coordinates:

$$\begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} \sqrt{x^2 + y^2} \\ \\ \tan^{-1}\left(\frac{y}{x}\right) \end{pmatrix}.$$

In this case the Jacobian matrix, as a 2×2 matrix function of x and y, is

$$\begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix}$$

whose Jacobian, the determinant of the matrix, is $\frac{1}{\sqrt{x^2+y^2}} = \frac{1}{r}$.

Chain Rule If we have two functions:

$$Y = F(X), \quad Z = G(Y),$$

with $F : \mathbf{R}^n \to \mathbf{R}^m$ and $G : \mathbf{R}^m \to \mathbf{R}^p$, we can compose the two into a single function

$$Z = G(F(X)) \equiv H(X); \quad H : \mathbf{R}^n \to \mathbf{R}^p.$$

The domain of $H(X)$ consists of those X in the domain of $F(X)$ for which $Y = F(X)$ is in the domain of $G(Y)$. Assuming $F(X)$ and $G(Y)$ continuously differentiable, the composite function $H(X) \equiv G(H(X))$ is also continuously differentiable and the Jacobian matrix of $H(X)$ is the matrix product of the Jacobian matrix of G, with respect to Y, with the substitution $Y = F(X)$, times the Jacobian matrix of F, with respect to X, in that order:

$$\underset{p \times n}{\nabla H(X)} = \underset{p \times m}{\nabla G(F(X))} \; \underset{m \times n}{\nabla F(X)}.$$

We will not give the proof here; it is much the same as the proof of the Chain Rule for functions $f(t) = \phi(X(t))$ given in the gradient section.

Newton's Method for Systems of Equations In elementary calculus students are introduced to Newton's method for iterative solution of the scalar equation

$$f(x) = y, \; y, \text{ given},$$

under the assumption that $f(x)$ is a continuously differentiable function of x and the solution sought, $\hat{x}$, is a point where $f'(\hat{x}) \neq 0$. We suppose that the approximate location of $\hat{x}$ is known to the extent that we can choose an initial "guess", x_0, fairly close (deliberately vague; but one can be precise!) to $\hat{x}$. We then generate a sequence of points $\{x_k \,|\, k = 1, 2, ...\}$ by taking x_0 to be the point just described and defining, recursively,

$$x_{k+1} = x_k - \frac{f(x_k) - y}{f'(x_k)}, \; k = 1, 2, ... \; .$$

Under the circumstances described it can be shown that if $|x_0 - \hat{x}|$ is sufficiently small, the generated sequence x_k converges, rather rapidly in fact, to the solution $\hat{x}$ of the equation.

Essentially the same procedure works for n-vector equations, i.e., **systems** of n equations in n unknowns. Such systems take the form

$$F(x) = Y, \text{ i.e., } \begin{pmatrix} f_1(x_1, x_2, ..., x_n) \\ f_2(x_1, x_2, ..., x_n) \\ \vdots \\ f_n(x_1, x_2, ..., x_n) \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

where Y is a known vector. The general description of the method is the same as in the scalar case but some details have to be modified. We suppose that $\hat{X} = (\hat{x_1}, \hat{x_2}, ..., \hat{x_n})$ is a solution of this system of equations. We suppose that the vector function, which is in fact a vector field, $F(X)$ is continuously differentiable and, further, that the Jacobian (determinant) of $F(X)$ at the point $\hat{X}$ is non-zero, i.e.,

$$\det \nabla F(\hat{X}) \neq 0.$$

Again we suppose that we can find a point X_0 reasonably close to the solution $\hat{X}$ we seek. Since $F(X)$ is continuously differentiable, we can write

$$F(X) = F(X_0) + \nabla F(X_0)(X - X_0) + o(\|X - X_0\|).$$

Since the point X_0 is supposed to be close to $\hat{X}$, we argue that the error term at $\hat{X}$, i.e., $o(\left\|\hat{X} - X_0\right\|$, should be small. We eliminate that error term and consider the modified equation

$$F(X_0) + \nabla F(X_0)(X - X_0) = Y, \text{i.e.,} \nabla F(X_0)(X - X_0) = Y - F(X_0).$$

rather than the original equation. From the continuity of the partial derivatives of the components of $F(X)$ one can show that if X_0 is sufficiently close to the point $\hat{X}$, where we have assumed $\det \nabla F(\hat{X}) \neq 0$, it will also be true that $\det \nabla F(X_0) \neq 0$. It then follows from a familiar theorem of linear algebra that the last equation above, which we can rewrite as

$$\nabla F(X_0) X = \nabla F(X_0) X_0 - (F(X_0) - Y),$$

has a unique solution, X_1, which can be shown to be closer to $\hat{X}$ than X_0 is, provided X_0 is "sufficiently" (vague again) close to $\hat{X}$. This solution can be found by the familiar elimination procedures of linear algebra. Thus we have

$$\nabla F(X_0) X_1 = \nabla F(X_0) X_0 - (F(X_0) - Y).$$

In terms of the **inverse matrix** for $\nabla F(X_0)$, i.e., the matrix denoted $\nabla F(X_0)^{-1}$ for which

$$\nabla F(X_0) \nabla F(X_0)^{-1} = \nabla F(X_0)^{-1} \nabla F(X_0) = \mathbf{I}_n = \begin{pmatrix} 1 & 1 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix},$$

we can write

$$X_1 = X_0 - \nabla F(X_0)^{-1}(F(X_0) - Y),$$

or

$$\nabla F(X_0)\,X_1 = \nabla F(X_0)\,X_0 - (F(X_0) - Y).$$

Continuing, subsequent points are generated by the recursion formula

$$X_{k+1} = X_k - \nabla F(X_k)^{-1}(F(X_k) - Y),$$

or

$$\nabla F(X_k)\,X_{k+1} = \nabla F(X_k)\,X_k - (F(X_k) - Y),$$

which is a rather direct generalization of the scalar Newton method formula. If the initial point X_0 is sufficiently close to $\hat{X}$, the sequence of vectors $\{X_k\,|k = 1, 2, ...\}$ converges to the desired solution $\hat{X}$. In fact, it can be further shown that the convergence is *quadratic* in the following sense: there is a positive number M such for $k = 1, 2, \ldots$, we have

$$\left\|X_{k+1} - \hat{X}\right\| \leq M \left\|X_k - \hat{X}\right\|^2.$$

Example 2 If $x_0 = 2$ and $y_0 = 1$ we clearly have

$$x_0^2 - y_0^3 = 3,$$

$$x_0^3 + y_0^2 = 9.$$

Suppose we wish to find $\hat{X} = \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix}$ such that

$$F(\hat{X}) \equiv \begin{pmatrix} \hat{x}^2 - \hat{y}^3 \\ \\ \hat{x}^3 + \hat{y}^2 \end{pmatrix} = \begin{pmatrix} 4 \\ \\ 10 \end{pmatrix}.$$

Let us see how the solution comes out of the presently proposed method. The Jacobian matrix at $(x_0, y_0) = (2, 1)$ is clearly

$$\begin{pmatrix} 2x_0 & -3y_0^2 \\ 3x_0^2 & 2y_0 \end{pmatrix} = \begin{pmatrix} 4 & -3 \\ 12 & 2 \end{pmatrix}$$

whose determinant is $8 + 36 = 44$. Then we can solve the equation $\nabla F(X_0)\, X_1 = \nabla F(X_0)\, X_0 - (F(X_0) - Y)$, which in this case reduces to

$$\begin{pmatrix} 4 & -3 \\ 12 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 4 & -3 \\ 12 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ 9 \end{pmatrix} + \begin{pmatrix} 4 \\ 10 \end{pmatrix} = \begin{pmatrix} 6 \\ 27 \end{pmatrix}.$$

Solving, we have $x_1 = 2.1136,\ \ y_1 = .8182$. Continuing, we obtain

$$\begin{aligned} x_2 &= 2.1109, & y_2 &= .7724, \\ x_3 &= 2.1110, & y_3 &= .7699, \\ x_4 &= 2.1110, & y_4 &= .7699; \end{aligned}$$

we accept the last answer as an adequate approximation to $\hat{X}$ within four decimal place accuracy.