

From Compact Actuator Identification to Policy Recovery: Quadratic Structure and a Matched-Controller Boundary

Daniel Schmitter

Abstract

Efficient physical-model adaptation does not imply successful policy recovery. We analyze a scalar stable-lag actuator with a monotone cardinal response, showing where quadratic identification applies and where clipping changes the sufficient statistics. A finite-region formulation reduces the clipped objective to constrained quadratic subproblems, while a finite-difference counterexample shows that ordinary empirical Grams need not retain a parameter-dependent clipped loss. We then report a frozen-policy comparison on sixteen fresh simulated actuator laws. A shared-prior spline inverse achieves median normalized restoration 0.9084 and 0.8966 in two families, with modest gains over recursive least-squares controls. It fails the complete capability gate, including the required advantage over a matched polynomial prior; true inversion under the same actor also misses two-family recovery. The result identifies a boundary between a fast model-maintenance component and useful behavior. It is not a safety guarantee, unknown-change detection system, or demonstration of robotic self-improvement.

1 Introduction

An actuator change creates at least two problems: identify its new response and make the controller useful under that response. The first can be low dimensional while the second remains difficult. We study this separation under deliberately favorable information: known mechanics, continuous force sensing, a source prior, and an inherited policy.

The contribution combines an explicit objective analysis with a matched-controller experiment. It preserves two useful mechanisms-conditional quadratic fitting and compact physical response geometry-without treating either as a control-value guarantee. The complete negative capability result is retained rather than selecting only the strongest model metric.

2 Related work

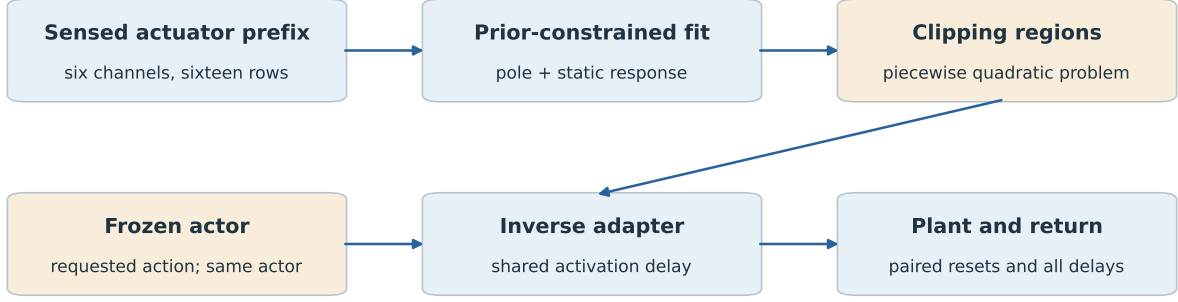
Hammerstein identification, kernels on dynamical systems, and policy learning are established fields. Dynamic-system kernels already compare trajectories through inner products [1]; kernel-based Hammerstein identification predates this construction [2]. Soft actor-critic [3] is the inherited policy-learning method. The present specialization uses a fixed cardinal carrier and a supplied low-dimensional prior; it is not a new general convex method for arbitrary pole learning or a replacement for policy optimization.

3 Architecture and information flow

The adaptation pipeline fits a response model and exposes an inverse adapter to an unchanged actor. Identification uses supplied force sensing and prior response directions. Every arm pays a common activation delay before the adapter is used; controller rewards during that delay remain in the score. The model-fit calculation and the policy rollout are separate stages, and the latter changes which commands the former is asked to invert.

The exact quadratic reduction applies only within consistent clipping regions. A coefficient-dependent saturation boundary is not a fixed feature map, which is why ordinary historical Grams can omit information required by the deployed objective. The same separation appears at task level: a response distance averaged over uniform commands need not reflect errors under commands visited by the actor. The true-inverse intervention tests whether inverse fitting alone is plausibly the bottleneck.

A model repair still has to help the actor



Identification configures the adapter. During control, the frozen actor supplies its requested action.

Figure 1: A model repair still has to help the actor. Boxes distinguish supplied information, fitted components, and the quantity evaluated. Arrows show computation or data dependence, not a newly trained deep network.

4 Identification with a stable lag

Let force evolve as a stable first-order recurrence with a clipped static drive. The pole is constrained between zero and 0.85. For a fixed source mean μ and directions D , parameterize the static coefficients using a scaled correction a :

$$p^+ = \rho p + (1 - \rho) \text{clip}(Bc, -1, 1), \quad (1)$$

$$c = \mu + \frac{Da}{1 - \rho}. \quad (2)$$

Without clipping, the prediction is affine in the joint variables ρ and a . Squared residuals are therefore quadratic and coefficient-monotonicity conditions are linear. However, a regularizer on Da corresponds to a pole-dependent regularizer on the static coefficients. The unclipped objective is not the deployed clipped loss; calling both the same convex fit would be incorrect.

$$\hat{p}_{\text{raw}} = B\mu + \begin{bmatrix} p - B\mu & BD \end{bmatrix} \begin{bmatrix} \rho \\ a \end{bmatrix}. \quad (3)$$

For sorted scalar commands and a coefficient-monotone curve, saturated observations form a lower prefix and upper suffix, with one interior block. Enumerating the $(n+1)(n+2)/2$ assignments yields affine predictions and linear region-consistency constraints within each assignment. Each subproblem is a convex quadratic fit under the unchanged scaled penalties. This is a finite-region specialization, not a convex objective over unrestricted nonlinear dynamics.

Dropping selected constraints gives lower bounds for pruning assignments. Batched Hessian construction and solves reduce generic optimization calls. On a reused 576-fit panel, ninety of 96 six-actuator cases finish below 100 ms, but the worst is 133.39 ms: the all-case timing gate fails. Numerical optimization statuses and residual guards are retained; floating-point pruning is not described as a formal exact-arithmetic certificate.

5 A memory boundary caused by clipping

Proposition 1. An ordinary cubic-feature empirical Gram and target right-hand side do not generally determine every clipped squared loss. At eight points $u_i = i/112$, give one zero-target dataset the binomial weights $C(7, i)$ on even indices and the other the same weights on odd indices. Their moments through degree six agree by the seventh finite-difference identity. Their cubic-feature Grams, sample counts, and zero target statistics consequently agree.

$$\sum_{i=0}^7 (-1)^i \binom{7}{i} u_i^m = 0 \quad (0 \leq m \leq 6), \quad (4)$$

$$q_i = \text{clip}(64u_i - 1, -1, 1)^2, \quad (5)$$

$$\left| \sum_{i=0}^7 (-1)^i \binom{7}{i} q_i \right| = \frac{16}{7}. \quad (6)$$

Thus no decoder of only those old statistics can evaluate both losses correctly for this drive. The obstruction is the parameter-dependent clipping partition, not floating-point error. Retaining observations, ordered region information, or additional statistics can change the information available. The counterexample does not invalidate fixed-feature quadratic retention or immutable archived models.

6 Response geometry and its limit

For a stable scalar response with pole ρ_i , bounded drive ϕ_i , and zero-command equilibrium b_i , a single command followed by zero commands gives a geometric transient. Averaging over independent uniform initial force and command in minus one to one yields an exact response Gram:

$$K_{ij} = b_i b_j + \frac{\rho_i \rho_j / 3 + \mathbb{E}_u[\delta_i(u) \delta_j(u)]}{1 - \rho_i \rho_j}, \quad \delta_i = (1 - \rho_i) \phi_i - b_i. \quad (7)$$

This follows by summing the geometric series and using the initial-force first and second moments. It is positive semidefinite because it is an inner product. It need not be invertible or circulant. The equilibrium is counted once, not summed forever. Exact clipped-polynomial products avoid time unrolling, but the uniform reference measure is not the controller’s trajectory distribution.

7 Experimental methods

The decisive matched-policy panel fixes the calibrated zero-context actor and compares identity, frozen sixteen-row RLS, online RLS, a six-coordinate polynomial-prior inverse, a six-coordinate shared-prior inverse, true inverse, and nominal dynamics. Two actuator families each contain eight fresh law seeds, with eight paired resets per law and 1,000 steps per episode. Shared and polynomial fits use the same cardinal carrier but different subspaces; geometry and source-subspace selection are not separately isolated.

Exactly sixteen sensed prefix rows fit each six-channel candidate. Common activation charges the time of both fits and RLS construction; all delay rewards count. Mechanics and ongoing noisy force sensing are supplied. The actor inherits 611,088 source training transitions, 126,000 validation transitions, and earlier 20-million-transition pretraining. No target policy-gradient update is performed. These costs preclude describing the complete agent as learned from sixteen examples.

Returns are paired-reset normalized, averaged within law, and summarized by family median. Uncertainty uses the law as the bootstrap unit, with 10,000 fixed resamples. Predeclared criteria include 0.90 restoration, a 0.80 all-law floor, and at least 0.02 gain over each cheap comparator. Contrasts are medians of paired law differences, not differences of marginal medians.

8 Results

Shared-prior paired gains over frozen RLS are 0.02124 and 0.02292; their descriptive law-bootstrap lower bounds are positive. Gains over the polynomial prior are only 0.00747 and 0.01164, below the declared two-point criterion. Compound lag also misses restoration and the all-law floor. True inversion under the same actor fails two-family restoration, weakening the hypothesis that further fit refinements alone would unlock the desired capability.

All 898,048 evaluator transitions and 1,536 scalar-fit reproductions pass the retained recurrence and immutability audits. Maximum common fitting takes 15.488 ms, activation is always step seventeen, numeric state is 651,900 bytes, and evaluator RSS is 317.672 MiB. These are simulation-run receipts on the local machine, not a hardware control deadline or a certified safe-probing procedure.

Table 1: Complete same-policy restoration comparison; one is nominal performance.

Inverse	Asymmetric lag	Compound lag
Identity	0.2295	0.2579
Frozen RLS	0.8850	0.8722
Online RLS	0.8766	0.8568
Polynomial prior	0.8999	0.8901
Shared prior	0.9084	0.8966
True inverse	0.9123	0.8977

The true inverse does not fully restore this actor

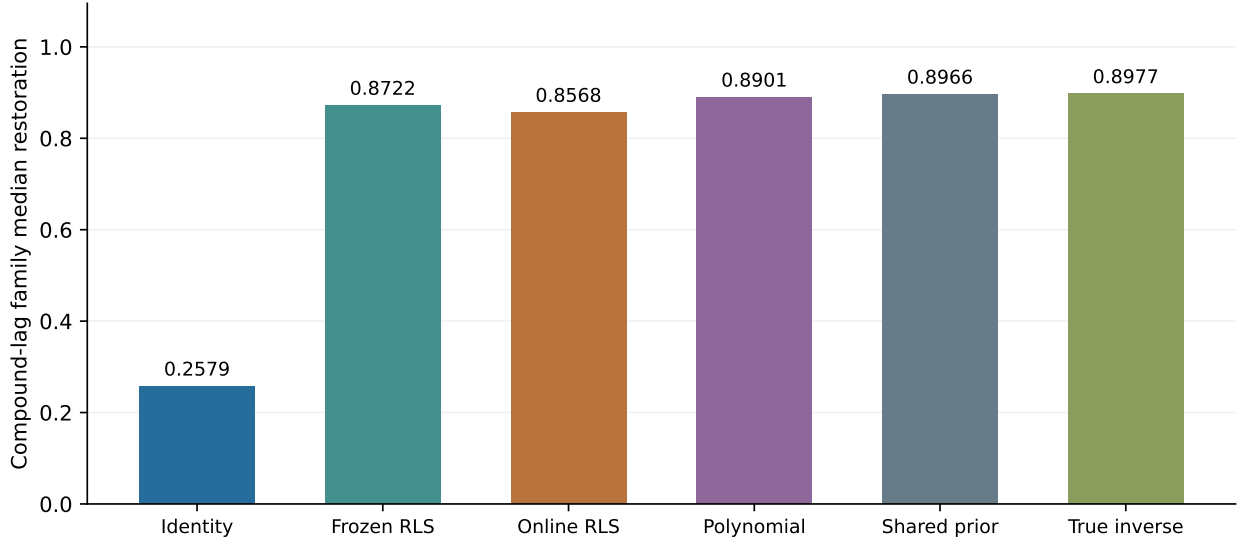


Figure 2: Complete compound-lag family medians under the same frozen actor. The true inverse is an intervention, not an optimal-policy upper bound. The asymmetric family remains in the table.

9 Discussion

The positive result is modest matched-controller improvement over two RLS baselines. The negative result is failure of the stronger reusable-recovery claim. A true inverse is only an intervention under this fixed actor, not an optimal-policy upper bound. Better policy learning or a different sensing interface could change the result, but that would be a different experiment.

The theoretical tools remain useful when their boundaries are respected: solve the quadratic part, retain statistics for the objective actually used, and measure geometry under an appropriate reference. None supplies missing behavior automatically. The branch is closed without retuning these fresh-law outcomes, and no new policy experiment is run for this reconstruction.

10 Application boundary and research implication

The current evidence supports modest adaptation benefits under this actor, not the larger recovery objective. The missing gain from true inversion makes further coefficient tuning an especially weak rescue strategy. A different controller, observation interface, or disturbance family would be a different study, not a reinterpretation of these failed gates.

11 Conclusion

Efficient actuator fitting is a valid component capability. Its exact algebra and compact state do not establish robust behavioral recovery. The matched comparison and clipping counterexample specify two concrete limits that a future model-adaptation system must address.

References

- [1] S. V. N. Vishwanathan, A. J. Smola, and R. Vidal. Binet-Cauchy Kernels on Dynamical Systems and its Application to the Analysis of Dynamic Scenes. IJCV 73, 95-119, 2007.
<https://doi.org/10.1007/s11263-006-9352-0>
- [2] R. S. Risuleo, G. Bottegal, and H. Hjalmarsson. A Kernel-Based Approach to Hammerstein System Identification. 2016. <https://arxiv.org/abs/1412.4055>
- [3] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. ICML, 2018.
<https://proceedings.mlr.press/v80/haarnoja18b.html>