The Derivative

Our goal is to give a definition of the derivative that is equivalent to the old definition, but will provide better guidance when using the derivative. We first examine the derivative in the one dimensional case.

The Derivative in Dimension One

Here are four guises in which the derivative appears.

⋆ Let \( y = f(x) \) be a function. Then its derivative at the point \( x \) is

\[
f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}
\]

⋆ The derivative of \( f(x) \) at \( x = x_0 \) is the slope of the tangent to the graph of \( y = f(x) \) at \( x = x_0 \).

⋆ The best linear approximation to \( f \) near \( x = x_0 \) is

\[
f(x) - f(x_0) \sim f'(x_0)(x - x_0).
\]

⋆ If \( t \) denotes time, and \( s \) denotes the position of a particle, then the instantaneous velocity of the particle at time \( t_0 \) is

\[
s'(t_0).
\]

We look at the interpretation of the derivative as instantaneous velocity more carefully. Let \( v \) denote \( s'(t_0) \). We have

\[
s(t) - s(t_0) \sim v \cdot (t - t_0).
\]

We can think of the right hand side of this relation as a function with

• input equal to change in time \( t - t_0 \),

• output equal to change in postion or displacement,

• the 'rule' governing the function is multiplication by \( v \), and

• And this function is a good approximation to \( s(t) - s(t_0) \).
Note that this is a linear function \((v) : \mathbb{R} \to \mathbb{R}\). It is given by a \(1 \times 1\) matrix with entry equal to \(v\).

We now give a first approximation to the definition of the derivative in the abstract case. We start with \(f : \mathbb{R} \to \mathbb{R}\) and \(x_0 \in \mathbb{R}\).

**Definition 1.** The derivative of \(f\) is a linear function \(L\) with domain equal to a copy of real numbers measuring change from \(x = x_0\) and range equal to a copy of the real numbers measuring displacement from \(f(x_0)\). This function has the property that \(L(x - x_0) \sim f(x) - f(x_0)\), that is, \(L(x - x_0)\) is the best approximation to \(f(x) - f(x_0)\).

Of course \(L(x - x_0) = f'(x_0) \cdot (x - x_0)\). We make this concrete. Define \(\epsilon\) by

\[
 f(x) - f(x_0) = m \cdot (x - x_0) + \epsilon
\]

Note that \(\epsilon\) depends on \(x\) and \(m\) so we could write something like \(\epsilon_{m,x}\).

**Definition 2.** The derivative of \(f(x)\) at \(x = x_0\) is \(m\) provided

\[
 f(x) - f(x_0) = m \cdot (x - x_0) + \epsilon
\]

and \(\frac{\epsilon}{x - x_0} \to 0\) as \(x \to x_0\).

We show that this is the same as the classical definition. Assume that \(m\) is the derivative of \(f\) at \(x = x_0\) according to the above definition. Dividing equation (1) by \(x - x_0\) gives

\[
 \frac{f(x) - f(x_0)}{x - x_0} = m + \frac{\epsilon}{x - x_0}.
\]

According to our definition of the derivative \(\lim_{x \to x_0} \frac{\epsilon}{x - x_0} = 0\). This says that

\[
 \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = m.
\]

This is the classical definition of the derivative in the one variable case.

**Example 1.** We give a numerical example. Let \(f(x) = 2x^2 - 3x + 1\) so \(f(2) = 3\), and \(f'(2) = 5\). We calculate. The entry \(\epsilon(h)\) in the table is the difference \((f(2 + h) - f(2)) - f'(2)h\).

<table>
<thead>
<tr>
<th>(h)</th>
<th>(f(2 + h) - f(2))</th>
<th>(f'(2)h)</th>
<th>(\epsilon(h))</th>
<th>(\epsilon(h)/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10^{-1}</td>
<td>.52</td>
<td>.5</td>
<td>.02</td>
<td>.2</td>
</tr>
<tr>
<td>10^{-2}</td>
<td>.0502</td>
<td>.05</td>
<td>.002</td>
<td>.02</td>
</tr>
<tr>
<td>10^{-3}</td>
<td>.005002</td>
<td>.005</td>
<td>2 * 10^{-6}</td>
<td>.002</td>
</tr>
</tbody>
</table>

**Definition of the Derivative**

We now give the definition of the derivative in the general case. Let \(F : \mathbb{R}^m \to \mathbb{R}^n, F(P) = Q\). We denote the vector space of displacements from \(P\) by \(\mathbb{R}^m_P\) and the vector space of displacements from \(Q\) by \(\mathbb{R}^n_Q\). You can think of the vectors space \(\mathbb{R}^m_P\) as the set arrows with tail at \(P\). If \(v \in \mathbb{R}^n, \|v\|_m\) denotes, for the moment, the maximum absolute value of the components of \(x\).
**Definition 3.** The derivative of $F$ at $P$ is the linear map

$$DF(P) : \mathbb{R}^m_p \to \mathbb{R}^n_Q$$

such that

$$F(x) - F(P) = DF(P)(x - P) + \epsilon$$

where $||x - P||_m \to 0$, as $||x - P||_m \to 0$.

In this case we say that $DF(P)$ is the best linear approximation to $F$ near $P$.

**Computing the Derivative**

**Case 1:**
Assume that $f : \mathbb{R} \to \mathbb{R}$. Write $Df(x)$ as the $1 \times 1$ matrix $(f'(x))$. We have

$$f(x + h) - f(x) = Df(x) \cdot (h) + \epsilon$$

with

$$\frac{\epsilon}{||h||_m} \to 0, \text{ as } ||h||_m \to 0.$$

**Case 2:**
Let $f : \mathbb{R} \to \mathbb{R}^n$.

Write $f(t_0) = s_0 \in \mathbb{R}^n$. Then the derivative of $f$ at $t_0 \in \mathbb{R}$ is a linear function $Df(t_0)$ that maps $\mathbb{R}_{t_0}$ to $\mathbb{R}^n_{s_0}$. In terms of matrices it is matrix of size $n \times 1$, that is it is a column vector of length $n$. It predicts how $f(t)$ changes near $t = t_0$. Write

$$f : t \to \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{pmatrix}.$$

We handle each $f_i$ separately using the one variable case. We get

$$f(t_0 + h) - f(t_0) = \begin{pmatrix} f_1(t_0 + h) - f(t_0) \\ f_2(t_0 + h) - f(t_0) \\ \vdots \\ f_n(t_0 + h) - f(t_0) \end{pmatrix} = \begin{pmatrix} \frac{df_1}{dt}(t_0)(h) + \epsilon_1 \\ \frac{df_2}{dt}(t_0)(h) + \epsilon_2 \\ \vdots \\ \frac{df_n}{dt}(t_0)(h) + \epsilon_n \end{pmatrix}$$

By the one variable case all the $\frac{\epsilon_i}{h} \to 0$.

$$Df(t_0) = \begin{pmatrix} \frac{df_1}{dt} \\ \frac{df_2}{dt} \\ \vdots \\ \frac{df_n}{dt} \end{pmatrix}(t_0).$$
Case 3:
Let \( f : \mathbb{R}^m \to \mathbb{R} \) and write \( f(p) = f(x_1, x_2, \ldots, x_m) = q \in \mathbb{R} \). Then the derivative of \( f \) at \( p \) is a \( 1 \times m \) matrix that predicts how \( f \) changes when \( p \) changes. The change in \( f \) comes is the sum of the changes from the change in each variable \( x_i \). Let \( h = (h_1, h_2, \ldots, h_m) \). Then the change in \( f \) as \( x \) changes from \( p \) to \( p + h \) is approximately

\[
f(p + h) - f(p) = \left( \frac{\partial f}{\partial x_0} \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_m} \right) (p) \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{pmatrix}.
\]

The derivative of \( f \) at \( p \) is

\[
\left( \frac{\partial f}{\partial x_0} \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_m} \right) (p).
\]

In words we are saying that

\[
(\text{The change in } f) = \sum_i (\text{change in } f \text{ due to change in } x_i).
\]

Case 4: We write out the general case. Let

\[
f : \mathbb{R}^m \to \mathbb{R}^n, \quad f : x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \mapsto \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}
\]

By using Case 3 to treat each of the \( f_i \) separately we get that the derivative of \( f \) at a point

\[
x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_m \end{pmatrix}
\]

is

\[
Df(P) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots \\ \vdots & \vdots & \ddots \\ \vdots & \vdots & \ddots \end{pmatrix} (p).
\]
Example: View

\[ s : \mathbb{R} \to \mathbb{R}^3 \]
\[ t \to \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} \]
as defining the position \( s(t) \) of a particle at time \( t \). The derivative of \( s(t) \) at \( t = t_0 \) is a linear function from a copy of \( \mathbb{R} \) to a copy of \( \mathbb{R}^3 \). We denote the copy of \( \mathbb{R} \) as \( \mathbb{R}_{t_0} \) since it measures changes in \( t \) from \( t = t_0 \). Let \( s(t_0) = s_0 \in \mathbb{R}^3 \). We denote the copy of \( \mathbb{R}^3 \) by \( \mathbb{R}_{s_0}^3 \) since it measures changes in position from \( s = s_0 \). Then the derivative is the linear function

\[ Ds(t_0) : \mathbb{R}_{t_0} \to \mathbb{R}_{s_0}^3 \]
so that

\[ s(t) - s(t_0) \approx Ds(t_0)(t - t_0) \]
or

\[ s(t) \approx s(t_0) + Ds(t_0)(t - t_0). \]
For one thing this says that \( Ds(t_0) \) is a linear function from \( \mathbb{R} \to \mathbb{R}^3 \). This means that it is given by a \( 3 \times 1 \) matrix, that is, a length 3 column vector. Indeed it is

\[ Ds(t_0) = \begin{pmatrix} x'(t_0) \\ y'(t_0) \\ z'(t_0) \end{pmatrix}. \]

We can interpret this as the velocity vector at \( t = t_0 \).

Example Let \( F : \mathbb{R}^3 \to \mathbb{R} \) be a nice function. We will think of it as giving the temperature at point \( \begin{pmatrix} x \\ y \\ z \end{pmatrix} \) in the ocean. Assume that the temperature at point \( p \) is \( T \). The derivative of \( F \) at \( p \) is a linear function from \( \mathbb{R}_p^3 \) to \( \mathbb{R}_T \). Thus it is given by a matrix of size \( 1 \times 3 \), a row vector. Its entries are

\[ \left( \frac{\partial F}{\partial x}(p), \frac{\partial F}{\partial y}(p), \frac{\partial F}{\partial z}(p) \right). \]
This is called the gradient of \( F \). We have

\[ F(p + \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}) \approx F(p) + \left( \frac{\partial F}{\partial x}(p), \frac{\partial F}{\partial y}(p), \frac{\partial F}{\partial z}(p) \right) \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = h_1 \frac{\partial F}{\partial x}(p) + h_2 \frac{\partial F}{\partial y}(p) + \frac{\partial F}{\partial z}(p)h_3. \]

We read this as

\[ \text{Change in temperature} = \]
\[ \text{Change from change in } x \text{ position} + \]
\[ \text{Change from change in } y \text{ position} + \]
\[ \text{Change from change in } z \text{ position.} \]
**Notation** In the case of a function \( f : \mathbb{R} \to \mathbb{R} \) we call the derivative the gradient and denote it by

\[
\nabla F(P) =
\]

**Example:** Let

\[
F : \mathbb{R}^2 \to \mathbb{R}^2 \quad \text{and} \quad F : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix}.
\]

The derivative of \( F \) at a general point is

\[
\begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}.
\]

We choose a point arbitrarily, say \( p = (1, 3) \), so that

\[
DF(p) = \begin{pmatrix} 2 & -6 \\ 6 & 2 \end{pmatrix}.
\]

Thus the best linear approximation to \( F \) near \( p \) is given by the function

\[
a f(x, y) = F(p) + DF(p) \begin{pmatrix} x - 1 \\ y - 3 \end{pmatrix}.
\]

In the figure below labeled “Derivative Versus Function” we have shown in red the image under \( F \) of a small square near \( p = (1, 3) \). The blue figure is the image of the same small square under the best linear approximation to \( F \) near \( p \).

**Question:** How does \( F \) distort area?

**Answer:**

Recall that the determinant of a \( 2 \times 2 \) matrix is the area of the parallelogram spanned by the column vectors of the matrix, at least up to sign. Let \( u, v \) be two short vectors based at \( P \in \mathbb{R}^2 \). Then \( F \) maps

\[
P \to F(P),
\]

\[
P + u \to F(p + u) \quad \text{which is super close to} \quad F(P) + DF(p)(v),
\]

\[
P + v \to f(P + v) \quad \text{which is super close to} \quad DF(P)(v).
\]

Hence the parallelogram based at \( P \) and spanned by \( u, v \) is mapped to the parallelogram based at \( F(P) \) and spanned by \( DF(P)(u), DF(P)(v) \). The ratio of the areas of these two parallelograms is the area distortion. This is the absolute value of the determinant of \( DF(P) \).

**Example 2.** We work with the function \( F \) above:

\[
F : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix}.
\]
Figure 1: Domain Squares, One Red, One Green
Figure 2: Image of Red Square
Figure 3: Image of Small Green Square
Figure 4: Derivative Versus Function
Let $u = \begin{pmatrix} t \\ 0 \end{pmatrix}, v = \begin{pmatrix} 0 \\ t \end{pmatrix}$ with $t$ small. The area of the parallelogram spanned by $u,v$ is $t^2$. The image parallelogram is spanned by the vectors $u_1 = DF(P)(u), v_1 = DF(P)(v)$. These two vectors are the column vectors of the matrix $t \cdot DF(P)$. The area is $t^2$ times $|det(DF(P))|$. 