A primer on differentials & tensor calculus

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 $h \in \mathbb{R}$:

then $\delta f[x^*; h] := f'(x^*) \cdot h = \left.\frac{df}{dx}\right|_{x=x^*} \cdot h$

→ If $f: \mathbb{R}^n \to \mathbb{R}$ (write $f(\underline{x})$), then for $\underline{h} \in \mathbb{R}^n$

$$\delta f[\underline{x}^*; \underline{h}] := \nabla f(\underline{x}^*) \cdot \underline{h} = \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i}\right|_{\underline{x}=\underline{x}^*} \cdot h_i$$

dot product

→ If $\underline{f}: \mathbb{R}^n \to \mathbb{R}^m$ (write $\underline{f}(\underline{x})$), then $\forall \underline{h} \in \mathbb{R}^n$

$$\delta \underline{f}[\underline{x}^*; \underline{h}] := \left.\frac{\partial \underline{f}}{\partial \underline{x}}\right|_{\underline{x}=\underline{x}^*} \cdot \underline{h}$$

Matrix-Vector Product

Jacobian

or, in "index notation"

$$\delta f_i[\underline{x}^*; \underline{h}] = \sum_{j=1}^{n} \left.\frac{\partial f_i}{\partial x_j}\right|_{\underline{x}=\underline{x}^*} h_j \qquad 1 \le i \le m$$

specific component

You might remember seeing such expressions before ... they show up in Taylor (1st order) formulas:

e.g.

$f: \mathbb{R} \to \mathbb{R}$ $$f(x^* + h) \approx f(x^*) + \underbrace{f'(x^*) h}_{\delta f[x^*; h]} + o(h^2)$$

$f: \mathbb{R}^n \to \mathbb{R}^n$ $$\underline{f}(\underline{x}^* + \underline{h}) \cong \underline{f}(\underline{x}^*) + \underbrace{\left.\frac{\partial \underline{f}}{\partial \underline{x}}\right|_{\underline{x}^*} \cdot \underline{h}}_{\delta \underline{f}[\underline{x}^*; \underline{h}]} + o(\|\underline{h}\|^2)$$

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:

$$\delta f[x^*; h] = \lim_{\epsilon \to 0} \left[\frac{f(x^* + \epsilon h) - f(x^*)}{\epsilon} \right] \tag{1}$$

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.:

$f(M) = \det M$, or $F(M) = M^{-1}$ (more on this later)

There are many convenient features of differentials, including:

→ In many cases, we can compute a differential <u>without</u> needing the derivative

⇒ The derivative might be a matrix (e.g. $\partial f / \partial x$) which is cumbersome to store / compute (we saw the pseudocode that computes δf)

⇒ The derivative might be a "weird" algebraic object, but the differential is always the same "type" as $f(x)$

e.g. $F(M) = M^{-1}$

we will see $\delta F[M^*; H] = -(M^*)^{-1} H M^*$

the derivative $\partial F / \partial M$ is a <u>4th order tensor</u>

→ Derivative properties are inherited to differentials

<u>Addition</u>

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

<u>Multiplication</u>

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

Also works for dot products matrix products, etc

Composition ("Chain Rule")

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

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

→ Differentials can help us compute derivatives!

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

then A has to be $\left.\partial f/\partial x\right|_{x^*}$!

Examples

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

St. Venant Kirchhoff

$$P = F\{2\mu E + \lambda\, \mathrm{tr}\, E\cdot I\}$$

where $E = \frac{1}{2}(F^TF + I)$

$$\Rightarrow \quad \delta E = \frac{1}{2}(\delta F^T F + F^T\delta F)$$

$$\delta P = \delta F\{2\mu E + \lambda\, \mathrm{tr}\, E\cdot I\}$$
$$+ F\{2\mu\,\delta E + \lambda\, \mathrm{tr}(\delta E)\cdot I\}$$