

Oracle and Jet Semantics of Stateful Optimization

Anonymous

Version 1.2 - 12 October 2026 - Prepared for editorial review

Author: Anonymous

Abstract

We distinguish four objective interfaces: pointwise finite jets, pointwise infinite jets, smooth germs, and global holonomic jet sections. Every extensional optimizer on smooth objectives admits a global holonomic representation, already at order zero, with runtime and auxiliary data retained. A separate fiber-constancy criterion characterizes descent to pointwise jets. We prove adaptive oracle indistinguishability and a finite-query obstruction to universal exact global minimization of smooth coercive objectives. Counterexamples exhibit flat germs and remote value queries. A coordinate-transport theorem explains why equivalent forward parameterizations need not produce equivalent Euclidean gradient dynamics. Exact rational polynomial examples accompany the interface results. The contribution is a precise organization of optimizer information dependence; no universal local-jet or convergence claim is inferred from the representation theorem.

1. Interfaces and prior work

Classical jet geometry separates differential order from the smooth germ that supplies it. Peetre-Slovak theory identifies finite-order behavior for local operators under specified regularity hypotheses [1]. Categorical gradient learning retains optimizer structure and parameters [2,3]. We combine these distinctions into a small semantics of stateful objective queries. The quotient and oracle proofs below are elementary; their value here is to make the limits of a universal jet description explicit.

Fix an open domain $\Theta \subseteq \mathbb{R}^d$, $d \geq 1$, and write $\mathcal{L} = C^\infty(\Theta, \mathbb{R})$. The objectives and all queried derivatives are ideal mathematical data. An effective implementation supplies only its declared, computable query interface; an infinite jet is not assumed to be a stored infinite array. Unless extra structure is stated, representation and quotient factorizations below are set-theoretic: they do not assert computability, continuity, or an algorithm for reading a global function.

Definition 1.1 (Runtime and auxiliary data). Runtime state Ξ contains current parameters, memory, query history, candidate populations, and random state. Auxiliary data C includes constraints, schedules, precision rules, objective presentations, and oracle contracts. An extensional deterministic step is a possibly partial map

$$A : \mathcal{L} \times \Xi \times C \rightharpoonup Y$$

whose definedness and result depend only on the objective function when runtime and auxiliary data are fixed. Y may contain a next state or a stopping result. Code-inspection behavior belongs in the auxiliary presentation data when it affects the algorithm. Extensionality in the scalar objective does not identify different such data.

Interface hierarchy. At a fixed point, germ equality implies equality of every finite jet and the infinite jet; the converse from infinite jets fails for smooth functions. A global holonomic section retains values at every point. Runtime state and auxiliary data remain additional inputs at every level. Sections 2-3 separate representation from locality; Section 4 states the oracle information bound.

2. Jets, germs, and global sections

Definition 2.1 (Germ and jet). The smooth germ algebra G_θ is the direct limit of $C^\infty(U, \mathbb{R})$ over neighborhoods of θ , with restriction maps. Germs agree if their representatives agree on some neighborhood. Their order- r jets agree if all derivatives of orders at most r agree at θ . Write J_θ^r for the quotient and

$$J_\theta^\infty = \lim_r J_\theta^r.$$

In coordinates, $j_\theta^r L$ records $\partial^\alpha L(\theta)$ for multiindices $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = \sum_i \alpha_i \leq r$. The maps $J^{r+1} \rightarrow J^r$ forget the highest derivatives. When a quotient argument uses infinite jets, its codomain is the image of $G_\theta \rightarrow J_\theta^\infty$, so no realization theorem for arbitrary formal data is needed.

Lemma 2.2 (An infinite jet need not determine a smooth germ). The Taylor map $G_0 \rightarrow J_0^\infty$ is not injective.

Proof. In one variable let $h(x) = e^{-1/x^2}$ for $x > 0$ and $h(x) = 0$ for $x \leq 0$. Each derivative on $x > 0$ is a polynomial in $1/x$ times the exponential. For every N , $x^{-N}e^{-1/x^2} \rightarrow 0$ as $x \downarrow 0$, so the function extends smoothly with every derivative zero at 0. Its germ is nonzero because it is positive on every sufficiently small positive interval. It has the same infinite jet as zero. $\square$

For analytic germs the locally convergent Taylor series determines the germ. The smooth and analytic statements have different hypotheses.

Definition 2.3 (Global holonomic representation). Let $P^r(L)$ be the section $\theta \mapsto j_\theta^r L$ over the entire domain, with r finite or infinite. Define $\text{Hol}^r(\Theta)$ to be precisely the sections arising this way. The zeroth-order projection is

$$\varepsilon_0 : \text{Hol}^r(\Theta) \rightarrow \mathcal{L}, \quad \varepsilon_0(s)(\theta) = s_0(\theta).$$

It satisfies $\varepsilon_0 P^r = \text{id}$ and $P^r \varepsilon_0 = \text{id}$. For infinite sections with smooth coordinate components, holonomicity is equivalent to $\partial_i s_\alpha = s_{\alpha+e_i}$: induction identifies each component with the corresponding derivative of s_0 .

Theorem 2.4 (Global representation). Every step of Definition 1.1 factors uniquely through P^r , for every $r \geq 0$, including $r = \infty$:

$$\hat{A}(s, \xi, c) = A(\varepsilon_0(s), \xi, c), \quad A = \hat{A} \circ (P^r \times \text{id}).$$

The same factorization preserves partial domains and every iterated trajectory with matched initial state and auxiliary streams.

Proof. The inverse identities of Definition 2.3 give existence and uniqueness. Transport the partial domain by that bijection. At every step the two representations recover the same objective and therefore the same definedness, result, and next runtime state. Induction gives trajectory and stopping equivalence. $\square$

The theorem is a representation result. It holds at order zero because the values at all points already retain the complete function. It neither asserts that all optimizers compute derivatives nor identifies a global representation with one point's jet.

3. Descent to local objective data

Definition 3.1 (Pointwise jet dependence). Fix θ . A germ-local rule is a map $A_\theta : G_\theta \times \Xi \times C \rightarrow Y$, with Y a set. It is order- r determined when it descends along the jet quotient $q_r : G_\theta \rightarrow \text{im}(q_r)$.

Theorem 3.2 (Fiber criterion). Such a factorization exists uniquely exactly when, at fixed runtime and auxiliary data, its domain is a union of q_r -fibers and its value is constant on every fiber in the domain.

Proof. A factorization has identical definedness and value on equivalent representatives. Conversely these two conditions give a well-defined domain and value on each quotient class. Surjectivity onto the image makes the descended rule unique. This is the universal property of a set quotient. $\square$

The set-valued hypothesis matters: higher-valued descent requires coherent identification data, not merely an equality of connected components. For example, the germ-local rule testing whether a germ is zero is not infinite-jet determined by Lemma 2.2. This is a mathematical rule; no finite effective germ-zero decision procedure is claimed.

Proposition 3.3 (Remote queries need not be germ-local). A legitimate value-based search can fail to depend on the current germ.

Proof. Choose a smooth bump β supported in $(1/2, 3/2)$ with $\beta(1) = 1$. The objectives $L_0(x) = x^2$ and $L_1(x) = x^2 - 2\beta(x)$ agree on a neighborhood of zero. Searching the candidates $\{0, 1\}$ with a fixed tie rule chooses 0 for L_0 and 1 for L_1 . Thus the result is not determined by the germ, or by any jet, at zero. $\square$

Peetre-Slovak theory [1] supplies stronger finite-order conclusions for suitably regular morphisms between sheaves of smooth sections. Its locality and smooth-family regularity assumptions must be checked for a particular update rule. Adaptive queries outside a neighborhood and arbitrary stateful oracle programs do not acquire those assumptions merely by being optimizers.

4. Adaptive oracle semantics

Definition 4.1 (Query machine). Fix auxiliary data and a random seed if used. A runtime state chooses either a stopping output or a query (θ, r) for the value and derivatives of orders at most r at θ . The oracle answer and current state determine the next state. Objective-independent internal computation may be included in the state transition. Each query is total on the declared smooth objective class; any declared failures must be matched as observations. Randomized runs are considered after fixing their complete random stream.

Theorem 4.2 (Adaptive indistinguishability). Suppose two objectives give identical answers to every query along a run on the first objective, and start from the same state, auxiliary data, and random stream. The second run chooses the same queries and has identical states and stopping behavior for the matched prefix. If the premise holds indefinitely, the entire runs agree.

Proof. Identical initial states choose the same first action. At a query, the assumed equal answer gives identical successor states. Those states choose the same next query or stopping result. Induction gives every finite prefix, and equality at all finite stages gives the full run. The proof does not assume in advance that the second run follows the first query schedule; induction establishes that schedule. $\square$

Theorem 4.3 (Finite-query obstruction). No deterministic value-and-derivative oracle algorithm that terminates after finitely many point queries on every smooth coercive function $\mathbb{R}^d \rightarrow \mathbb{R}$ can always return an exact global minimizer. The statement still holds if each query supplies the entire pointwise infinite jet.

Proof. Run the purported algorithm on $L_0(x) = \|x\|^2$. It must return its unique minimizer 0 after a finite query set Q . Choose x_0 outside $Q \cup \{0\}$ and a smooth compactly supported bump β with $\beta(x_0) = 1$, whose support avoids that finite set. For $M > \|x_0\|^2$, put $L_1 = L_0 - M\beta$. At each queried

point, the perturbation is zero on a neighborhood, so every derivative agrees with L_0 . Theorem 4.2 forces the same terminating run and output 0. But $L_1(x_0) < 0 = L_1(0)$, so 0 is not a global minimizer. The perturbation is bounded and compactly supported, so $L_1(x) \rightarrow +\infty$ as $\|x\| \rightarrow \infty$. By continuity its minimum is attained on a sufficiently large compact ball. This verifies both coercivity and attainment within the theorem’s domain. This contradicts the guarantee within the stated class. $\square$

This is a restricted-oracle information obstruction. Analytic objectives, known parametric families, convexity and regularity bounds, global representation inspection, and approximate guarantees can change the question. The theorem supplies no convergence rate for a particular optimizer. Conditioning on a fixed random stream gives the same adversarial statement for that deterministic run; a separate probability argument is needed for probabilistic guarantees.

5. Composition and finite-order calculus

For fixed auxiliary data, define a category whose objects are runtime state spaces and whose arrows are total objective-indexed steps $A : \mathcal{L} \times \Xi \rightarrow \Xi'$. Composition shares the objective:

$$(B \star A)(L, \xi) = B(L, A(L, \xi)).$$

Proposition 5.1 (Representation respects composition). Replacing $\mathcal{L}$ by $\text{Hol}^r(\Theta)$ gives an isomorphic category under the factorization of Theorem 2.4. Partial maps have the corresponding transported domains.

Proof. Associativity is substitution and the identity returns the state. Both composed arrows recover the same $L = \varepsilon_0(s)$, so $\widehat{B \star A} = \widehat{B} \star \widehat{A}$. Precomposition with P^r is inverse. The partial statement follows by transporting the same sequential domain condition. $\square$

Finite jets of pointed smooth maps also compose by truncated Taylor substitution: the chain rule determines the order- r composite from the order- r factors. The points and coordinate transformations are part of this construction. It is independent of a homotopy truncation level.

Lemma 5.2 (Differentiation lowers finite order). For $r \geq 1$, differentiation of one-variable germs induces $D : J^r \rightarrow J^{r-1}$. It is not an endomorphism obtained by differentiating representatives in $\mathbb{R}[z]/(z^{r+1})$.

Proof. If two germs have equal derivatives through order r , their derivatives agree through order $r - 1$. In the truncated polynomial quotient, z^{r+1} represents zero, while its derivative $(r + 1)z^r$ is nonzero. Representative differentiation therefore does not descend as an endomorphism of that same quotient. $\square$

The normalized Taylor coefficient product is convolution, $(ab)_k = \sum_{j=0}^k a_j b_{k-j}$, and derivative coefficients shift as $(Da)_k = (k + 1)a_{k+1}$. Selecting polynomial representatives and differentiating them is an explicit finite approximation, not an exact operation retaining the missing next derivative.

6. Parameter equivalence and optimizer geometry

Let $h : \Phi \rightarrow \Theta$ be a diffeomorphism between open parameter domains. A preconditioner M_θ maps objective covectors to parameter velocities. Gradients below use the declared Euclidean coordinates.

Theorem 6.1 (Transport of gradient dynamics). For $L' = L \circ h$, the flows $\dot{\theta} = -M_\theta \nabla_\theta L$ and $\dot{\phi} = -M_\phi \nabla_\phi L'$ correspond under $\theta = h(\phi)$ when

$$M_\phi = (Dh_\phi)^{-1} M_{h(\phi)} (Dh_\phi)^{-\top}.$$

Proof. The chain rule gives $\nabla_\phi L' = (Dh_\phi)^\top \nabla_\theta L$. Multiplying the second flow by Dh_ϕ and substituting the displayed preconditioner yields the first flow. Corresponding initial conditions identify trajectories

on their common interval of existence when the induced vector fields are locally Lipschitz; for example, smooth h , L , and preconditioner fields supply this local uniqueness hypothesis. No global-in-time existence is inferred. $\square$

The theorem concerns continuous-time vector fields. A finite Euler update need not commute with a nonlinear change of coordinates: generally $h(\phi + \eta v) \neq h(\phi) + \eta Dh_\phi v$. Exact discrete conjugacy requires transporting the complete update by $h^{-1} \circ \Phi \circ h$, or a separate discretization analysis.

Setting both coordinate preconditioners to the identity generally violates this condition. For $F(\theta) = \theta$ and $\tilde{F}(\phi) = a\phi$, $a > 0$, squared output error gives $\dot{F} = -(F - y)$ but $\dot{\tilde{F}} = -a^2(\tilde{F} - y)$ under their respective Euclidean flows. They represent the same possible outputs with a diffeomorphic parameterization and different learning rates. Transporting $M_\phi = a^{-2}$ restores correspondence.

For an output map $F : \Theta \rightarrow \mathbb{R}^m$, loss $\ell(F)$, and $M \succeq 0$, the exact chain-rule dynamics is

$$\dot{F} = -K_M \nabla_F \ell, \quad K_M = DF M (DF)^\top \succeq 0.$$

This kernel depends on the optimizer geometry and generally evolves with the trajectory. It is not a fixed-kernel description of every architecture or optimizer. Extra drift, memory, sampled objectives, and nonsmooth responses need explicit additional data.

7. Sheaf perspective and executable evidence

The assignment $U \mapsto C^\infty(U, \mathbb{R})$ is a sheaf: compatible local functions glue uniquely. Its stalks are the germ algebras. Prolongation and zeroth-order projection commute with restriction, so global holonomic representations form an isomorphic sheaf. A chosen smooth site embeds this discussion in a presheaf or sheaf topos; differentiation is additional smooth structure, and an arbitrary infinity-topos does not supply it [4].

The companion uses exact rational polynomial arithmetic to check finite-jet quotient operations and adaptive indistinguishability. It constructs a perturbation divisible by $(x - q)^4$ at each of six adaptive rational query points q , matching all 24 derivative coordinates through order 3; both runs end at $1/64$. It checks 84 finite-jet compositions at orders 0-6 and eight differentiation-order-loss cases. It also checks that optimizer memory changes an update at fixed current objective jet, verifies the scalar coordinate-transport calculation, and evaluates the remote-query example. The flatness and smooth bump arguments are proved above; their truth is not inferred from floating-point sampling. Program version 1.0.0, bounds, and hashes are recorded in `verification-results.json`.

Automation disclosure. OpenAI ChatGPT/Codex assisted with drafting, proof review, references, and the executable examples. The submission byline is Anonymous; no affiliation or independent referee approval is asserted. Written proofs have not been formalized in a proof assistant, and bounded program checks do not replace the general arguments.

Reproduction. Run `python3 code/jets_example.py` using the Python standard library. Polynomial identities and the recorded adaptive trace use `Fractions.Fraction`. The remote bump example evaluates two points with ordinary floating point and is separate from those exact checks. The polynomial perturbation matching six order-three queries is not the compactly supported perturbation of Theorem 4.3: its leading negative term need not preserve coercivity. It demonstrates finite-order oracle indistinguishability only; the global-minimization obstruction is the written smooth-bump proof. The complete runner and separately supplied results file retain this distinction.

References

1. J. Navarro and J. B. Sancho. **Peetre-Slovák’s theorem revisited**. 2014; revised 2016. [arXiv:1411.7499](#).
2. Brendan Fong, David I. Spivak, Rémy Tuyéras. **Backprop as Functor: A compositional perspective on supervised learning**. LICS 2019. [arXiv:1711.10455](#).
3. G. S. H. Cruttwell, Bruno Gavranović, Neil Ghani, Paul Wilson, Fabio Zanasi. **Categorical Foundations of Gradient-Based Learning**. [arXiv:2103.01931](#).
4. Jacob Lurie. **Higher Topos Theory**. Princeton University Press, 2009. [Electronic edition](#).