Matrices and Derivatives

For derivatives of functions of more than one variable, matrices play the role that numbers play for derivatives of functions of one variable.

First we will remember the setup for functions of one variable.
Then we'll see the generalization.
We'll use notation $y = f(x)$.

A function $f : \mathbb{R} \to \mathbb{R}$.

Here f is a function of one variable. The inputs and outputs are numbers. (The domain could be smaller, e.g. $(0, \infty)$ for $f(x) = \ln x$. We write the domain as $\mathbb{R}$ to avoid a distraction.)

Let a be some number.
Given a (small) number Δx, define $\Delta y = f(a + \Delta x) - f(a)$.

There are TWO KEY FEATURES to $f'(a)$, the derivative of f at $x = a$:

(1) The derivative is a NUMBER (defined by a certain limit).
(2) $\Delta y \approx m\Delta x$, if m is the derivative.
(I.e., for the linear approximation to Δy, multiply Δx by m.)

Example. Suppose $y = f(x)$ is the function defined by the rule $f(x) = x^2$. Suppose $a = 3$. Then the derivative $f'(a)$ is the number 6; $\Delta y = (3+\Delta x)^2 - 3^2$, and the linear approximation is

$$\Delta y \approx f'(3)\Delta x = 6\Delta x \ .$$

"Cancellation mnemonic." One notation for $f'(x)$ is dy/dx. In this notation, the linear approximation above looks like

$$\Delta y \approx \frac{dy}{dx}\Delta x \ .$$

If we think of dx as being like Δx, and "cancel", the right side is dy.

Functions $f : \mathbb{R}^2$ to $\mathbb{R}^2$.

Now we have a function of two variables, which we'll call x_1 and x_2. Instead of thinking of two inputs, we think of a single input $x = (x_1, x_2)$ (visualized as a point in the plane).

We write $y = f(x)$, just as we did for $f : \mathbb{R} \to \mathbb{R}$. Now, instead of x and y being numbers, we have $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Δx and Δy make sense; each now has two entries. Below, we write them as columns rather than rows. (Actually, in order to have the matrix multiplications work out, we always think of x and y as being columns, even though we write rows inside sentences to avoid ugly displays.)

There are THREE KEY FEATURES to $f'(a)$, the derivative of f at $x = a$:

(1) The derivative is a MATRIX (defined by a certain limit).
(2) $\Delta y \approx M\Delta x$, if M is the derivative.
(I.e., for the linear approximation to Δy, multiply Δx by M.)

(3) Computation. The entries of that matrix M are partial derivatives:

$$M = \begin{pmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 \end{pmatrix}$$

and the linear approximation is

$$\Delta y \approx M\Delta x = \begin{pmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix}$$

$$\begin{pmatrix} \Delta y_1 \\ \Delta y_2 \end{pmatrix} \approx \begin{pmatrix} (\partial y_1/\partial x_1)\Delta x_1 + (\partial y_1/\partial x_2)\Delta x_2 \\ (\partial y_2/\partial x_1)\Delta x_1 + (\partial y_2/\partial x_2)\Delta x_2 \end{pmatrix}.$$

Notice that the multiplication gives you another cancellation mnemonic.

Example. Suppose $y = (y_1, y_2)$, $x = (x_1, x_2)$ and $y = f(x)$ is defined by

$$y_1 = 5(x_1)^2 x_2 \qquad \text{and} \qquad y_2 = (x_1)^3(x_2)^2 \quad .$$

Then

$$\begin{pmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 \end{pmatrix} = \begin{pmatrix} 10x_1x_2 & 5(x_1)^2 \\ 3(x_1)^2(x_2)^2 & (x_1)^3 2x_2 \end{pmatrix} \quad .$$

Let us consider the particular input $a = (1, 2)$ and a particular $\Delta x = (.1, .2)$. Then the derivative at a is

$$\begin{pmatrix} 20 & 5 \\ 12 & 4 \end{pmatrix}$$

and the linear approximation to $\Delta y = f(a + \Delta x) - f(x)$ is

$$\Delta y = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \end{pmatrix} \approx \begin{pmatrix} 20 & 5 \\ 12 & 4 \end{pmatrix} \begin{pmatrix} .1 \\ .2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

giving

$$f(1.1, 2.2) \approx f(1, 2) + (3, 2) = (10, 4) + (3, 2) = (13, 6) \ .$$

The general case. A function $f : \mathbb{R}^n \to \mathbb{R}^m$.

Here m and n are positive integers. Now an input is $x = (x_1, \dots x_n)$ and an output is $(y_1, \dots, y_m)$. The derivative is an $m \times n$ matrix. The entry of the matrix in row i and column j is $\partial y_i/\partial x_j$. Then everything goes as in the 2 by 2 case above.

A function $f : \mathbb{R}^2 \to \mathbb{R}$.

This corresponds to $n = 2, m = 1$ in the general case above.

Example. $y = f(x) = x_1(x_2)^3$. In the general notation we used, $y = (y_1)$, and the derivative is the 1×2 matrix

$$M = \begin{pmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \end{pmatrix} = \begin{pmatrix} (x_2)^3 & 3x_1(x_2)^2 \end{pmatrix} .$$

For the particular input $a = (1, 2)$ and a particular $\Delta x = (.3, .2)$, the derivative is $M = (8, 12)$ and we have the linear approximation

$$\begin{aligned} \Delta y \approx M\Delta x &= \begin{pmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} \\ &= \begin{pmatrix} 8 & 12 \end{pmatrix} \begin{pmatrix} .3 \\ .2 \end{pmatrix} \\ &= (2.4 + 2.4) = (4.8) . \end{aligned}$$

So, $f(1.3, 2.2) \approx f(1, 3) + M\Delta x = 27 + 4.8 = 31.8$.

Sample problem.
For the function $y = f(x)$ from $\mathbb{R}^2 \to \mathbb{R}^2$ given by the rule $y_1 = 4x_1 \cos(x_1 x_2)$, $y_2 = (x_1) \sin(3x_2)$, compute the derivative at the input $(1, \pi/2)$ and use it to compute the linear approximation to $f(1 + .02, \pi/2 + .03)$.

Remark. When we are being systematic, it's very helpful to use subscripted coordinate variables like x_i, y_j. On the other hand, when we only have a few variables, then it's often easier to have different letters for them; e.g., $z = f(x, y)$ instead of $(y_1) = f(x_1, x_2)$. In the end, when you're free to name, use the notation which works best for you for your particular situation.