

Uniformly Accurate EDMD Generators Can Give Unbounded Auxiliary Optima

Alper Ferudun*

25 October 2026

Abstract

We construct noiseless data for which extended dynamic mode decomposition (EDMD) approximates an infinitesimal generator uniformly on a fixed polynomial auxiliary space, while the associated unregularized auxiliary optimization has value $-\infty$ instead of the correct value 1. For $\dot{x} = -x^5$ on $[-1, 1]$, we use quadratic auxiliary functions and a degree-six regression dictionary. The data are noiseless and the regression has full column rank. The result holds for every set of $N \geq 9$ Chebyshev nodes and, almost surely, for all sufficiently large datasets sampled independently from the full-support arcsine measure. An explicit sign certificate holds simultaneously for every $0 < \tau \leq 2^{-26}$, while the generator error is at most 156τ in the stated coefficient norm. A valid lower constraint on the auxiliary bound makes the approximate value finite but still incorrect. This realizes a strict-feasibility obstruction already identified in earlier work; it does not contradict that work's convergence theorems or settle the general optimal-value error-estimation problem. We also record elementary coefficient-cap and robust-penalty repairs.

1 The question and the role of earlier work

The auxiliary-function approach bounds long-time averages by minimizing a pointwise upper bound involving the generator of a dynamical system. Bramburger and Fantuzzi [1] developed a data-driven version using extended dynamic mode decomposition (EDMD). Fantuzzi's Oberwolfach contribution [3] asks for quantitative errors in the resulting optimal values as functions of data count, dictionary degree and sampling time. Operator approximation estimates, including those in [4], are a natural input to this question.

There is already an essential warning in the literature. Bramburger and Fantuzzi [2, Section 4.4] explain that convergence of approximate generator coefficients need not imply convergence of optimization values without strict feasibility. Their illustrative perturbation of a discrete quadratic map cannot, however, arise from exact data once the polynomial regression is sufficiently rich, as they explicitly observe. Their circular-attractor example in [1, Section 6.5] exhibits an incorrect ergodic bound when the sampling hypothesis fails.

The example below supplies an actual finite-timestep EDMD realization with exact flow data, a full-rank regression, full-support sampling and uniform generator convergence on the entire auxiliary space. The mechanism is the known loss of optimization stability; the explicit realization and uniform sign estimate are the point of this note. The auxiliary and dictionary degrees remain fixed. We make no claim about a limit that increases the dictionary degree first, nor about the full quantitative program in [3].

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: [AlperTheKing](https://github.com/AlperTheKing).

2 Exact and data-driven auxiliary problems

Write $\mathcal{P}_k$ for the real polynomials of degree at most k , restricted to $X = [-1, 1]$. Consider

$$\dot{x} = -x^5, \quad \Phi_t(x) = x(1 + 4tx^4)^{-1/4}, \quad g(x) = 1 - x^2. \quad (1)$$

The formula gives a smooth forward flow on $\mathbb{R}$. The interval X is forward invariant and absorbing: for $t > 0$, $|\Phi_t(x)| \leq (4t)^{-1/4}$. Every trajectory tends to zero, so the long-time average of g is 1.

Let $V = \mathcal{P}_2$, $W = \mathcal{P}_6$, and $\mathcal{L}v = -x^5v'$. For a linear map $A : V \rightarrow C(X)$ that annihilates constants, define

$$\alpha(A) = \inf_{v \in V} \max_{x \in X} \{g(x) + (Av)(x)\}. \quad (2)$$

Equivalently, this is the infimum of U subject to $U - g - Av \geq 0$ on X . Since $(\mathcal{L}v)(0) = 0$ and $g(0) = 1$, while $v = 0$ is admissible,

$$\alpha(\mathcal{L}) = 1. \quad (3)$$

The same reasoning gives value 1 when V is replaced by $C^1(X)$. Thus the chosen auxiliary degree causes no gap in the exact comparator.

Let μ be the arcsine probability measure

$$d\mu(x) = \frac{dx}{\pi\sqrt{1-x^2}},$$

and let P be the $L^2(\mu)$ orthogonal projection onto W . For data nodes $x_1, \dots, x_N$ with a full column-rank degree-six evaluation matrix, let P_N be the equal-weight least-squares projection onto W . Define

$$A_{\tau,Q}v = Q \left(\frac{v \circ \Phi_\tau - v}{\tau} \right), \quad Q = P \text{ or } P_N. \quad (4)$$

This is precisely the rectangular EDMD construction from [3], with auxiliary basis $(1, x, x^2)$ and dictionary $(1, x, \dots, x^6)$. Indeed, fitting $v \circ \Phi_\tau$ and then subtracting v agrees with (4), because $Qv = v$. The snapshots are $(x_j, \Phi_\tau(x_j))$, without observation noise.

For $v = c_0 + c_1x + c_2x^2$, use the quotient coefficient norm $\|v\|_* = |c_1| + |c_2|$, which ignores constants.

Theorem 1. *Set $\tau_0 = 2^{-26}$. Each of the following choices of Q has the properties stated below:*

1. *the population projection $Q = P$;*
2. *$Q = P_N$ at the equally weighted nodes*

$$x_j = \cos \left(\frac{(2j-1)\pi}{2N} \right), \quad 1 \leq j \leq N, \quad N \geq 9;$$

3. *$Q = P_N$ for iid nodes with law μ , almost surely for all $N \geq N_0$, where N_0 is finite and does not depend on τ .*

In all three cases, simultaneously for every $0 < \tau \leq \tau_0$,

$$A_{\tau,Q}(x^2)(x) \leq -\frac{\tau}{64} \quad (x \in X), \quad (5)$$

$$\|(A_{\tau,Q} - \mathcal{L})v\|_\infty \leq 156\tau\|v\|_* \quad (v \in V), \quad (6)$$

$$\alpha(A_{\tau,Q}) = -\infty, \quad \alpha(\mathcal{L}) = 1. \quad (7)$$

All empirical regressions in these regimes have full column rank. If the approximate optimization is additionally constrained by $U \geq \min_X g = 0$, its value is 0, still different from the exact value.

3 A negative projected derivative

The Chebyshev polynomials satisfy $T_k(\cos \theta) = \cos(k\theta)$. The functions $e_0 = 1$ and $e_k = \sqrt{2}T_k$, $1 \leq k \leq 6$, form an orthonormal basis of W in $L^2(\mu)$. Since $\sum_{k=0}^6 e_k(x)^2 \leq 13$, Cauchy–Schwarz gives

$$\|Ph\|_\infty \leq \sqrt{13}\|h\|_{L^2(\mu)} \leq \sqrt{13}\|h\|_\infty < 4\|h\|_\infty. \quad (8)$$

Expanding a tenth power of a cosine gives

$$x^{10} = \frac{126 + 210T_2 + 120T_4 + 45T_6 + 10T_8 + T_{10}}{512}.$$

Consequently

$$q(x) := P(x^{10}) = -\frac{9}{512} + \frac{135}{256}x^2 - \frac{75}{32}x^4 + \frac{45}{16}x^6. \quad (9)$$

For $\delta = 1/8$, discarding the negative quartic term shows that

$$q(x) \leq -\frac{9}{512} + \frac{135}{256}\delta^2 + \frac{45}{16}\delta^6 = -\frac{39123}{4194304} \leq -\frac{1}{128}, \quad |x| \leq \delta. \quad (10)$$

Lemma 2. *Suppose that $Q : C(X) \rightarrow W$ is linear, reproduces W , and satisfies*

$$\|Q\|_{\infty \rightarrow \infty} \leq 6, \quad \|Q(x^{10}) - q\|_\infty \leq \frac{1}{256}. \quad (11)$$

Then (5) and (6) hold for $0 < \tau \leq \tau_0$.

Proof. Set $h_\tau = (\Phi_\tau^2 - x^2)/\tau$. Taylor’s theorem applied to $(1+z)^{-1/2}$ for $z \geq 0$ gives

$$h_\tau(x) = -2x^6 + 6\tau x^{10} + r_\tau(x), \quad |r_\tau(x)| \leq 20\tau^2|x|^{14} \leq 20\tau^2. \quad (12)$$

The remainder follows from $|[(1+z)^{-1/2}]'''| \leq 15/8$ and $z = 4\tau x^4$. Because Q reproduces x^6 , the polynomial $p_\tau = Qh_\tau$ satisfies

$$\|p_\tau + 2x^6\|_\infty \leq 36\tau + 120\tau^2 \leq 156\tau \quad (\tau \leq 1). \quad (13)$$

On $|x| \leq \delta$, equations (10) and (11) imply $Q(x^{10}) \leq -1/256$. Hence

$$p_\tau(x) \leq -\frac{3\tau}{128} + 120\tau^2 \leq -\frac{\tau}{64},$$

where the last inequality uses $120\tau_0 \leq 1/128$. On $|x| \geq \delta$, equation (13) gives

$$p_\tau(x) \leq -2\delta^6 + 156\tau \leq -\delta^6 \leq -\frac{\tau}{64}.$$

Here $\delta^6 = 2^{-18}$, $156\tau_0 \leq 2^{-18}$, and $\tau_0/64 \leq 2^{-18}$. This proves the uniform sign estimate.

For the other nonconstant basis function, the second derivative bound $|[(1+z)^{-1/4}]''| \leq 5/16$ yields

$$\left| \frac{\Phi_\tau(x) - x}{\tau} + x^5 \right| \leq \frac{5}{2}\tau|x|^9.$$

Since Q reproduces x^5 , we obtain $\|A_{\tau,Q}x - \mathcal{L}x\|_\infty \leq 15\tau$. Together with (13) and $A_{\tau,Q}1 = \mathcal{L}1 = 0$, this proves (6). $\square$

4 Population, finite and iid sampling

Equation (8) and the definition of q show that $Q = P$ satisfies (11). We verify that the same lemma applies to the two empirical regimes.

4.1 Explicit equally weighted data

For the nodes in Theorem 1(2), the equal-weight quadrature rule integrates every polynomial of degree at most $2N - 1$ exactly against μ . For completeness, the average of $T_k(x_j)$ vanishes when $1 \leq k < 2N$: odd k cancel by reflection, while even $k = 2\ell$ reduce to the real part of a nontrivial geometric sum of N th roots of unity. This proves the assertion since $T_0, \dots, T_{2N-1}$ span that polynomial space.

For $N \geq 9$, this exactness includes the products $e_i e_j$, of degree at most 12, and $x^{10} e_i$, of degree at most 16. The empirical Gram matrix is therefore I , and $P_N(x^{10}) = q$. By orthogonal projection in the empirical inner product,

$$|P_N h(x)| \leq \left(\sum_{k=0}^6 e_k(x)^2 \right)^{1/2} \left(\frac{1}{N} \sum_{j=1}^N |h(x_j)|^2 \right)^{1/2} \leq \sqrt{13} \|h\|_\infty.$$

Thus (11) holds for every $N \geq 9$. The nodes are distinct, so the degree-six evaluation matrix has full column rank already for $N \geq 7$: a nonzero polynomial of degree at most six cannot have seven distinct zeros. For odd N , one data point is the equilibrium snapshot $(0, 0)$. Even then, unconstrained overdetermined least squares need not interpolate that particular snapshot.

4.2 Independent samples of full support

Now let $x_1, x_2, \dots$ be iid with distribution μ . Since μ is nonatomic, almost surely all nodes are distinct. Put

$$e = (e_0, \dots, e_6)^\top, \quad G_N = \frac{1}{N} \sum_{j=1}^N e(x_j) e(x_j)^\top.$$

The strong law applied to finitely many bounded functions gives, on one event of probability one,

$$G_N \longrightarrow I, \quad \frac{1}{N} \sum_{j=1}^N x_j^{10} e(x_j) \longrightarrow \int_X x^{10} e(x) d\mu(x).$$

Continuity of matrix inversion and equivalence of norms in the fixed polynomial space imply $P_N(x^{10}) \rightarrow q$ uniformly. Thus there is a finite random N_0 such that, for every $N \geq N_0$,

$$\lambda_{\min}(G_N) \geq \frac{1}{2}, \quad \|P_N(x^{10}) - q\|_\infty \leq \frac{1}{256}. \quad (14)$$

For any bounded h , the empirical projection estimate gives

$$|P_N h(x)| \leq (e(x)^\top G_N^{-1} e(x))^{1/2} \|h\|_{\text{emp}} \leq \sqrt{26} \|h\|_\infty < 6 \|h\|_\infty.$$

One way to see the first inequality is to express evaluation at x as an inner product in W with its empirical reproducing vector and then use the contraction of P_N in the empirical norm. Consequently (11) holds.

Crucially, the event in (14) and the integer N_0 involve only the nodes, not τ . Lemma 2 therefore applies simultaneously to every $0 < \tau \leq \tau_0$, without an uncountable intersection of almost-sure events.

4.3 Completion of the optimization argument

For $v = \lambda x^2$ with $\lambda \geq 0$, equation (5) implies

$$\max_X (g + A_{\tau, Q} v) \leq 1 - \frac{\lambda \tau}{64}.$$

Letting $\lambda \rightarrow \infty$ proves $\alpha(A_{\tau,Q}) = -\infty$. If $U \geq 0$ is also required, the choice $\lambda = 64/\tau$ makes $U = 0$ feasible, proving the last assertion of Theorem 1.

The only invariant probability measure of (1) on X is δ_0 . Indeed, invariance implies $\int \Phi_t(x)^2 d\nu = \int x^2 d\nu$; dominated convergence as $t \rightarrow \infty$ makes the right side zero. Its moment matrices are singular. In contrast, no probability measure can satisfy $\int A_{\tau,Q} v d\nu = 0$ for all $v \in V$, since $v = x^2$ makes this integral strictly negative. The strict invariant-moment feasibility used in [2, Assumption 4.1(ii) and Section 4.3] is therefore absent. Full rank of a sampling Gram matrix is a different property.

5 A convex criterion and elementary repairs

The next observations are standard consequences of separation and perturbation inequalities. We include them to identify the precise source of the instability and a simple way to control it.

Proposition 3. *Let X be compact, let $g, a_1, \dots, a_r \in C(X)$, and write $a = (a_1, \dots, a_r)$. Then*

$$\inf_{c \in \mathbb{R}^r} \max_{x \in X} \{g(x) + c \cdot a(x)\}$$

is finite if and only if $0 \in \text{conv}(a(X))$.

Proof. If 0 does not belong to that compact convex set, strict separation provides c and $\epsilon > 0$ such that $c \cdot a(x) \leq -\epsilon$ for every x . Scaling c makes the infimum $-\infty$. If 0 belongs to the convex hull, a finite convex combination supplies a probability measure ν with $\int a d\nu = 0$. Integration then gives the finite lower bound $\int g d\nu$, while $c = 0$ gives the finite upper bound $\max g$. $\square$

For a basis b modulo constants and any coefficient norm, define

$$\alpha_R(A) = \inf_{\|c\| \leq R} \max_X \{g + A(c \cdot b)\}.$$

If $\|(A - \mathcal{L})(c \cdot b)\|_\infty \leq \epsilon \|c\|$ for every c , comparing the two maxima for the same vector and then taking infima gives

$$|\alpha_R(A) - \alpha_R(\mathcal{L})| \leq R\epsilon. \quad (15)$$

Similarly,

$$\inf_c \left[\max_X \{g + A(c \cdot b)\} + \epsilon \|c\| \right] \geq \alpha(\mathcal{L}). \quad (16)$$

For our example, take $b = (x, x^2)$, the ℓ^1 norm and $\epsilon = 156\tau$. The robust value in (16) is exactly 1: the inequality gives the lower bound, and $c = 0$ attains it. These observations do not supply the general N, m, τ theory requested in [3]; they show explicitly why a condition controlling coefficients or feasible invariant moments is substantive.

Remark 4. Exact-generator EDMD, which fits samples of $\mathcal{L}v$ instead of flow finite differences, recovers $\mathcal{L}$ here as soon as the regression has full rank, since $\mathcal{L}(\mathcal{P}_2) \subseteq \mathcal{P}_6$. The counterexample concerns the finite-timestep construction (4). The constant 2^{-26} is a convenient proved threshold, not a proposed sharp value.

6 Verification and scope

The proof is analytic. The accompanying standard-library Python verifier checks the projection both by rational moment equations and by a Chebyshev recurrence, all rational margins

in the sign estimate, and quadrature identities through independent Newton power sums for several data counts. It passes 297 exact checks in ordinary and optimized Python. The all- N conclusion and the almost-sure quantifiers are established by the written proof, not by finite experiments.

This is a complete counterexample within its stated fixed-dictionary regimes. The broader error-estimation question remains open here. The qualitative strict-feasibility obstruction is credited to [2]. A targeted literature search found no matching construction, but establishes no absolute priority claim.

AI tools assisted the mathematical derivation, literature search, verification code and exposition. This is an unrefereed preprint; no independent review or proof-assistant formalization is claimed.

References

- [1] J. J. Bramburger and G. Fantuzzi, *Auxiliary functions as Koopman observables: Data-driven analysis of dynamical systems via polynomial optimization*, Journal of Nonlinear Science **34** (2024), article 8. [doi:10.1007/s00332-023-09990-2](https://doi.org/10.1007/s00332-023-09990-2). See also [arXiv:2303.01483v4](https://arxiv.org/abs/2303.01483v4).
- [2] J. J. Bramburger and G. Fantuzzi, *Data-driven discovery of invariant measures*, Proceedings of the Royal Society A **480** (2024), 20230627. [doi:10.1098/rspa.2023.0627](https://doi.org/10.1098/rspa.2023.0627). See also [arXiv:2308.15318v2](https://arxiv.org/abs/2308.15318v2).
- [3] G. Fantuzzi (joint work with J. Bramburger), *Time averages, polynomial optimization and Koopman*, in *Mini-Workshop: Data-driven Modeling, Analysis, and Control of Dynamical Systems*, Oberwolfach Report 57/2024, pp. 3291–3292. [doi:10.4171/OWR/2024/57](https://doi.org/10.4171/OWR/2024/57).
- [4] L. Llamazares-Elias, S. Llamazares-Elias, J. Latz and S. Klus, *Data-driven approximation of Koopman operators and generators: Convergence rates and error bounds* (2026). [doi:10.3934/jcd.2026012](https://doi.org/10.3934/jcd.2026012); [arXiv:2405.00539v3](https://arxiv.org/abs/2405.00539v3).