Short Explanation: Multivariate Taylor Mode

Say we have a function $\boldsymbol{f} : \mathbb{R}^D \to \mathbb{R}^C$ and are interested in calculating its partial derivative with respect to some multi-index $\mathbf{k}$, which is a D-tuple of natural numbers $k_i \in \mathbb{N}$, where $i = 1, \ldots, D$. Given a point of interest $\mathbf{x_0}$, we write:

$$\partial^{\mathbf{k}} \boldsymbol{f}(\mathbf{x_0}) := \boldsymbol{f}_{\mathbf{k}}.$$

Faà di Bruno Adapting the formula from [1] to the tensor notation of [2], we get:

$$\boldsymbol{f}_{\mathbf{k}} = \sum_{\sigma \in \text{multpart}(\tau_{\mathbf{k}})} \nu(\sigma) \left\langle \partial^{|\sigma|} \boldsymbol{f}(\mathbf{x}_0), \bigotimes_{\tau_{\mathbf{s}} \in \sigma} \mathbf{x_s} \right\rangle \quad \text{with} \quad \nu(\sigma) = \frac{\mathbf{k}!}{\left(\prod_{\tau_{\mathbf{s}} \in \sigma} n(\tau_{\mathbf{s}})!!\right)\left(\prod_{\tau_{\mathbf{s}} \in \bar{\sigma}} m_{\tau_{\mathbf{s}}}(\bar{\sigma})!\right)}. \tag{1}$$

To this a few key points:

- A *multi-set* to a given multi-index is defined as: $\tau_{\mathbf{k}} := \{\underbrace{1, \ldots, 1}_{k_1 \text{ times}}, \ldots, \underbrace{D, \ldots, D}_{k_D \text{ times}}\}$, e.g. if $\mathbf{k} = (3, 0, 2)$, then $\tau_{\mathbf{k}} = \{1, 1, 1, 3, 3\}$.
- A *partition* of a multi-set or a *multi-partition* is an expression of a multi-set as a *sum* of multi-sets. Here a *sum* of multi-sets is a union of multi-sets where we add the multiplicities, e.g. a possible partition of $\tau_{\mathbf{k}} = \{1, 1, 1, 3, 3\}$ is $\{1, 1, 1, 3, 3\} = \{1, 1, 3\} + \{1, 3\}$.
- $\text{multpart}(\tau_{\mathbf{k}})$ is a set containing all the possible multi-partitions of $\tau_{\mathbf{k}}$ corresponding to $\mathbf{k}$.
- If we define the number of occurrences of an integer i in a multi-set σ as $n_i(\sigma)$, we can define:

  $$n(\sigma)!! := \prod_{i \in \mathbb{N}} n_i(\sigma)!.$$

  So, for example, $n(\{1, 1, 3\}) = 2! \cdot 1!$.
- Finally, given a multi-set partition $\sigma = \tau_1 + \cdots + \tau_1 + \tau_2 + \cdots + \tau_2 + \ldots$ with repeated summands we define the number of occurrences of those summands as $m_{\tau_i}(\bar{\sigma})$. So, if $\sigma = \{1, 1\} + \{1, 1\} + \{2\}$, then $m_{\{1,1\}}(\bar{\sigma}) = 2$.

To get a better idea, of where this is all coming from, one should consult [1] (which can be found here: `https://arxiv.org/pdf/math/0601149`).

Multijets In multivariate Taylor Mode one generally propagates multivariate Taylor Polynomials starting from the simplest cases for functions. Let $\boldsymbol{f} : \mathbb{R}^D \to \mathbb{R}$

- $\boldsymbol{f} = \alpha$, for $\alpha \in \mathbb{R}^D$. In this case every partial derivative to a given multi-index is zero, except $\partial^{\mathbf{0}} \boldsymbol{f} = \boldsymbol{f} = \alpha$.

- $\boldsymbol{f} = x_i$, for $i = 1, \dots, D$. In this case every partial derivative is zero, except the derivative corresponding to the "canonical" multi-index $\mathbf{e}_i$, i.e. $\partial^{\mathbf{e}_j} \boldsymbol{f} = \boldsymbol{f}_{\mathbf{e}_i} = \delta_{ij}$ with the Kronecker-Delta.

Now, if $\boldsymbol{f} = \boldsymbol{g} \circ \boldsymbol{h} : \mathbb{R}^D \to \mathbb{R}$ in general one would propagate the Taylor Polynomial up to some degree $K \in \mathbb{N}$:

$$\sum_{|\mathbf{k}| \leq K} \frac{1}{k_1! \cdot \dots \cdot k_D!} \boldsymbol{h}_{\mathbf{k}} (\mathbf{x} - \mathbf{x_0})^{\mathbf{k}} \quad \to \quad \sum_{|\mathbf{k}| \leq K} \frac{1}{k_1! \cdot \dots \cdot k_D!} \boldsymbol{g}_{\mathbf{k}} (\mathbf{x} - \mathbf{x_0})^{\mathbf{k}}$$

It turns out, you do not need to propagate the whole thing all the time. To calculate a specific $\boldsymbol{g}_{\mathbf{k}}$ all you need are the partial derivatives with multi-indices component-wise less than or equal to $\mathbf{k}$, i.e. you need $\boldsymbol{h}_{\mathbf{j}}$, such that $\mathbf{j} \leq \mathbf{k}$. A multi-jet is now just a list of those necessary partial derivatives.
Thus, one starts with

$$\mathbf{x} \to (\mathbf{x_0}, \mathbf{x}_{\mathbf{e}_1}, \mathbf{x}_{\mathbf{e}_2}, \dots, \mathbf{x}_{\mathbf{e}_D}, \mathbf{x}_{\mathbf{e}_1+\mathbf{e}_2}, \dots) = (\mathbf{x_0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_D, \mathbf{0}, \dots)$$

and then propagates using (1). It becomes the following propagation scheme:

$$\begin{pmatrix} \mathbf{x_0} \\ \mathbf{x}_{\mathbf{e}_1} \\ \mathbf{x}_{\mathbf{e}_2} \\ \vdots \\ \mathbf{x}_{\mathbf{e}_1+\mathbf{e}_2} \\ \mathbf{x}_{\mathbf{e}_1+\mathbf{e}_3} \\ \vdots \\ \mathbf{x}_{\mathbf{k}} \end{pmatrix} \to \begin{pmatrix} \boldsymbol{h} = \boldsymbol{h}(\mathbf{x_0}) \\ \boldsymbol{h}_{\mathbf{e}_1} = \langle \partial \boldsymbol{h}(\mathbf{x_0}), \mathbf{x}_{\mathbf{e}_1} \rangle \\ \boldsymbol{h}_{\mathbf{e}_2} = \langle \partial \boldsymbol{h}(\mathbf{x_0}), \mathbf{x}_{\mathbf{e}_2} \rangle \\ \vdots \\ \boldsymbol{h}_{\mathbf{e}_1+\mathbf{e}_2} = \langle \partial^2 \boldsymbol{h}(\mathbf{x_0}), \mathbf{x}_{\mathbf{e}_1} \otimes \mathbf{x}_{\mathbf{e}_2} \rangle + \langle \partial \boldsymbol{h}(\mathbf{x_0}), \mathbf{x}_{\mathbf{e}_1+\mathbf{e}_2} \rangle \\ \boldsymbol{h}_{\mathbf{e}_1+\mathbf{e}_3} = \langle \partial^2 \boldsymbol{h}(\mathbf{x_0}), \mathbf{x}_{\mathbf{e}_1} \otimes \mathbf{x}_{\mathbf{e}_3} \rangle + \langle \partial \boldsymbol{h}(\mathbf{x_0}), \mathbf{x}_{\mathbf{e}_1+\mathbf{e}_3} \rangle \\ \vdots \\ \boldsymbol{h}_{\mathbf{k}} = \sum_{\sigma \in \mathrm{multpart}(\tau_{\mathbf{k}})} \nu(\sigma) \left\langle \partial^{|\sigma|} \boldsymbol{h}(\mathbf{x_0}), \bigotimes_{\tau_{\mathbf{s}} \in \sigma} \mathbf{x}_{\mathbf{s}} \right\rangle \end{pmatrix}$$
$$\to \begin{pmatrix} \boldsymbol{g}_{\mathbf{0}} = \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}) \\ \boldsymbol{g}_{\mathbf{e}_1} = \langle \partial \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}), \boldsymbol{h}_{\mathbf{e}_1} \rangle \\ \boldsymbol{g}_{\mathbf{e}_2} = \langle \partial \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}), \boldsymbol{h}_{\mathbf{e}_2} \rangle \\ \vdots \\ \boldsymbol{g}_{\mathbf{e}_1+\mathbf{e}_2} = \langle \partial^2 \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}), \boldsymbol{h}_{\mathbf{e}_1} \otimes \boldsymbol{h}_{\mathbf{e}_2} \rangle + \langle \partial \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}), \boldsymbol{h}_{\mathbf{e}_1+\mathbf{e}_2} \rangle \\ \boldsymbol{g}_{\mathbf{e}_1+\mathbf{e}_3} = \langle \partial^2 \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}), \boldsymbol{h}_{\mathbf{e}_1} \otimes \boldsymbol{h}_{\mathbf{e}_3} \rangle + \langle \partial \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}), \boldsymbol{h}_{\mathbf{e}_1+\mathbf{e}_3} \rangle \\ \vdots \\ \boldsymbol{g}_{\mathbf{k}} = \sum_{\sigma \in \mathrm{multpart}(\tau_{\mathbf{k}})} \nu(\sigma) \left\langle \partial^{|\sigma|} \boldsymbol{g}(\boldsymbol{h}_{\mathbf{0}}), \bigotimes_{\tau_{\mathbf{s}} \in \sigma} \boldsymbol{h}_{\mathbf{s}} \right\rangle \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_{\mathbf{0}} \\ \boldsymbol{f}_{\mathbf{e}_1} \\ \boldsymbol{f}_{\mathbf{e}_2} \\ \vdots \\ \boldsymbol{f}_{\mathbf{e}_1+\mathbf{e}_2} \\ \boldsymbol{f}_{\mathbf{e}_1+\mathbf{e}_3} \\ \vdots \\ \boldsymbol{f}_{\mathbf{k}} \end{pmatrix} \tag{2}$$

References

[1] Michael Hardy. *Combinatorics of Partial Derivatives.* arXiv, 2006.

[2] Felix Dangel et al. "Collapsing Taylor Mode Automatic Differentiation". In: *arXiv* (2025).