Let's attempt to understand what differentials mean, and what is the intuition behind them.

→ If \( f : \mathbb{R} \to \mathbb{R} \) (written as \( f(x) \)), for any \( h \in \mathbb{R} \):
  \[ \delta f [ x^* ; h ] := f'(x^*) \cdot h = \frac{df}{dx} \bigg|_{x=x^*} \cdot h \]

→ If \( f : \mathbb{R}^n \to \mathbb{R} \) (write \( f(x) \)), then for \( h \in \mathbb{R}^n \):
  \[ \delta f [ x^* ; h ] := \nabla f(x^*) \cdot h = \sum_{i=1}^{n} \frac{df}{dx_i} \bigg|_{x=x^*} \cdot h_i \]

→ If \( f : \mathbb{R}^n \to \mathbb{R}^m \) (write \( f(x) \)), then \( \forall h \in \mathbb{R}^n \):
  \[ \delta f_j [ x^* ; h ] := \frac{df_j}{dx} \bigg|_{x=x^*} \cdot h \]

  *Matrix-Vector Product*

  *Jacobian*

  or, in "index notation"

  \[ \delta f_i [ x^* ; h ] = \sum_{j=1}^{m} \frac{df_i}{dx_j} \bigg|_{x=x^*} \cdot h_j \quad 1 \leq i \leq m \]
You might remember seeing such expressions before... they show up in Taylor (1st order) formulas:

e.g.

\[ f: \mathbb{R} \to \mathbb{R} \]

\[ f(x^* + h) \approx f(x^*) + f'(x^*)h + o(h^2) \]

\[ \delta f[x^*; h] \]

\[ f: \mathbb{R}^n \to \mathbb{R}^n \]

\[ f(x^* + h) \approx f(x^*) + \frac{\partial f}{\partial x} \bigg|_{x^*} h + o(\|h\|^2) \]

\[ \delta f[x^*; h] \]

In this sense, the differential is the linear term in the Taylor approximation of the difference

\[ f(x^* + h) - f(x^*) \]

This also gives us a concrete formula to define the differential:

\[
\delta f[x^*; h] = \lim_{\epsilon \to 0} \left[ \frac{f(x^* + \epsilon h) - f(x^*)}{\epsilon} \right]
\]

Since eq. (1) doesn't (explicitly) mention derivatives, we could consider taking differentials of more complex functions, e.g. those taking matrices as input (or output), e.g.:

\[ f(M) = \text{det} M \]

\[ F(M) = M^{-1} \]
There are many convenient features of differentials, including:

- In many cases, we can compute a differential without needing the derivative.
- The derivative might be a matrix (e.g. $\frac{df}{dx}$) which is cumbersome to store/compute (we saw the pseudocode that computes $\delta f$).
- The derivative might be a "weird" algebraic object, but the differential is always the same "type" as $f(x)$.
  
  \[ F(M) = M^{-1} \]
  
  We will see \[ \delta F [M^*; H] = -(M^*)^{-1} H M^* \]
  
  the derivative $\frac{df}{dM}$ is a 4th order tensor.

- Derivative properties are inherited to differentials.

**Addition**

\[ \delta (f+g) [x^*; h] = \delta f [x^*; h] + \delta g [x^*; h] \]

**Multiplication**

\[ \delta (f \cdot g) [x^*; h] = \delta f [x^*; h] \cdot g(x^*) + f(x^*) \cdot \delta g [x^*; h] \]

Also works for dot products, matrix products, etc.
Composition ("Chain Rule")

\[ \delta \{ f \circ g \} [x^*; h] = \delta f \left[ g(x^*) \right] \delta g [x^*; h] \]

(i.e. \( \delta \{ f(g) \} \))

\[ \Rightarrow \text{Differentials can help us compute derivatives!} \]

If we somehow derive \( \delta f [x^*; h] = A \cdot h \)

then \( A \) has to be \( \left[ \frac{df}{dx} \right]_{x^*} \)

Examples

W.l.o.g we use \( \delta x \) in place of \( h \), and shorten \( \delta f [x; \delta x] \) to just \( \delta f \) ...

St. Venant Kirchhoff

\[ P = F \left\{ 2\mu E + 2\lambda I \cdot E \cdot I \right\} \]

where \( E = \frac{1}{2} (FF + I) \)

\[ \Rightarrow \delta E = \frac{1}{2} \left( \delta F^T F + F^T \delta F \right) \]

\[ \delta P = \delta F \left\{ 2\mu \delta E + 2\lambda I \cdot I \right\} \]

\[ + F \left\{ 2\mu \delta E + 2\lambda I \cdot (\delta E) \cdot I \right\} \]