NOTE ON JACOBIANS OF FUNCTIONS ON LIE GROUPS

ADAM WILLIAMS

1. Derivatives as Sensitivity Measurements

Consider a scalar-valued function of a scalar variable, $y = f(x)$. The derivative of f can be viewed as a "sensitivity measurement" of f, in the sense that it maps **first-order perturbations** δ of x to first-order perturbations ϵ of y. A first-order perturbation of a scalar x is another scalar δ that is "infinitesimally small" in the sense that $\delta^2 = 0$. The derivative of f at x satisfies the relation

$$f(x+\delta) \approx f(x) + df_x\delta,$$

where $\approx$ indicates equality up to first order (i.e. equality assuming $\delta^2 = 0$).

For example. consider $f(x) = x^2$. How does f change under a first-order perturbation of x?

$$\begin{aligned} f(x+\delta) &= (x+\delta)^2 \\ &= x^2 + 2x\delta + \delta^2 \\ &\approx x^2 + 2x\delta \\ &= f(x) + 2x\delta. \end{aligned}$$

Hence df_x is the linear mapping $\delta \mapsto 2x\delta$. The interpretation of this is: "if we change x by a small amount δ, then x^2 will (approximately) change by the small amount $2x\delta$."

You can derive various "rules" for computing derivatives of complicated scalar functions:

(1) **Sum rule:**

$$\begin{aligned} (f+g)(x+\delta) &= f(x+\delta) + g(x+\delta) \\ &\approx f(x) + df_x\delta + g(x) + dg_x\delta \\ &= (f+g)(x) + (df_x + dg_x)\delta. \end{aligned}$$

Hence

$$d(f+g) = df + dg.$$

(2) **Product rule:**

$$\begin{aligned} (fg)(x+\delta) &= f(x+\delta)g(x+\delta) \\ &\approx (f(x) + df_x\delta)(g(x) + dg_x\delta) \\ &\approx f(x)g(x) + (df_x\delta)g(x) + f(x)(dg_x\delta). \end{aligned}$$

Hence

$$d(fg) = dfg + fdg.$$

(3) **Chain rule:**

$$\begin{aligned}(f \circ g)(x+\delta) &= f(g(x+\delta)) \\ &\approx f(g(x) + dg_x \delta) \\ &\approx f(g(x)) + df_{g(x)} dg_x \delta.\end{aligned}$$

Hence

$$d(f \circ g) = df_g dg.$$

A similar story holds if we replace scalars with general vectors, where the "first-order condition" is that $\|\delta\|^2 = 0$ for vector perturbations δ. For example, we can compute the derivative of the squared-norm function $n(x) = x^\top x$:

$$\begin{aligned}n(x+\delta) &= (x+\delta)^\top (x+\delta) \\ &\approx x^\top x + 2x^\top \delta \\ &= n(x) + 2x^\top \delta.\end{aligned}$$

So $dn_x = 2x^\top$.

2. Derivatives on Manifolds

The story from the previous section generalizes to manifolds, as long as we make the distinction that perturbations and manifold points no longer have the same type. On manifolds, perturbations of a point $p \in M$ are vectors in the tangent space at p. So, if we have a function between manifolds $f : M \to N$, then its derivative at p is a linear mapping between tangent spaces:

$$df_p : T_p M \to T_{f(p)} N.$$

If we have some method $\oplus$ of mapping perturbations to manifold points on each of M and N (for instance an exponential map), then the derivative satisfies the same "first-order approximation property":

$$f(p \oplus X_p) \approx f(p) \oplus df_p X_p.$$

where $X_p \in T_p M$. Once again, it measures how much $f(p)$ is perturbed when you perturb p.

3. Derivatives on Lie Groups

For Lie groups, the story is simpler than the story for general manifolds, since a Lie group has the "universal tangent space" (the tangent space at the identity) that can be used at each point in the group. Recall that a transformation $g \in G$ admits both left and right perturbations by a tangent vector $X \in T_e G$:

$$\begin{aligned}X \oplus g &= \exp(X) g \\ g \oplus X &= g \exp(X)\end{aligned}$$

Thus, for a function $\phi : G \to H$ (where G and H are both Lie groups), we have four different types of derivative, each being a linear mapping from $T_e G$ to $T_e G$:

- Mapping left-perturbations of the input to left-perturbations of the output:

$$\phi(X \oplus g) \approx Y \oplus \phi(g).$$

- Mapping right-perturbations of the input to right-perturbations of the output:

$$\phi(g \oplus X) \approx \phi(g) \oplus Y.$$

- Mapping left-perturbations of the input to right-perturbations of the output:

$$\phi(X \oplus g) \approx \phi(g) \oplus Y.$$

- Mapping right-perturbations of the input to left-perturbations of the output:

$$\phi(g \oplus X) \approx Y \oplus \phi(g).$$

Clearly, devising a notation that distinguishes between these types of derivatives becomes very cumbersome (and might look something like $d_g^{L,R}\phi$ to indicate "derivative mapping left-perturbations to right-perturbations" at g of ϕ"). Fortunately, using the frame-based notation makes everything clearer.

From the perspective of frames, derivatives are taken with respect to frames, rather than with respect to group elements, and the derivative simply describes how a perturbation in a given frame affects another frame. For example, consider the product of two group elements: $(g, h) \mapsto gh$. One might want to know how perturbations of g and h affect the product. So say that $g = (\mathfrak{c} \leftarrow \mathfrak{b})$ and $h = (\mathfrak{b} \leftarrow \mathfrak{a})$. Then $gh = (\mathfrak{c} \leftarrow \mathfrak{a})$ and we can ask "what is the derivative of $(\mathfrak{c} \leftarrow \mathfrak{a})$ with respect to either $\mathfrak{a}$, $\mathfrak{b}$, or $\mathfrak{c}$? And we can express the result in either frame $\mathfrak{c}$ or in frame $\mathfrak{a}$. Hence to specify a derivative (as a mapping from $T_eG \to T_eG$) we have specify

- Which frame we apply the perturbation in.
- Which frame the output of the derivative mapping is expressed in.

So what happens to $(\mathfrak{c} \leftarrow \mathfrak{a})$ under a perturbation of each frame?

- Perturbing $\mathfrak{c}$:

$$\begin{aligned} X^{\mathfrak{c}} \oplus (\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) &= X^{\mathfrak{c}} \oplus (\mathfrak{c} \leftarrow \mathfrak{a}) \\ &= (\mathfrak{c} \leftarrow \mathfrak{a}) \oplus \mathrm{Ad}_{\mathfrak{a} \leftarrow \mathfrak{c}} X^{\mathfrak{c}}. \end{aligned}$$

- Perturbing $\mathfrak{a}$:

$$\begin{aligned} (\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) \oplus X^{\mathfrak{a}} &= (\mathfrak{c} \leftarrow \mathfrak{a}) \oplus X^{\mathfrak{a}} \\ &= \mathrm{Ad}_{\mathfrak{c} \leftarrow \mathfrak{a}} X^{\mathfrak{a}} \oplus (\mathfrak{c} \leftarrow \mathfrak{a}). \end{aligned}$$

- Perturbing $\mathfrak{b}$:

$$\begin{aligned} (\mathfrak{c} \leftarrow \mathfrak{b}) \oplus X^{\mathfrak{b}}(\mathfrak{b} \leftarrow \mathfrak{a}) = (\mathfrak{c} \leftarrow \mathfrak{b}) X^{\mathfrak{b}} \oplus (\mathfrak{b} \leftarrow \mathfrak{a}) &= (\mathfrak{c} \leftarrow \mathfrak{b}) \exp(X^{\mathfrak{b}})(\mathfrak{b} \leftarrow \mathfrak{a}) \\ &= \mathrm{Ad}_{\mathfrak{c} \leftarrow \mathfrak{b}} X^{\mathfrak{b}} \oplus (\mathfrak{c} \leftarrow \mathfrak{a}) \\ &= (\mathfrak{c} \leftarrow \mathfrak{a}) \oplus \mathrm{Ad}_{\mathfrak{a} \leftarrow \mathfrak{b}} X^{\mathfrak{b}}. \end{aligned}$$

So, say that $D^{\mathfrak{c}}_{\mathfrak{a}}$ is the derivative whose input perturbation is expressed in $\mathfrak{a}$ and whose output is expressed in $\mathfrak{c}$ (etc). Then

$$\begin{aligned}
D^{\mathfrak{c}}_{\mathfrak{c}}(\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) &= \mathrm{Id}\\
D^{\mathfrak{a}}_{\mathfrak{c}}(\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) &= \mathrm{Ad}_{\mathfrak{a}\leftarrow\mathfrak{c}}\\
D^{\mathfrak{c}}_{\mathfrak{a}}(\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) &= \mathrm{Ad}_{\mathfrak{c}\leftarrow\mathfrak{a}}\\
D^{\mathfrak{a}}_{\mathfrak{a}}(\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) &= \mathrm{Id}\\
D^{\mathfrak{c}}_{\mathfrak{b}}(\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) &= \mathrm{Ad}_{\mathfrak{c}\leftarrow\mathfrak{b}}\\
D^{\mathfrak{a}}_{\mathfrak{b}}(\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) &= \mathrm{Ad}_{\mathfrak{a}\leftarrow\mathfrak{b}}
\end{aligned}$$

Notice the general rule

$$D^{\mathfrak{g}}_{\mathfrak{f}}(\mathfrak{c} \leftarrow \mathfrak{b})(\mathfrak{b} \leftarrow \mathfrak{a}) = \mathrm{Ad}_{\mathfrak{g}\leftarrow\mathfrak{f}},$$

as long as $\mathfrak{f}, \mathfrak{g} \in \{\mathfrak{a}, \mathfrak{b}, \mathfrak{c}\}$.