A APPENDIX

A.1 Clarke Jacobian

One might be tempted to think that the Clarke Jacobian simply just lifts all the rules of the ordinary Jacobian to operate on convex sets, but this is not the case. There are key cases where the Clarke Jacobian does *not* satisfy the analogous rule for ordinary Jacobians. For instance, the notion of partial Clarke Jacobians (denoted as ∂_c^i when taken with respect to variable x_i) does not always obey an equality or containment relation, meaning in general, they are incomparable:

$$\partial_c(f,(x_1,x_2)) \not\gtreqless \partial_c^1(f,(x_1,x_2)) \times \partial_c^2(f,(x_1,x_2))$$

Consequently, as pointed out by Khan and Barton [2013], this means that computing the Clarke Jacobian for multi-variate functions is not as simple as just concatenating the partial Clarke Jacobians ∂_c^i into a matrix (as is done with the classical Jacobian). Likewise, a chain rule for the Clarke Jacobian only holds in select cases. Because of this, our language only uses expressions where the Clarke Jacobian *does* always obey an explicit, closed-form rule analogous to the ordinary Jacobian, which ensures that the analysis is fully compositional, *by construction*. Hence, by restricting to these primitives, we do not need to compute gradients of arbitrary composite functions in situations where the classical Jacobian would require differentiating with respect to a single variable while holding all others as constants (since the Clarke Jacobian does not obey this rule).

However, as restrictive as this sounds, this still allows for us to have a very expressive set of language primitives (that goes beyond existing work [Edalat and Maleki 2017, 2018; Mangal et al. 2020]) where the semantics can be defined precisely.

A.2 Soundness

To prove soundness, we first state the following corollaries that follow from Definition 5.3.

Corollary A.1. *γ is monotonic with respect to $\subseteq$.*

Corollary A.2. *For any $c \in \mathbb{R}$ and $S \in \mathbb{IR}^{n\times m}$, $c \cdot \gamma(S) = \gamma(cS)$ and $\gamma(S^T) = \gamma(S)^T$.*

Corollary A.3. *For any $S_1, S_2 \in \mathbb{IR}^{n\times m}$, $\gamma(S_1 \times S_2) = \gamma(S_1) \times \gamma(S_2)$. Likewise, $\mathbf{co}(\gamma(S_1), \gamma(S_2)) \subseteq \gamma(S_1 \sqcup S_2)$, $\gamma(S_1) +_c \gamma(S_2) = \gamma(S_1 +_{\mathbb{IR}} S_2)$ and $\gamma(S_1) \cdot_c \gamma(S_2) = \gamma(S_1 \cdot_{\mathbb{IR}} S_2)$.*

The proof of soundness is now given case-wise for each primitive in the language (where each case is given as a separate theorem).

Constants. Given that our language is inductively defined, we first prove the soundness of the interval domain over-approximation of the Clarke Jacobian for the base cases, the first one being for constant functions $f = c$ for any $c \in \mathbb{R}$.

Theorem A.4. *(Soundness of ∂^{Int} for Constants) Let $c \in \mathbb{R}$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$.*

$$\partial^{Int}(c, \widehat{\mathbf{x_0}}) = \widehat{\mathbf{0}} = [0,0] \times \ldots \times [0,0]$$

Further, by the rules for ∂_c, we know explicitly that

$$\partial_c(c, \mathbf{x_0}) = \mathbf{0} = \{(0, \ldots, 0)\}$$

Thus

$$\gamma(\partial^{Int}(c, \widehat{\mathbf{x_0}})) = \gamma([0,0] \times \ldots \times [0,0]) = \{(0, \ldots, 0)\} = \partial_c(c, \mathbf{x_0})$$

Hence, because of the equality, the weaker statement trivially follows:

$$\partial_c(c, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(c, \widehat{\mathbf{x_0}}))$$

Variables. The next base case for which we must prove the soundness of the interval domain over-approximation of the Clarke Jacobian is for functions of a single variable $f = x_i$.

Theorem A.5. *(Soundness of ∂^{Int} for Variables) Let $f = x_i$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$.*

$$\partial^{Int}(x_i, \widehat{\mathbf{x_0}}) = \widehat{\mathbf{e}}_i = [0,0] \times ... \times [0,0] \times [1,1] \times [0,0]...$$

$$\partial_c(x_i, \mathbf{x_0}) = \mathbf{e}_i = \{(0, ...0, 1, 0...)\}$$

Hence

$$\gamma(\partial^{Int}(c, \widehat{\mathbf{x_0}})) = \gamma([0,0] \times ... \times [0,0] \times [1,1] \times [0,0]...) = \{(0, ...0, 1, 0...)\} = \partial_c(x_i, \mathbf{x_0})$$

Again, because of the equality, the weaker statement trivially follows:

$$\partial_c(x_i, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(x_i, \widehat{\mathbf{x_0}}))$$

Addition. Having proven the soundness of the interval domain over-approximation of the Clarke Jacobian for the base cases of constant functions and functions of a single variable, we can now inductively prove the soundness of ∂^{Int} for the sum of functions.

Theorem A.6. *(Soundness of ∂^{Int} for Addition of Functions (Linearity)) Let $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$. By the inductive assumption,*

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$

$$\partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Thus

$$\partial_c(f_1, \mathbf{x_0}) +_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_c \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

However, when the arguments are interval matrices, $+_c$ and $+_{\mathbb{IR}}$ coincide exactly. Then, since $\gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$ and $\gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$ are interval matrices, we have

$$\partial_c(f_1, \mathbf{x_0}) +_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_{\mathbb{IR}} \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Since γ factors over $+_{\mathbb{IR}}$, we get:

$$\partial_c(f_1, \mathbf{x_0}) +_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}) +_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

By the definition of $\partial^{Int}(f_1 + f_2, \widehat{\mathbf{x_0}})$, we get:

$$\partial_c(f_1, \mathbf{x_0}) +_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1 + f_2, \widehat{\mathbf{x_0}}))$$

Lastly, since $\partial_c(f_1 + f_2, \mathbf{x_0}) \subseteq \partial_c(f_1, \mathbf{x_0}) +_c \partial_c(f_2, \mathbf{x_0})$, we get:

$$\partial_c(f_1 + f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1 + f_2, \widehat{\mathbf{x_0}}))$$

Multiplication. Proving the soundness of the interval domain over-approximation of the Clarke Jacobian for the multiplication of two functions (i.e., ensuring the correctness of the interval domain-lifted product rule) also requires the inductive assumption.

Theorem A.7. *(Soundness of ∂^{Int} for Multiplication of Functions (Lifted Product Rule)) Let $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$. By the inductive assumption,*

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$

$$\partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Furthermore, $f_1(\mathbf{x_0})$ and $f_2(\mathbf{x_0})$ are just vectors in $\mathbb{R}^n$, hence

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) \subseteq f_2(\mathbf{x_0}) \cdot_c \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$

$$f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq f_1(\mathbf{x_0}) \cdot_c \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Because γ distributes over scalar multiplication,

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(f_2(\mathbf{x_0}) \cdot_c \partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$

$$f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(f_1(\mathbf{x_0}) \cdot_c \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Thus we now have

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) +_c f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(f_2(\mathbf{x_0}) \cdot_c \partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_c \gamma(f_1(\mathbf{x_0}) \cdot_c \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Again, when the arguments are interval matrices, $+_c$ and $+_{\mathbb{IR}}$ coincide exactly, so because $\gamma(f_2(\mathbf{x_0}) \cdot_c \partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$ and $\gamma(f_1(\mathbf{x_0}) \cdot_c \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$ are interval matrices, we have

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) +_c f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(f_2(\mathbf{x_0}) \cdot_c \partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_{\mathbb{IR}} \gamma(f_1(\mathbf{x_0}) \cdot_c \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Likewise, since $f_1(\mathbf{x_0})$ and $f_2(\mathbf{x_0})$ are vectors, $\cdot_c$ and $\cdot_{\mathbb{IR}}$ coincide giving

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) +_c f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(f_2(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_{\mathbb{IR}} \gamma(f_1(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Furthermore, by the soundness of interval arithmetic,

$$f_1(\mathbf{x_0}) \in f_1(\widehat{\mathbf{x_0}})$$

$$f_2(\mathbf{x_0}) \in f_2(\widehat{\mathbf{x_0}})$$

Hence

$$f_2(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) \subseteq f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})$$

$$f_1(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}) \subseteq f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}})$$

By the monotonicity of γ,

$$\gamma(f_2(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})) \subseteq \gamma(f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$

$$\gamma(f_1(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}})) \subseteq \gamma(f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Hence

$$\gamma(f_2(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_{\mathbb{IR}} \gamma(f_1(\mathbf{x_0}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$
$$\subseteq \gamma(f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_{\mathbb{IR}} \gamma(f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Therefore

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) +_c f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})) +_{\mathbb{IR}} \gamma(f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Since γ distributes over $+_{\mathbb{IR}}$, we get

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) +_c f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) +_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

By the definition of $\partial^{Int}(f_1 \cdot f_2, \widehat{\mathbf{x_0}})$, we have

$$f_2(\mathbf{x_0}) \cdot_c \partial_c(f_1, \mathbf{x_0}) +_c f_1(\mathbf{x_0}) \cdot_c \partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1 \cdot f_2, \widehat{\mathbf{x_0}}))$$

Lastly, by the definition of $\partial_c(f_1 \cdot f_2, \widehat{\mathbf{x_0}})$, we get

$$\partial_c(f_1 \cdot f_2, \widehat{\mathbf{x_0}}) \subseteq \gamma(\partial^{Int}(f_1 \cdot f_2, \widehat{\mathbf{x_0}}))$$

Division. We now detail the soundness of the interval domain over-approximation of the Clarke Jacobian for the division of two functions, thus ensuring the correctness of the lifted quotient rule. This also requires the inductive assumption.

Theorem A.8. *(Soundness of ∂^{Int} for Division of Functions (Lifted Quotient Rule)) Let $f_1 : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$. By the inductive assumption,*

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$

Since $-1/f_1(\mathbf{x_0})^2$ is a constant, $/_c$ simply scales each element of both sets by this constant, which still preserves the inequality

$$-\partial_c(f_1, \mathbf{x_0})/_c f_1(\mathbf{x_0})^2 \subseteq -\gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))/_c f_1(\mathbf{x_0})^2$$

Since γ distributes over scaling,

$$-\partial_c(f_1, \mathbf{x_0})/_c f_1(\mathbf{x_0})^2 \subseteq \gamma(-\partial^{Int}(f_1, \widehat{\mathbf{x_0}})/_{\mathbb{IR}} f_1(\mathbf{x_0}))^2$$

Then, by the soundness of interval arithmetic,

$$f_1(\mathbf{x_0})^2 \in f_1(\widehat{\mathbf{x_0}})^2$$

Hence

$$-\partial^{Int}(f_1, \widehat{\mathbf{x_0}})/_{\mathbb{IR}} f_1(\mathbf{x_0})^2 \subseteq -\partial^{Int}(f_1, \widehat{\mathbf{x_0}})/_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}})^2$$

Thus by the monotonicity of γ, we have

$$-\partial_c(f_1, \mathbf{x_0})/_c f_1(\mathbf{x_0})^2 \subseteq \gamma(-\partial^{Int}(f_1, \widehat{\mathbf{x_0}})/_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}})^2)$$

Which by definition

$$\partial_c(1/f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(1/f_1, \widehat{\mathbf{x_0}}))$$

Composition. We now detail the soundness of the interval domain over-approximation of the Clarke Jacobian for compositions with C^1 functions, thus ensuring the correctness of the lifted chain rule.

Theorem A.9. *(Soundness of ∂^{Int} for Compositions with C^1 functions (Lifted Chain Rule)) Let $f_1 : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$. By the inductive assumption,*

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma\big(\partial^{Int}(f_1, \widehat{\mathbf{x_0}})\big)$$

Since each C^1 function maps $\mathbb{R} \to \mathbb{R}$, this means that $\mathbf{J}(C^1, f_1(\mathbf{x_0}))$ is a constant, hence we can scale both sides by this constant and preserve the $\subseteq$ inequality:

$$\mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \partial_c(f_1, \mathbf{x_0}) \subseteq \mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \gamma\big(\partial^{Int}(f_1, \widehat{\mathbf{x_0}})\big)$$

Since $\mathbf{J}(C^1, f_1(\mathbf{x_0}))$ is a constant, we can move it inside γ, giving

$$\mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \partial_c(f_1, \mathbf{x_0}) \subseteq \gamma\big(\mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \partial^{Int}(f_1, \widehat{\mathbf{x_0}})\big)$$

By the soundness of the interval arithmetic,

$$\mathbf{J}(C^1, f_1(\mathbf{x_0})) \in \mathbf{J}^{Int}(C^1, f_1(\widehat{\mathbf{x_0}}))$$

Hence

$$\mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) \subseteq \mathbf{J}^{Int}(C^1, f_1(\widehat{\mathbf{x_0}})) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})$$

Then, by monotonicity of γ, we have

$$\mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \partial_c(f_1, \mathbf{x_0}) \subseteq \gamma\big(\mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \partial^{Int}(f_1, \widehat{\mathbf{x_0}})\big) \subseteq \gamma\big(\mathbf{J}^{Int}(C^1, f_1(\widehat{\mathbf{x_0}})) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})\big)$$

By definition of $\partial_c(C^1 \circ f_1, \mathbf{x_0})$, we have

$$\partial_c(C^1 \circ f_1, \mathbf{x_0}) \subseteq \mathbf{J}(C^1, f_1(\mathbf{x_0})) \cdot \partial_c(f_1, \mathbf{x_0}) \subseteq \gamma\big(\mathbf{J}^{Int}(C^1, f_1(\widehat{\mathbf{x_0}})) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})\big)$$

Lastly, by the definition of $\partial^{Int}(C^1 \circ f_1, \widehat{\mathbf{x_0}})$, we obtain

$$\partial_c(C^1 \circ f_1, \mathbf{x_0}) \subseteq \gamma\big(\mathbf{J}^{Int}(C^1, f_1(\widehat{\mathbf{x_0}})) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})\big) = \gamma\big(\partial^{Int}(C^1 \circ f_1, \widehat{\mathbf{x_0}})\big)$$

Cartesian Product. We now describe the soundness of the interval domain over-approximation of the Clarke Jacobian for the Cartesian product of two functions, which also requires the inductive assumption.

Theorem A.10. *(Soundness of ∂^{Int} for Cartesian Products) Let $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$. By the inductive assumption,*

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$
$$\partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

Starting from the definition of $\partial_c(f_1 \times f_2, \mathbf{x_0})$ gives:

$$\partial_c(f_1 \times f_2, \mathbf{x_0}) \subseteq \begin{bmatrix} \partial_c(f_1, \mathbf{x_0}) \\ \partial_c(f_2, \mathbf{x_0}) \end{bmatrix}$$

By the inductive assumption,

$$\begin{bmatrix} \partial_c(f_1, \mathbf{x_0}) \\ \partial_c(f_2, \mathbf{x_0}) \end{bmatrix} \subseteq \begin{bmatrix} \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}})) \\ \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}})) \end{bmatrix}$$

However, γ commutes over blocks in a block matrix, hence

$$\begin{bmatrix} \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}})) \\ \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}})) \end{bmatrix} = \gamma\left(\begin{bmatrix} \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) \\ \partial^{Int}(f_2, \widehat{\mathbf{x_0}}) \end{bmatrix} \right) = \gamma(\partial^{Int}(f_1 \times f_2, \widehat{\mathbf{x_0}}))$$

Branching. Finally, we describe the soundness of the interval domain over-approximation of the Clarke Jacobian for the branching primitive.

Theorem A.11. *(Soundness of ∂^{Int} for Branching) Let $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{x_0} \in \mathbb{R}^m$, $\widehat{\mathbf{x_0}} \in \mathbb{IR}^m$ with $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$.*

Case 1. *If $f_0(\mathbf{x_0}) > c$, then*

$$\partial_c(f_0 > c ?\, f_1 : f_2, \mathbf{x_0}) = \partial_c(f_1, \mathbf{x_0})$$

and either $lb(f_0(\widehat{\mathbf{x_0}})) > c$ or $c \in f_0(\widehat{\mathbf{x_0}})$. If $lb(f_0(\widehat{\mathbf{x_0}})) > c$, then

$$\partial^{Int}(f_0 > c ?\, f_1 : f_2, \widehat{\mathbf{x_0}}) = \partial^{Int}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}))$$

Thus all that remains to be shown is that $\mathbf{x_0} \in [\![f_0 > c]\!](\widehat{\mathbf{x_0}})$; we can invoke the inductive assumption. If $f_0 \neq x_i$, then $[\![f_0 > c]\!](\widehat{\mathbf{x_0}}) = \widehat{\mathbf{x_0}}$, and by assumption $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$, thus trivially,

$$\mathbf{x_0} \in [\![f_0 > c]\!](\widehat{\mathbf{x_0}})$$

If $f_0 = x_i$, then $f_0(\mathbf{x_0}) > c \implies \mathbf{x_0}[i] > c$, thus

$$\mathbf{x_0} \in [-\infty, \infty] \times ... \times [c, \infty] \times ... \times [-\infty, \infty]$$

But by assumption $\mathbf{x_0} \in \widehat{\mathbf{x_0}}$, thus by the definition of $[\![x_i > c]\!]$,

$$\mathbf{x_0} \in \widehat{\mathbf{x_0}} \cap [-\infty, \infty] \times ... \times [c, \infty] \times ... \times [-\infty, \infty]$$
$$\implies \mathbf{x_0} \in [\![f_0 > c]\!](\widehat{\mathbf{x_0}})$$

Thus by the inductive assumption,

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}})))$$

Hence, given the assumptions that $f_0(\mathbf{x_0}) > c \wedge lb(f_0(\widehat{\mathbf{x_0}})) > c$,

$$\partial_c(f_0 > c ?\, f_1 : f_2, \mathbf{x_0}) = \partial_c(f_1, \mathbf{x_0}) \subseteq$$

$$\gamma(\partial^{Int}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}))) = \gamma(\partial^{Int}(f_0 > c?\ f_1 : f_2, \widehat{\mathbf{x_0}}))$$

$$f_0(\mathbf{x_0}) > c \wedge lb(f_0(\mathbf{x_0})) > c \implies \partial_c(f_0 > c?\ f_1 : f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_0 > c?\ f_1 : f_2, \widehat{\mathbf{x_0}}))$$

Now, assuming $c \in f_0(\widehat{\mathbf{x_0}})$, then

$$\partial^{Int}(f_0 > c?\ f_1 : f_2, \widehat{\mathbf{x_0}}) = \partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})) \sqcup \partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}))$$

However, by the same logic as above, regardless of if $f_0 = x_i$ or $f_0 \neq x_i$, since $f_0(\mathbf{x_0}) > c$,

$$\implies \mathbf{x_0} \in [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})$$

Again invoking the inductive assumption,

$$\implies \partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})))$$

However, by the monotonocity of γ,

$$\gamma(\partial^{Int}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}))) \subseteq \gamma(\partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})) \sqcup \partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}})))$$

which means

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}))) \subseteq \gamma(\partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})) \sqcup \partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}))$$

Thus substituting in, gives:

$$\partial_c(f_0 > c?\ f_1 : f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_0 > c?\ f_1 : f_2, \widehat{\mathbf{x_0}}))$$

Case 2. *If $f_0(\mathbf{x_0}) \geq c$, the proof proceeds exactly the same albeit replacing $>$ with $\geq$.*

Case 3. *If $f_0(\mathbf{x_0}) < c$, then $-f_0(\mathbf{x_0}) \geq -c$, which would be covered by the previous case.*

Case 4. *Likewise, if $f_0(\mathbf{x_0}) \leq c$, then $-f_0(\mathbf{x_0}) > -c$, which would be covered by the first case.*

Case 5. *If $f_0(\mathbf{x_0}) = c$, then*

$$\partial_c(f_0 > c?\ f_1 : f_2, \mathbf{x_0}) = \mathbf{co}(\partial_c(f_1, \mathbf{x_0}), \partial_c(f_2, \mathbf{x_0}))$$

which also implies $c \in f_0(\widehat{\mathbf{x_0}})$. Thus

$$\partial^{Int}(f_0 > c?\ f_1 : f_2, \widehat{\mathbf{x_0}}) = \partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})) \sqcup \partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}))$$

If $f_0 \neq x_i$, then

$$\partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})) \sqcup \partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}})) = \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) \sqcup \partial^{Int}(f_2, \widehat{\mathbf{x_0}})$$

But by the inductive assumption ,

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, \widehat{\mathbf{x_0}}))$$
$$\partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_2, \widehat{\mathbf{x_0}}))$$

If $f_0 = x_i$, then $\mathbf{x_0}[i] = c$ and

$$[\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) = \widehat{\mathbf{x_0}} \cap [-\infty, \infty] \times ... \times [-\infty, c] \times ... \times [-\infty, \infty]$$
$$[\![f_0 > c]\!](\widehat{\mathbf{x_0}}) = \widehat{\mathbf{x_0}} \cap [-\infty, \infty] \times ... \times [c, \infty] \times ... \times [-\infty, \infty]$$

However, because both interval vectors contain c in their i^{th} component interval,

$$\mathbf{x_0} \in [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}})$$
$$\mathbf{x_0} \in [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})$$

Thus we can again invoke the inductive assumption:

$$\partial_c(f_1, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}))$$
$$\partial_c(f_2, \mathbf{x_0}) \subseteq \gamma(\partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}))$$

Because **co** *is monotonic,*

$$\mathbf{co}(\partial_c(f_1,\mathbf{x_0}),\partial_c(f_2,\mathbf{x_0}))\subseteq\mathbf{co}(\gamma(\partial^{Int}(f_1,[\![f_0\geq c]\!](\widehat{\mathbf{x_0}})),\gamma(\partial^{Int}(f_2,[\![f_0\leq c]\!](\widehat{\mathbf{x_0}})))$$

By Corollary A.3

$$\mathbf{co}(\gamma(\partial^{Int}(f_1,[\![f_0\geq c]\!](\widehat{\mathbf{x_0}})),\gamma(\partial^{Int}(f_2,[\![f_0\leq c]\!](\widehat{\mathbf{x_0}}))\subseteq\gamma(\partial^{Int}(f_1,[\![f_0\geq c]\!](\widehat{\mathbf{x_0}}))\sqcup\partial^{Int}(f_2,[\![f_0\leq c]\!](\widehat{\mathbf{x_0}})))$$

Hence by transitivity:

$$\mathbf{co}(\partial_c(f_1,\mathbf{x_0}),\partial_c(f_2,\mathbf{x_0}))\subseteq\gamma(\partial^{Int}(f_1,[\![f_0\geq c]\!](\widehat{\mathbf{x_0}}))\sqcup\partial^{Int}(f_2,[\![f_0\leq c]\!](\widehat{\mathbf{x_0}})))$$

Substituting in the definitions:

$$\partial_c(f_0>c?\ f_1:f_2,\mathbf{x_0})\subseteq\gamma(\partial^{Int}(f_0>c?\ f_1:f_2,\widehat{\mathbf{x_0}}))$$

Thus in all cases (when $f_0(\mathbf{x_0})$ is greater than, less than, or equal to c) and sub-cases (when f_0 is either the trivial function x_i or not), the following result holds:

$$\partial_c(f_0>c?\ f_1:f_2,\mathbf{x_0})\subseteq\gamma(\partial^{Int}(f_0>c?\ f_1:f_2,\widehat{\mathbf{x_0}}))$$

A.3 Obtaining a Lipschitz Constant

Having proven that ∂^{Int} soundly over-approximates ∂_c, we can now use ∂^{Int} to bound the Lipschitz constant. As ∂^{Int} evaluates to an interval matrix, we first define the norm of interval matrices.

Definition A.12. (Operator Norm of Interval Matrices) Let $\|\cdot\|_{\alpha,\beta}$ be an induced operator norm for $\mathbb{R}^{n\times m}$. For an interval matrix $M\in\mathbb{IR}^{n\times m}$, we define $\widehat{\|M\|}_{\alpha,\beta}$ as

$$\widehat{\|M\|}_{\alpha,\beta}=sup_{m\in M}\|m\|_{\alpha,\beta}$$

Preliminary: Obtaining Lipschitz Constants. As with the standard Jacobian, one can recover the Lipschitz constant from the Clarke Generalized Jacobian, thanks to a result by Jordan et al. [Jordan and Dimakis 2020].

Theorem A.13. *(Jordan et al. Theorem 1 [Jordan and Dimakis 2020]) Let $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ be arbitary convex norms over $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and let $f:\mathbb{R}^m\to\mathbb{R}^n$ be Lipschitz continuous over an open set $O\subseteq\mathbb{R}^m$. The following equality holds:*

$$K^{(\alpha,\beta)}(f,O)=sup_{G\in\partial_c(f,O)}\|G^T\|_{\alpha,\beta}$$

where $\partial_c(f,O)=\{G\in\partial_c(f,x)\mid x\in O\}$ and $\|M\|_{\alpha,\beta}=sup_{\|v\|_\alpha\leq 1}\|Mv\|_\beta$.

Thankfully, for certain α,β, we can compute this value exactly, as we detail below.

Theorem A.14. *(Definition 7.1 [Moore et al. [n.d.]], presented using our notation) For an interval matrix $M\in\mathbb{IR}^{n\times m}$,*

$$\widehat{\|M\|}_{1,1}=\max_{1\leq j\leq m}\left(\sum_{i=1}^{n}\max\left(|lb(M_{i,j})|,|ub(M_{i,j})|\right)\right)$$

$$\widehat{\|M\|}_{\infty,\infty}=\max_{1\leq i\leq n}\left(\sum_{j=1}^{m}\max\left(|lb(M_{i,j})|,|ub(M_{i,j})|\right)\right)$$

We can now relate this to the Lipschitz constant by combining it with Theorem A.13.

Theorem A.15. *For a function $f:\mathbb{R}^m\to\mathbb{R}^n$ and an open set $\mathcal{X}\subseteq\mathbb{R}^m$ such that $\mathcal{X}\subseteq X$ where $X\in\mathbb{IR}^m$, the local Lipschitz constant on $\mathcal{X}$ satisfies*

$$K^{\alpha,\beta}\leq\widehat{\|\partial^{Int}(f,X)\|}_{\alpha,\beta}$$

Proof. By Theorem A.13, we know that

$$K^{\alpha,\beta} = sup_{G \in \partial_c(f,\mathcal{X})} \|G\|_{\alpha,\beta}$$

But by the soundness of the over-approximation (Thm. 5.4), we know that $\partial_c(f,\mathcal{X}) \subseteq \gamma(\partial^{Int}(f,X))$. But $\gamma(\partial^{Int}(f,X)) = \partial^{Int}(f,X)$, hence

$$sup_{G \in \partial_c(f,\mathcal{X})} \|G\|_{\alpha,\beta} \leq sup_{G \in \partial^{Int}(f,X)} \|G\|_{\alpha,\beta}$$

And the RHS is just the definition of $\|\partial^{\widehat{Int}(f,X)}\|_{\alpha,\beta}$. □

A.4 Equivalence of $Eval_{\mathbb{ID}}$ and ∂^{Int}

We now give the proof that forward mode dual interval evaluation, $Eval_{\mathbb{ID}}$, yields the same answer as the recursively defined ∂^{Int}, thus giving a full proof of Theorem 6.3.

Proof. We start with the base cases.

Case 1. ***Base Cases - Constants.*** *The first base case is when $f = c$. We start with*

$$\partial^{Int}(c, \widehat{\mathbf{x_0}}) = \widehat{\mathbf{0}}$$

But $Eval_{\mathbb{ID}}(c, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon)^T = [c, c] + [0, 0]\epsilon$ *for any* $i \in \{1, .., m\}$, *thus*

$$snd\Big(Eval_{\mathbb{ID}}(c, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(c, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big) = [0, 0] \times ... \times [0, 0]$$

But by our notation, $\widehat{\mathbf{0}} \triangleq [0, 0] \times ... \times [0, 0]$. *Hence*

$$\partial^{Int}(c, \widehat{\mathbf{x_0}}) = \widehat{\mathbf{0}} = snd\Big(Eval_{\mathbb{ID}}(c, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(c, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Case 2. ***Base Cases - Variables.*** *We next analyze the base case where $f = x_i$. We start with*

$$\partial^{Int}(x_i, \widehat{\mathbf{x_0}}) = \widehat{\mathbf{e}}_i$$

where $\widehat{\mathbf{e}}_i = [0, 0], \times ... \times [0, 0] \times [1, 1] \times [0, 0]...$ *(the* $[1, 1]$ *term is in the* i^{th} *index). We also know that*

$$Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon) = \widehat{\mathbf{x_0}}[i] + \widehat{\mathbf{e}}_i[i]\epsilon = \widehat{\mathbf{x_0}}[i] + [1, 1]\epsilon$$

Thus

$$snd\big(Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon)\big) = [1, 1] = snd\big(Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon)^T\big)$$

And for any $j \neq i$,

$$Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_j\epsilon) = \widehat{\mathbf{x_0}}[i] + \widehat{\mathbf{e}}_j[i]\epsilon = \widehat{\mathbf{x_0}}[i] + [0, 0]\epsilon$$

Thus

$$snd\big(Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_j\epsilon)\big) = [0, 0] = snd\big(Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_j\epsilon)^T\big)$$

Hence

$$snd\Big(Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(x_i, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big) = [0, 0] \times ...[0, 0] \times [1, 1] \times [0, 0]...$$

$$= \widehat{\mathbf{e}}_i = \partial^{Int}(x_i, \widehat{\mathbf{x_0}})$$

We now proceed to the inductive cases.

Case 3. ***Addition.*** *We require both $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}^n$. We start with*

$$\partial^{Int}(f_1 + f_2, \widehat{\mathbf{x_0}}) = \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) +_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}})$$

By the inductive hypothesis,

$$= snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

$$+_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Distributing the snd then grouping terms element-wise,

$$= \Big(snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big) +_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big)\Big) \times ...$$

$$... \times \Big(snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big) +_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)\Big)$$

By the definition of dual interval arithmetic evaluation of addition,

$$= snd\Big(Eval_{\mathbb{ID}}(f_1 + f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times ...$$

$$... \times snd\Big(Eval_{\mathbb{ID}}(f_1 + f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Undistributing the snd,

$$= snd\Big(Eval_{\mathbb{ID}}(f_1 + f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_1 + f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Case 4. ***Multiplication.*** *We require both* $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}^n$. *We start with*

$$\partial^{Int}(f_1 \cdot f_2, \widehat{\mathbf{x_0}}) = f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) +_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} \partial^{Int}(f_2, \widehat{\mathbf{x_0}})$$

By the inductive hypothesis,

$$= f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

$$+_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Distributing the snd then distributing f_1 *and* f_2 *by the definition of* $\cdot_{\mathbb{IR}}$,

$$= \Big(f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} snd(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T) \times ... \times (f_2(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} snd(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T)\Big)$$

$$+_{\mathbb{IR}} \Big(f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times ... \times f_1(\widehat{\mathbf{x_0}}) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)\Big)$$

Because $f_2(\widehat{\mathbf{x_0}}) = fst\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \mathbf{e_i}\epsilon)\Big)$ *for any* $i \in \{1, ..., m\}$, *and likewise for* $f_1(\widehat{\mathbf{x_0}})$, *we have*

$$= \Big(fst\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times ...$$

$$... \times fst\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)\Big)$$

$$+_{\mathbb{IR}} \Big(fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times ...$$

$$... \times fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)\Big)$$

By the definition of $+_{\mathbb{IR}}$, *we can group terms element-wise:*

$$= \Big(fst\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big)$$

$$+_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big)\Big) \times ...$$

$$... \times \Big(fst\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

$$+_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)\Big)$$

By the definition of $\cdot_{\mathbb{IR}}$,

$$= snd\Big(Eval_{\mathbb{ID}}(f_1 \cdot f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times ... \times snd\Big(Eval_{\mathbb{ID}}(f_1 \cdot f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Undistributing the snd,

$$= snd\Big(Eval_{\mathbb{ID}}(f_1 \cdot f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_1 \cdot f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

CASE 5. ***Division.*** *We require* $f_1 : \mathbb{R}^m \to \mathbb{R}$. *We start from the desired LHS:*

$$snd\Big(Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Distributing the snd,

$$= snd\Big(Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times ... \times snd\Big(Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Since $1/f$ *has one dimensional output,* $Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T$ *is a (dual-interval) scalar. Hence,* $Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon)^T = Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon)$, *and thus we can drop all transposes, giving:*

$$= snd\Big(Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \times ... \times snd\Big(Eval_{\mathbb{ID}}(1/f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)$$

By the definition of $Eval_{\mathbb{ID}}$,

$$= snd\Big([1,1] + [0,0]\epsilon /_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \times ... \times snd\Big([1,1] + [0,0]\epsilon /_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)$$

By the definition of dual interval division,

$$= snd\Big([1,1]/_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_1\epsilon)\Big) - snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_1\epsilon)\Big)/_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_1\epsilon)\Big)^2\epsilon\Big) \times ...$$

$$... \times snd\Big([1,1]/_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_m\epsilon)\Big) - snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_m\epsilon)\Big)/_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_m\epsilon)\Big)^2\epsilon\Big)$$

Applying the outermost snd operator element-wise,

$$= -snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big)/_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big)^2 \times ...$$

$$... \times -snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)/_{\mathbb{IR}} fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)^2$$

However, since $fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon)\Big) = f_1(\widehat{\mathbf{x_0}})$ *for any i, we have:*

$$= -snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big)/_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}})^2 \times ...$$

$$... \times -snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)/_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}})^2$$

We can pull out the $f_1(\widehat{\mathbf{x_0}})^2$ *term since it is a scalar common to each dimension,*

$$= \Big(-snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \times ... \times -snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)\Big)/_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}})^2$$

But by the inductive hypothesis, the numerator just reduces to $-\partial^{Int}(f_1, \widehat{\mathbf{x_0}})$, *hence giving*

$$= -\partial^{Int}(f_1, \widehat{\mathbf{x_0}})/_{\mathbb{IR}} f_1(\widehat{\mathbf{x_0}})^2$$
$$= \partial^{Int}(1/f_1, \widehat{\mathbf{x_0}})$$

CASE 6. ***Branching.***
If $lb(f_0(\widehat{\mathbf{x_0}})) > c$, *then we have*

$$\partial^{Int}(f_0 > c\ ?\ f_1 : f_2, \widehat{\mathbf{x_0}}) = \partial^{Int}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}))$$

However, $lb(f_0(\widehat{\mathbf{x_0}})) > c$ *also means that for any* $i \in \{1, ..., m\}$,

$$Eval_{\mathbb{ID}}(f_0 > c\ ?\ f_1 : f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon) = Eval_{\mathbb{ID}}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)$$

Hence, by the inductive hypothesis:

$$\partial^{Int}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}})) = snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_1, [\![f_0 > c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Thus substituting back in, we have that $lb(f_0(\widehat{\mathbf{x_0}})) > c$ *implies*

$$\partial^{Int}(f_0 > c\,?\,f_1 : f_2, \widehat{\mathbf{x_0}}) = snd\Big(Eval_{\mathbb{ID}}(f_0 > c\,?\,f_1 : f_2, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}_1}\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(f_0 > c\,?\,f_1 : f_2, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

The case where $ub(f_0(\widehat{\mathbf{x_0}})) < c$ *proceeds exactly the same by symmetry.*

The case where $c \in f_0(\widehat{\mathbf{x_0}})$*, then*

$$\partial^{Int}(f_0 > c\ ?\ f_1 : f_2, \widehat{\mathbf{x_0}}) = \partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})) \sqcup \partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}})$$

But if $c \in f_0(\widehat{\mathbf{x_0}})$*, then for any* $i \in \{1, ..., m\}$,

$$Eval_{\mathbb{ID}}(f_0 > c\ ?\ f_1 : f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon) = Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon) \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)$$

Furthermore, the transpose T *distributes over* $\sqcup_{\mathbb{ID}}$*:*

$$\big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon) \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)\big)^T$$
$$= Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)^T \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)^T$$

Thus starting from the desired RHS:

$$snd\Big(Eval_{\mathbb{ID}}(f_0 > c\ ?\ f_1 : f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}_1}\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(f_0 > c\ ?\ f_1 : f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$
$$= snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T \times \ldots$$
$$\ldots \times Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

We distribute snd to each term:

$$= snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times \ldots$$
$$\ldots \times snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

And since $snd(A \sqcup_{\mathbb{ID}} B) = snd(A) \sqcup snd(B)$*, we get (for any* i*):*

$$snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)^T \sqcup_{\mathbb{ID}} Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)^T\Big)$$
$$= snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)^T\Big) \sqcup snd\Big(Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_i\epsilon)^T\Big)$$

Thus applying this to each term gives:

$$= snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \sqcup snd\Big(Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times \ldots$$
$$\ldots \times snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big) \sqcup snd\Big(Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

But this is just $\sqcup$ *applied element-wise, hence we can rewrite it as*

$$= snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times \ldots \times snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$
$$\sqcup$$
$$snd\Big(Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T\Big) \times \ldots \times snd\Big(Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Undistributing the snd gives

$$= snd\Big(Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$
$$\sqcup$$
$$snd\Big(Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_1\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}) + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

By the inductive hypothesis, this is just

$$= \partial^{Int}(f_1, [\![f_0 \geq c]\!](\widehat{\mathbf{x_0}})) \sqcup \partial^{Int}(f_2, [\![f_0 \leq c]\!](\widehat{\mathbf{x_0}}))$$

which is of course just equal to

$$= \partial^{Int}(f_0 > c \,?\, f_1 : f_2, \widehat{\mathbf{x_0}})$$

CASE 7. ***Composition.*** *We require $f_1 : \mathbb{R}^m \to \mathbb{R}$. We start with the desired RHS:*

$$snd\Big(Eval_{\mathbb{ID}}(C^1(f_1), \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(C^1(f_1), \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

By the definition of $Eval_{\mathbb{ID}}$,

$$= snd\Big(Eval_{\mathbb{ID}}(C^1, Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon))^T \times \ldots \times Eval_{\mathbb{ID}}(C^1, Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon))^T\Big)$$

Distributing the snd element-wise:

$$= snd\Big(Eval_{\mathbb{ID}}(C^1, Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon))^T\Big) \times \ldots \times snd\Big(Eval_{\mathbb{ID}}(C^1, Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon))^T\Big)$$

Applying the definition of $Eval_{\mathbb{ID}}(C^1, \cdot)$ and then applying snd:

$$= J^{Int}\Big(C^1, fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \times \ldots$$

$$\ldots \times J^{Int}\Big(C^1, fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)$$

Simplifying each $fst\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_i\epsilon)\Big)$ term:

$$= J^{Int}\Big(C^1, f_1(\widehat{\mathbf{x_0}})\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \times \ldots$$

$$\ldots \times J^{Int}\Big(C^1, f_1(\widehat{\mathbf{x_0}})\Big) \cdot_{\mathbb{IR}} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)$$

Since $J^{Int}\Big(C^1, f_1(\widehat{\mathbf{x_0}})\Big)$ is common to each element, it can be factored out:

$$= J^{Int}\Big(C^1, f_1(\widehat{\mathbf{x_0}})\Big) \cdot_{\mathbb{IR}} \Big(snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)\Big) \times \ldots \times snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)\Big)\Big)$$

But by the inductive assumption, the rightmost term reduces to $\partial^{Int}(f_1, \widehat{\mathbf{x_0}})$, giving:

$$= J^{Int}(C^1, f_1(\widehat{\mathbf{x_0}})) \cdot_{\mathbb{IR}} \partial^{Int}(f_1, \widehat{\mathbf{x_0}})$$

$$= \partial^{Int}(C^1(f_1), \widehat{\mathbf{x_0}})$$

CASE 8. ***Cartesian Product.*** *We start with the desired LHS:*

$$snd\Big(Eval_{\mathbb{ID}}(f_1 \times f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(f_1 \times f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big)$$

Applying the definition of $Eval_{\mathbb{ID}}(f_1 \times f_2, \cdot)$,

$$= snd\Big(\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_1\epsilon) \times Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_1\epsilon)\Big)^T \times \ldots \times \Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_1\epsilon) \times Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}}+\widehat{\mathbf{e}}_m\epsilon)\Big)^T\Big)$$

By properties of transpose T:

$$= snd\Big(\begin{bmatrix} Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \\ Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \end{bmatrix} \times \ldots \times \begin{bmatrix} Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T \\ Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T \end{bmatrix} \Big)$$

By properties of $\times$:

$$= snd\Big(\begin{bmatrix} Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T \\ Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times \ldots \times Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T \end{bmatrix} \Big)$$

Distributing snd,

$$= \begin{bmatrix} snd\Big(Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_1, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big) \\ snd\Big(Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_1\epsilon)^T \times ... \times Eval_{\mathbb{ID}}(f_2, \widehat{\mathbf{x_0}} + \widehat{\mathbf{e}}_m\epsilon)^T\Big) \end{bmatrix}$$

By the inductive assumption:

$$= \begin{bmatrix} \partial^{Int}(f_1, \widehat{\mathbf{x_0}}) \\ \partial^{Int}(f_2, \widehat{\mathbf{x_0}}) \end{bmatrix}$$

$$= \partial^{Int}(f_1 \times f_2, \widehat{\mathbf{x_0}})$$

□

A.5 Evaluation

Detailed Network Architectures. For the local optimization landscape experiment, we trained 7 fully-connected networks on the MNIST dataset, varying the total number of layers from 3-9. Each hidden layer contained 30 neurons, and a ReLU activation was applied after every layer (including the final layer). For the Lipschitz experiments, we trained three network architectures on both the MNIST and CIFAR10 datasets, for a total of six networks. We denote $\text{Conv}(o, k, s, p)$ as a convolutional layer with o output channels, k x k kernel width and height, stride s, and padding p; we denote $\text{Fc}(o)$ as a fully-connected layer that outputs o neurons. $\text{ReLU}(l)$ indicates that a ReLU activation is applied element-wise to the outputs of layer l.

(1) FFNN: Seven-layer fully-connected network with 1024, 512, 256, 128, 64, and 32 hidden neurons and a ReLU activation after every layer (including the final layer).
(2) ConvMed: ReLU(Conv(16, 4, 2, 1)) → ReLU(Conv(32, 4, 2, 1)) → ReLU(Fc(100)) → Fc(10)
(3) ConvBig: ReLU(Conv(32, 3, 1, 1)) → ReLU(Conv(32, 4, 2, 1)) → ReLU(Conv(64, 3, 1, 1)) → ReLU(Conv(64, 4, 2, 1)) → ReLU(Fc(512)) → ReLU(Fc(512)) → Fc(10)

Additional Results. Starting on the next page, we show results for Lipschitz bounds and runtimes on additional networks.

Table 6. Runtimes in seconds to compute Interval Clarke Jacobians for individual perturbations over 100 images. For each network, the six rows represent the times for DeepJ vanilla, 2x, 5x, 9x, 25x, and Baseline.

	Haze			Contrast			Rotation		
	Min	Med	Max	Min	Med	Max	Min	Med	Max
MNIST FFNN	1.7	1.7	1.8	1.7	1.8	2.7	15.5	15.7	117.9
	3.4	3.4	3.5	3.3	3.4	3.5	30.9	31.0	107.3
	8.3	8.6	9.2	8.2	8.2	9.4	36.1	36.6	81.2
	15.0	15.1	15.3	14.8	14.9	93.9	57.3	57.9	96.5
	41.6	68.8	371.4	40.7	57.9	70.8	141.3	208.4	475.7
	0.05	0.06	0.07	0.04	0.06	0.08	13.6	14.0	114.9
MNIST ConvMed	0.6	0.6	0.6	0.6	0.6	0.7	14.9	14.9	120.5
	1.1	1.1	1.2	1.1	1.1	1.2	29.5	29.7	108.8
	2.7	2.7	2.8	2.7	2.7	2.7	31.2	31.3	77.4
	4.7	4.8	4.9	4.7	4.7	4.8	47.4	53.4	652.6
	13.0	13.4	22.3	12.8	13.0	22.0	112.4	150.2	210.4
	0.04	0.05	0.06	0.04	0.05	0.07	13.6	14.0	114.9
CIFAR FFNN	4.3	4.3	4.4	4.3	4.3	4.4	71.2	71.4	704.0
	8.5	8.6	10.2	8.5	8.6	9.1	140.6	142.2	599.6
	21.1	21.2	21.4	21.1	21.3	25.7	154.3	155.9	582.2
	38.4	38.6	38.8	38.4	38.7	198.7	246.8	265.0	470.9
	106.9	158.1	312.5	106.1	156.9	331.8	591.0	853.4	1104.1
	0.06	0.07	0.08	0.07	0.08	0.09	65.5	66.2	685.6

Table 7. Runtimes in seconds to compute Interval Clarke Jacobians for composite perturbations over 10 images. For each network, the six rows represent the times for DeepJ vanilla, 2x, 3x, 5x, 9x, and Baseline.

	Haze-Rotation			Contrast-Rotation			Contrast-Haze		
	Min	Med	Max	Min	Med	Max	Min	Med	Max
MNIST FFNN	9.4	9.7	15.3	9.3	9.4	15.1	1.0	1.1	1.1
	37.4	42.2	48.5	37.2	37.3	48.0	4.0	4.0	4.1
	47.9	48.3	48.8	47.9	48.2	49.2	5.6	5.8	6.1
	58.6	58.6	65.3	58.6	58.8	61.2	8.7	9.0	9.2
	125.9	128.6	129.9	124.6	126.6	133.9	26.5	26.7	27.1
	8.2	8.2	13.9	8.2	8.2	13.8	0.06	0.07	0.08
MNIST ConvMed	9.0	9.5	15.4	9.1	9.4	15.3	0.4	0.5	1.6
	35.5	35.6	46.7	35.5	35.6	46.5	1.3	1.3	1.4
	44.8	45.3	45.7	44.5	44.8	47.7	2.0	2.2	2.5
	53.3	53.5	56.3	53.1	53.2	55.4	2.8	2.9	3.1
	113.5	114.9	119.1	113.2	113.8	118.1	8.2	9.3	9.5
	8.2	8.2	13.9	8.2	8.2	13.8	0.05	0.06	0.07
CIFAR FFNN	42.7	42.7	80.2	42.7	42.8	80.6	2.6	2.6	2.7
	171.5	172.3	239.7	171.4	174.0	240.8	11.6	16.8	49.4
	217.8	218.6	224.6	217.9	218.2	223.9	14.2	14.6	19.5
	262.4	263.0	284.6	262.8	263.2	275.8	22.3	22.4	22.8
	568.8	571.4	600.8	565.5	569.0	606.0	67.6	68.0	73.1
	39.4	39.4	79.1	39.4	39.4	76.1	0.06	0.07	0.08

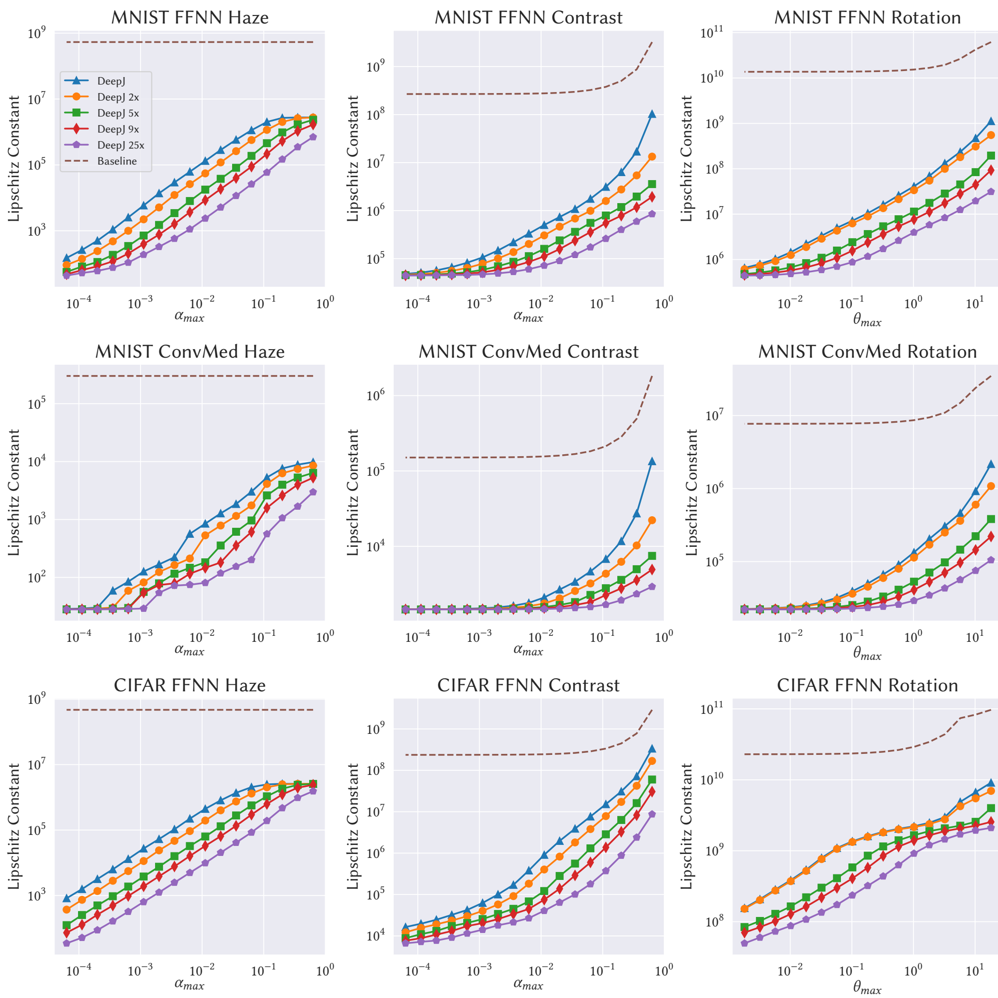

Fig. 10. Upper bounds on the Lipschitz constants with respect to individual perturbations.

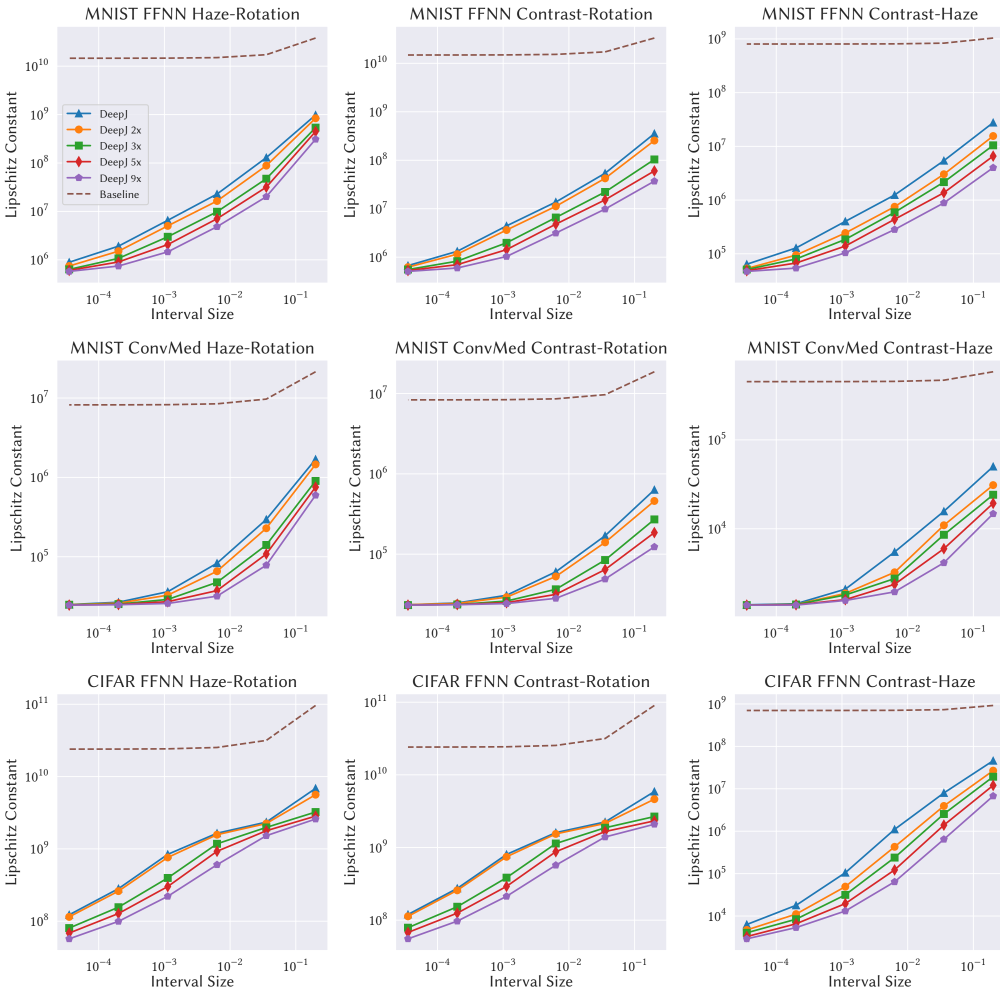

Fig. 11. Upper bounds on the Lipschitz constants with respect to composite perturbations.