# Basin Geometry of High-Order Trajectory Inference in Chaotic Hamiltonian Systems

**Angshul Majumdar**

*Indraprastha Institute of Information Technology Delhi (IIIT-Delhi), New Delhi 110020, India.*

**E-mail: angshul@iiitd.ac.in**

*(Received: Apr 27, 2026; Revised: May 15, 2026; Accepted: May 23, 2026; Published: Jun 30, 2026)*

**Abstract** This paper develops a rigorous basin-geometry theory for high-order stationary-point iterations applied to trajectory reconstruction in chaotic Hamiltonian systems. By formulating the inverse problem through a discrete action functional induced by symplectic dynamics, explicit bounds are derived for the second, third, and fourth derivatives of the objective on collision-free trajectory tubes. These estimates yield quantitative convergence radii for Newton, Halley, and higher-order Householder methods. The analysis separates forward-time chaotic instability from inverse conditioning by characterising coercivity on the identifiable subspace through observation Gramians and sensitivity growth. Specialisation to the planar three-body problem reveals explicit links between admissible initialization radii, finite-time Lyapunov exponents, and sampling density. The resulting framework provides a mechanism-level explanation for instability in practical trajectory reconstruction and establishes principled conditions for robust solver and sampling design in chaotic mechanical systems.

**Keywords:** Chaotic Hamiltonian systems, three-body problem, Householder methods, Newton methods, basin geometry, trajectory reconstruction, Lyapunov exponents, inverse problems.

## I Introduction

Throughout the paper, we consider a finite continuous time horizon $[0, T]$ and a fixed discretisation step $h > 0$, leading to a discrete trajectory $x_{0:N}$ with $N = T/h$. All variational and stationary-point analyses are performed at this discrete level, while constants governing derivative growth depend on the continuous horizon length $T$.

The Newtonian planar three-body problem remains one of the most studied nonlinear Hamiltonian systems in mathematical physics. Despite its compact formulation, the system exhibits strong sensitivity to initial conditions, non-integrability in the general case, and rich chaotic behaviour [1, 2, 3]. The exponential divergence of nearby trajectories, quantified through Lyapunov exponents, limits long-horizon prediction and renders inverse trajectory problems intrinsically delicate [4].

In this work we study the finite-horizon inverse problem: reconstructing a trajectory from partial noisy observations when the Hamiltonian is known. Rather than learning a surrogate dynamics, we formulate trajectory inference variationally by minimizing a discrete action functional consistent with the Hamiltonian flow. Variational data assimilation and weak-constraint formulations have demonstrated the effectiveness of such approaches in dynamical systems inference [5, 6]. However, for chaotic Hamiltonian systems, the geometric properties of the resulting objective function remain insufficiently understood.

From the perspective of nonlinear analysis, convergence of Newton-type methods is governed by local curvature and higher-order derivative bounds [7]. While classical results establish quadratic and higher-order convergence under strong regularity assumptions, explicit basin-of-attraction estimates in chaotic dynamical settings are largely absent. In Hamiltonian systems, structure-preserving discretizations maintain symplectic geometry and energy behaviour [8], yet their implications for inverse conditioning have not been quantified in terms of Lyapunov growth.

Recent machine-learning approaches such as Physics-Informed Neural Networks [9], Neural Ordinary Differential Equations [10], and Hamiltonian Neural Networks [11] aim to learn dynamics or trajectories from data while respecting physical structure. Although effective in forward prediction and short-term approximation, these approaches provide limited theoretical guarantees on basin geometry, statistical efficiency, or stability in chaotic regimes.

The present work develops a rigorous analytical framework linking Hamiltonian sensitivity, variational curvature, and high-order stationary iterations. We derive explicit derivative bounds for the discrete action functional induced by symplectic flow, establish coercivity estimates on the identifiable subspace, and prove quantitative basin-of-attraction theorems for Newton, Halley, and general Householder methods. Specialising the theory to the planar three-body problem, we obtain sampling-dependent lower bounds and show how Lyapunov growth governs admissible initialization radii.

This approach provides a unified geometric interpretation of inverse trajectory conditioning in chaotic Hamiltonian systems, bridging classical celestial mechanics, nonlinear optimization theory, and modern dynamical system inference.

# II Variational Formulation of Finite-Horizon Trajectory Inference

We consider a smooth Hamiltonian system on $\mathbb{R}^{2d}$,

$$\dot{x} = J\nabla H(x), \qquad x(0) = \theta \in \mathbb{R}^{2d}, \tag{1}$$

where $J$ is the canonical symplectic matrix and $H : \mathbb{R}^{2d} \to \mathbb{R}$ is $C^{p+1}$ with $p \geq 3$ [12, 8]. We denote by $\Phi_t(\theta)$ the associated flow map.

The planar three-body problem corresponds to the special case $d = 6$ with Newtonian Hamiltonian structure; see [1, 2]. Chaotic sensitivity and Lyapunov growth in Hamiltonian systems are well documented [4].

## II.a Observation model

We assume observations at times $0 < t_{k_1} < \cdots < t_{k_m} \leq T$. Let $\Pi_r : \mathbb{R}^{2d} \to \mathbb{R}^r$ be a fixed linear observation operator (e.g., position projection in the three-body case).

Observations satisfy

$$y_k = \Pi_r \Phi_{t_k}(\theta) + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, R), \tag{2}$$

with $R$ symmetric positive definite. Gaussian observational models of this type arise naturally in variational data assimilation and inverse problems [5, 6].

## II.b Trajectory-space formulation

Variational trajectory inference can be formulated via minimization of a discrete action functional consistent with a structure-preserving discretization of the Hamiltonian flow [13, 8]. Let $F$ be a $C^{p+1}$ symplectic integrator approximating $\Phi_h$. We define

$$\mathcal{J}(x_{0:N}) = \frac{1}{2}\|x_0 - \mu_0\|^2 + \frac{1}{2}\sum_{k=0}^{N-1}\|x_{k+1} - F(x_k)\|^2_{Q^{-1}} + \frac{1}{2}\sum_{k\in\mathcal{K}}\|y_k - \Pi_r x_k\|^2_{R^{-1}} + \frac{\mu}{2}\sum_{k=0}^{N}\|x_k\|^2. \tag{3}$$

Such weak-constraint formulations are standard in data assimilation [5].

## II.c Initial-condition pullback

Restricting to dynamically admissible trajectories $x_k = \Phi_{t_k}(\theta)$, the objective pulls back to parameter space:

$$\mathcal{J}_\theta(\theta) = \frac{1}{2}\sum_{k\in\mathcal{K}} \|y_k - \Pi_r \Phi_{t_k}(\theta)\|^2_{R^{-1}} + \frac{\mu}{2}\|\theta\|^2. \tag{4}$$

Differentiation yields

$$\nabla^2_\theta \mathcal{J}_\theta = \mathcal{I}(\theta) + \mathcal{R}(\theta) + \mu I, \tag{5}$$

where $\mathcal{I}(\theta)$ is the Fisher information operator [6] and $\mathcal{R}(\theta)$ collects higher-order residual terms.

## II.d Identifiable subspace

Define the stacked sensitivity matrix

$$A(\theta) = \begin{bmatrix} \Pi_r D\Phi_{t_{k_1}}(\theta) \\ \vdots \\ \Pi_r D\Phi_{t_{k_m}}(\theta) \end{bmatrix}. \tag{6}$$

The identifiable subspace is

$$\mathcal{V}(\theta) := \operatorname{range}(A(\theta)^\top). \tag{7}$$

Identifiability and information geometry in inverse problems are discussed in [6, 5].

## II.e Regularity assumptions

We assume $H \in C^{p+1}$ and that the reference trajectory remains in a compact set $\Omega$. Under compactness, derivative bounds on $D^j(J\nabla H)$ imply corresponding bounds on $D^j\Phi_t$ via standard ODE sensitivity theory [12, 8].

# III Differential Structure of the Action Functional

This section establishes explicit derivative representations and uniform bounds for the action functional (3) defined on the finite-horizon trajectory space. We work throughout with the Euclidean product structure on

$$\mathcal{X}_N := (\mathbb{R}^n)^{T+1}, \qquad n := 2d,$$

equipped with the inner product and norm

$$\langle x, z\rangle_{\mathcal{X}_N} := \sum_{k=0}^{T} \langle x_k, z_k\rangle_{\mathbb{R}^n}, \qquad \|x\|^2_{\mathcal{X}_N} := \sum_{k=0}^{T} \|x_k\|^2.$$

For a symmetric positive definite matrix $A \in \mathbb{R}^{m\times m}$ we use

$$\|v\|^2_A := v^\top A v, \qquad \lambda_{\min}(A),\ \lambda_{\max}(A) \text{ for its extremal eigenvalues.}$$

## III.a Standing assumptions and tube-local bounds

We fix a horizon $T \in \mathbb{N}$ and a tube $\mathcal{T}_T \subset \mathcal{X}_N$ on which uniform bounds hold.

**Assumption 1** (Smoothness and derivative bounds of the flow map)**.** *The discrete flow map $F : \mathbb{R}^n \to \mathbb{R}^n$ is of class $C^4$. There exist constants $M_j(T) \in (0, \infty)$ such that for all $x \in \mathbb{R}^n$ with $(x_0, \dots, x_N) \in \mathcal{T}_T$ and for all $j \in \{1, 2, 3, 4\}$,*

$$\|D^j F(x_k)\|_{\mathrm{op}} \le M_j(T), \qquad 0 \le k \le N-1, \tag{8}$$

*where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm induced by the Euclidean norm, and $D^j F$ is viewed as a $j$-linear map equipped with its natural operator norm.*

**Assumption 2** (Uniform bound on model residuals inside the tube)**.** *Define the model residuals $r_k(x) := x_{k+1} - F(x_k)$ for $0 \le k \le N-1$. There exists $\rho(T) \in [0, \infty)$ such that for all $x \in \mathcal{T}_T$,*

$$\max_{0 \le k \le N-1} \|r_k(x)\| \le \rho(T). \tag{9}$$

The bounds (8) are standard consequences of smooth Hamiltonian dynamics combined with finite-horizon tube control, and may be derived via Grönwall-type arguments for variational equations; see, e.g., [12, 14]. Assumption 2 is natural in finite-horizon inference when the tube is chosen around candidate trajectories with bounded model mismatch.

We also fix symmetric positive definite matrices $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{r \times r}$, and a linear observation operator $\Pi_r \in \mathbb{R}^{r \times n}$ (typically a coordinate projection). Let $\mathcal{K} \subseteq \{0, 1, \dots, T\}$ denote the observation index set. Define the spectral constants

$$q_{\min} := \lambda_{\min}(Q^{-1}), \quad q_{\max} := \lambda_{\max}(Q^{-1}), \qquad r_{\min} := \lambda_{\min}(R^{-1}), \quad r_{\max} := \lambda_{\max}(R^{-1}),$$

and

$$\pi_{\min} := \lambda_{\min}(\Pi_r^\top \Pi_r), \qquad \pi_{\max} := \lambda_{\max}(\Pi_r^\top \Pi_r).$$

## III.b Gradient representation

We work on a finite time horizon of length $T > 0$ and fix a step size $h > 0$ with $N := T/h \in \mathbb{N}$. All discrete trajectories are written $x_{0:N} := (x_0, \dots, x_N) \in (\mathbb{R}^n)^{N+1}$, and observation indices are subsets $\mathcal{K} \subseteq \{0, 1, \dots, N\}$. The constants $M_j(T)$ (and derived quantities such as $\mu_{\mathrm{eff}}(T)$) depend on the *continuous* horizon length $T$, whereas $N$ is the corresponding discrete length.

Recall the action functional

$$\mathcal{J}(x_{0:N}) = \frac{1}{2}\|x_0 - \mu_0\|^2 + \frac{1}{2}\sum_{k=0}^{T-1} \|x_{k+1} - F(x_k)\|^2_{Q^{-1}} + \frac{1}{2}\sum_{k \in \mathcal{K}} \|y_k - \Pi_r x_k\|^2_{R^{-1}} + \frac{\mu}{2}\sum_{k=0}^{T} \|x_k\|^2, \tag{10}$$

with $\mu > 0$.

**Proposition 1** (Explicit gradient)**.** *For $x \in \mathcal{X}_N$, define $r_k(x) := x_{k+1} - F(x_k)$ for $0 \le k \le N-1$. Then $\mathcal{J}$ is $C^1$ and its gradient $\nabla \mathcal{J}(x) \in \mathcal{X}_N$ has blocks*

$$(\nabla \mathcal{J}(x))_0 = (x_0 - \mu_0) - DF(x_0)^\top Q^{-1} r_0(x) + \mathbf{1}_{0 \in \mathcal{K}}\, \Pi_r^\top R^{-1}(\Pi_r x_0 - y_0) + \mu x_0, \tag{11}$$

$$(\nabla \mathcal{J}(x))_k = Q^{-1} r_{k-1}(x) - DF(x_k)^\top Q^{-1} r_k(x) + \mathbf{1}_{k \in \mathcal{K}}\, \Pi_r^\top R^{-1}(\Pi_r x_k - y_k) + \mu x_k, \quad 1 \le k \le N-1, \tag{12}$$

$$(\nabla \mathcal{J}(x))_T = Q^{-1} r_{T-1}(x) + \mathbf{1}_{T \in \mathcal{K}}\, \Pi_r^\top R^{-1}(\Pi_r x_N - y_T) + \mu x_N. \tag{13}$$

*Proof.* Write $\mathcal{J} = \mathcal{J}_{\text{prior}} + \mathcal{J}_{\text{model}} + \mathcal{J}_{\text{obs}} + \mathcal{J}_{\text{reg}}$, corresponding to the four terms in (10). The prior and regularization terms are smooth with gradients

$$\nabla_{x_0}\mathcal{J}_{\text{prior}}(x) = x_0 - \mu_0, \qquad \nabla_{x_k}\mathcal{J}_{\text{reg}}(x) = \mu x_k \ \ (0 \le k \le T).$$

For the observation term,

$$\mathcal{J}_{\text{obs}}(x) = \frac{1}{2}\sum_{k\in\mathcal{K}}(\Pi_r x_k - y_k)^\top R^{-1}(\Pi_r x_k - y_k),$$

hence by direct differentiation,

$$\nabla_{x_k}\mathcal{J}_{\text{obs}}(x) = \mathbf{1}_{k\in\mathcal{K}}\,\Pi_r^\top R^{-1}(\Pi_r x_k - y_k).$$

Finally, for the model term,

$$\mathcal{J}_{\text{model}}(x) = \frac{1}{2}\sum_{k=0}^{T-1} r_k(x)^\top Q^{-1} r_k(x), \qquad r_k(x) = x_{k+1} - F(x_k).$$

Fix $k$. The block $x_k$ appears only in $r_{k-1}$ and $r_k$ (with the convention that $r_{-1}$ does not exist). Using $\nabla_u \frac{1}{2}\|u\|^2_{Q^{-1}} = Q^{-1}u$ and the chain rule,

$$\nabla_{x_k}\frac{1}{2}\|r_{k-1}(x)\|^2_{Q^{-1}} = Q^{-1}r_{k-1}(x), \qquad (1 \le k \le T),$$

and

$$\nabla_{x_k}\frac{1}{2}\|r_k(x)\|^2_{Q^{-1}} = -DF(x_k)^\top Q^{-1} r_k(x), \qquad (0 \le k \le N-1).$$

Summing the contributions yields (11)–(13). □

The block-structured gradient in Proposition 1 is standard in variational formulations of finite-horizon inference and data assimilation; see, e.g., [15, 16].

## III.c Hessian structure

We now derive the block structure of the Hessian. For brevity write $A_k := DF(x_k)$ and $B_k := D^2F(x_k)$.

**Proposition 2** (Block tridiagonal Hessian with explicit remainder)**.** *The functional $\mathcal{J}$ is of class $C^2$. Its Hessian $\nabla^2\mathcal{J}(x)$ is a symmetric block-tridiagonal operator on $\mathcal{X}_N$, whose blocks satisfy, for $0 \le k \le T$,*

$$(\nabla^2\mathcal{J}(x))_{kk} = \mu I + \mathbf{1}_{k\in\mathcal{K}}\Pi_r^\top R^{-1}\Pi_r + \mathbf{1}_{k=0} I + \mathbf{1}_{k\le N-1}\, A_k^\top Q^{-1} A_k + \mathbf{1}_{k\ge 1}\, Q^{-1} + \mathcal{R}_k(x), \tag{14}$$

*and for $0 \le k \le N-1$,*

$$(\nabla^2\mathcal{J}(x))_{k,k+1} = -A_k^\top Q^{-1}, \qquad (\nabla^2\mathcal{J}(x))_{k+1,k} = -Q^{-1}A_k. \tag{15}$$

*Here $\mathcal{R}_k(x)$ is the (symmetric) remainder operator arising from second derivatives of $F$, defined by the bilinear form*

$$\langle u, \mathcal{R}_k(x)v\rangle = -\langle B_k[u,v],\, Q^{-1}r_k(x)\rangle \qquad (0 \le k \le N-1), \tag{16}$$

*and $\mathcal{R}_T(x) \equiv 0$. Moreover,*

$$\|\mathcal{R}_k(x)\|_{\text{op}} \le q_{\max}\,\|B_k\|_{\text{op}}\ \|r_k(x)\|. \tag{17}$$

*Proof.* The prior, observation, and regularization contributions are quadratic in the blocks and hence contribute constant Hessian blocks:

$$\nabla^2_{x_0x_0}\frac{1}{2}\|x_0-\mu_0\|^2 = I, \qquad \nabla^2_{x_kx_k}\frac{1}{2}\mathbf{1}_{k\in\mathcal{K}}\|y_k-\Pi_r x_k\|^2_{R^{-1}} = \mathbf{1}_{k\in\mathcal{K}}\Pi_r^\top R^{-1}\Pi_r,$$

and $\nabla^2_{x_kx_k}\frac{\mu}{2}\sum_{j=0}^T\|x_j\|^2 = \mu I$.

It remains to differentiate the model term

$$\mathcal{J}_{\text{model}}(x) = \frac{1}{2}\sum_{k=0}^{T-1} r_k(x)^\top Q^{-1} r_k(x), \qquad r_k(x) = x_{k+1} - F(x_k).$$

Fix $k\in\{0,\dots,T-1\}$. Consider a perturbation $h\in\mathcal{X}_N$. The first variation of the $k$th summand is

$$\delta\Big(\frac{1}{2}\|r_k(x)\|^2_{Q^{-1}}\Big)[h] = \langle Q^{-1}r_k(x),\, \delta r_k(x)[h]\rangle,$$

with

$$\delta r_k(x)[h] = h_{k+1} - A_k h_k.$$

Differentiating again in direction $g\in\mathcal{X}_N$ gives the second variation

$$\delta^2\Big(\frac{1}{2}\|r_k(x)\|^2_{Q^{-1}}\Big)[h,g] = \langle Q^{-1}\delta r_k(x)[h],\, \delta r_k(x)[g]\rangle + \langle Q^{-1}r_k(x),\, \delta^2 r_k(x)[h,g]\rangle,$$

where $\delta^2 r_k(x)[h,g] = -B_k[h_k,g_k]$ because only the $F(x_k)$ term is nonlinear. Thus

$$\delta^2\Big(\frac{1}{2}\|r_k(x)\|^2_{Q^{-1}}\Big)[h,g] = \langle h_{k+1}-A_kh_k,\ Q^{-1}(g_{k+1}-A_kg_k)\rangle - \langle B_k[h_k,g_k],\, Q^{-1}r_k(x)\rangle.$$

The first term generates the block-tridiagonal contributions $Q^{-1}$ on $(k+1,k+1)$, $A_k^\top Q^{-1}A_k$ on $(k,k)$, and $-A_k^\top Q^{-1}$ and $-Q^{-1}A_k$ on the off-diagonals $(k,k+1)$ and $(k+1,k)$, respectively. The second term is exactly the remainder form (16), which contributes only to the diagonal block $(k,k)$, and $\mathcal{R}_T\equiv 0$ since no $F(x_N)$ appears. Finally, the bound (17) follows from Cauchy–Schwarz and $\|Q^{-1}\|_{\text{op}} = q_{\max}$:

$$|\langle u, \mathcal{R}_k(x)v\rangle| = |\langle B_k[u,v],\, Q^{-1}r_k(x)\rangle| \le \|B_k\|_{\text{op}}\,\|u\|\,\|v\|\,q_{\max}\,\|r_k(x)\|.$$

□

## III.d Uniform coercivity of the Hessian

We isolate the observation contribution in operator form. Define the linear observation operator

$$\mathcal{S}:\mathcal{X}_N\to(\mathbb{R}^r)^{|\mathcal{K}|}, \qquad (\mathcal{S}h)_k := \Pi_r h_k, \quad k\in\mathcal{K},$$

and the corresponding (trajectory) observation information operator

$$\mathcal{G}_{\text{obs}} := \sum_{k\in\mathcal{K}} E_k^\top\,\Pi_r^\top R^{-1}\Pi_r\,E_k = \mathcal{S}^\top(I\otimes R^{-1})\,\mathcal{S}. \tag{18}$$

Since observations are partial, $\mathcal{G}_{\text{obs}}$ may be singular. We write $\lambda^+_{\min}(\mathcal{G}_{\text{obs}})$ for its smallest strictly positive eigenvalue (equivalently, the smallest eigenvalue on range$(\mathcal{G}_{\text{obs}})$).

**Identifiability viewpoint (link to Section V).** If the quadratic regulariser $\sum_{k=0}^T\phi(x_k) = \frac{\mu}{2}\sum_{k=0}^T\|x_k\|^2$ were absent (i.e. $\mu=0$) and only partial observations were available, then $\mathcal{G}_{\text{obs}}$ would in general be singular and all curvature lower bounds would hold only on the identifiable subspace range($\mathcal{G}_{\text{obs}}$) (equivalently, on the quotient by $\ker(\mathcal{G}_{\text{obs}})$). In Section V we adopt this identifiable-subspace convention explicitly for the initial-state parameter $\theta$. The trajectory-space and parameter-space information operators are linked by a pullback: if $x(\theta)$ denotes the (discrete or continuous) flow map from an initial condition to a trajectory, then

$$\mathcal{I}(\theta) = Dx(\theta)^\top \mathcal{G}_{\text{obs}}\, Dx(\theta),$$

so that eigenvalue bounds on $\mathcal{G}_{\text{obs}}$ transfer directly to Fisher-information bounds for $\theta$.

Define the linear operator $\mathcal{B}_x : \mathcal{X}_N \to (\mathbb{R}^n)^{T+1}$ by

$$(\mathcal{B}_x h)_0 := h_0, \qquad (\mathcal{B}_x h)_{k+1} := h_{k+1} - A_k h_k, \quad 0 \le k \le N-1.$$

It is block lower-triangular with identity blocks on the diagonal, hence invertible.

**Lemma 1** (A uniform bound on $\|\mathcal{B}_x^{-1}\|$)**.** *Assume $\|A_k\|_{\text{op}} \le M_1(T)$ for all $k$. Then for every $x \in \mathcal{T}_T$,*

$$\|\mathcal{B}_x^{-1}\|_{\text{op}} \le S_T(M_1(T)) := \sum_{j=0}^{T} (M_1(T))^j. \tag{19}$$

*Consequently,*

$$\sigma_{\min}(\mathcal{B}_x) \ge \frac{1}{S_T(M_1(T))}. \tag{20}$$

*Proof.* Solve $\mathcal{B}_x h = z$ by forward substitution: $h_0 = z_0$ and $h_{k+1} = z_{k+1} + A_k h_k$. Iterating yields products of $A_k$ with $\|A_k\| \le M_1(T)$, giving the stated geometric-series bound. □

**Theorem 1** (Uniform coercivity with explicit constant)**.** *Let Assumptions 1–2 hold. Then for every $x \in \mathcal{T}_T$,*

$$\nabla^2 \mathcal{J}(x) \;\succeq\; \mu_{\text{eff}}(T)\, I_{\mathcal{X}_N},$$

*where*

$$\mu_{\text{eff}}(T) := \mu + q_{\min}\, S_T(M_1(T))^{-2} + \lambda^+_{\min}(\mathcal{G}_{\text{obs}}) - q_{\max}\, M_2(T)\, \rho(T), \tag{21}$$

*and*

*Proof.* Combine the positive model-mismatch quadratic form with the lower bound $q_{\min}\|\mathcal{B}_x h\|^2 \ge q_{\min} S_T(M_1(T))^{-2}\|h\|^2$ (Lemma 1), add the $\mu\|h\|^2$ regularization and the observation quadratic form $\langle h, \mathcal{G}_{\text{obs}} h\rangle$, and subtract the worst-case remainder bound $\|\mathcal{R}_k(x)\| \le q_{\max} M_2(T)\rho(T)$ from Proposition 2. □

**Remark 1** (Role of partial observations in $\mu_{\text{eff}}(T)$)**.** *The term $\lambda^+_{\min}(\mathcal{G}_{\text{obs}})$ in (21) captures the additional curvature contributed by observations on the identifiable subspace. When $\mu > 0$ the functional $\mathcal{J}$ is strongly convex even if $\mathcal{G}_{\text{obs}}$ is singular; the $\lambda^+_{\min}(\mathcal{G}_{\text{obs}})$ term simply sharpens the lower bound. In contrast, for purely likelihood-based formulations ($\mu = 0$) coercivity can only be asserted on* range($\mathcal{G}_{\text{obs}}$)*, consistent with the identifiable-subspace Cramér–Rao bound in Section V.*

**Remark 2** (Link to initial-condition information in Section V)**.** *Section V treats the initial condition $\theta = x_0$ as the parameter and expresses information in terms of $D\Phi_{t_k}(\theta)$. The present coercivity result is stated on the trajectory space $\mathcal{X}_T$. The two viewpoints are consistent: restricting $\mathcal{J}$ to dynamically admissible trajectories $x(\theta) = (\Phi_{t_0}(\theta), \ldots, \Phi_{t_T}(\theta))$ and applying the chain rule yields a pulled-back curvature operator on $\theta$ of the form*

$$Dx(\theta)^\top\, \nabla^2 \mathcal{J}(x(\theta))\, Dx(\theta),$$

*whose observation component coincides with the Fisher information $\sum_{k\in\mathcal{K}} (D\Phi_{t_k}(\theta))^\top \Pi_r^\top R^{-1} \Pi_r\, D\Phi_{t_k}(\theta)$.*

### III.e Third derivative bound

**Theorem 2** (Uniform third derivative bound). *Under Assumptions 1–2, for all $x \in \mathcal{T}_T$,*

$$\|\nabla^3 \mathcal{J}(x)\|_{\mathrm{op}} \le L_3(T) := q_{\max}\Big(6M_1(T)M_2(T) + 2\rho(T)M_3(T)\Big). \tag{22}$$

*Proof.* Only the model mismatch term contributes. Differentiate $\frac{1}{2}r_k^\top Q^{-1} r_k$ three times using $\delta r_k[h] = h_{k+1} - A_k h_k$, $\delta^2 r_k[h, g] = -B_k[h_k, g_k]$ and $\delta^3 r_k[h, g, \ell] = -D^3 F(x_k)[h_k, g_k, \ell_k]$, then bound each multilinear term by Cauchy–Schwarz and operator norms. □

### III.f Fourth derivative bound

**Theorem 3** (Uniform fourth derivative bound). *Under Assumptions 1–2, for all $x \in \mathcal{T}_T$,*

$$\|\nabla^4 \mathcal{J}(x)\|_{\mathrm{op}} \le L_4(T) := q_{\max}\Big(18\, M_2(T)^2 + 12\, M_1(T)M_3(T) + 2\,\rho(T)M_4(T)\Big). \tag{23}$$

*Proof.* Differentiate the third-variation expression once more. The resulting terms involve $D^2F \cdot D^2F$, $DF \cdot D^3F$, and $r_k \cdot D^4F$, which are bounded using Assumptions 1–2. □

## IV Basin-of-Attraction Geometry of High-Order Iterations

We study stationary-point iterations for the nonlinear equation

$$f(x) := \nabla \mathcal{J}(x) = 0, \qquad x \in \mathcal{X}_T, \tag{24}$$

where $\mathcal{J}$ is defined in (10). We write

$$J(x) := Df(x) = \nabla^2 \mathcal{J}(x), \qquad D^2 f(x) = \nabla^3 \mathcal{J}(x), \qquad D^3 f(x) = \nabla^4 \mathcal{J}(x).$$

### IV.a Standing hypotheses for basin analysis

Fix a tube $\mathcal{T}_T \subset \mathcal{X}_T$ as in Section III. We assume that $\mathcal{T}_T$ is star-shaped with respect to a solution $x^\star \in \mathcal{T}_T$.

**Assumption 3** (Tube-local regularity and strong monotonicity). *There exists $x^\star \in \mathcal{T}_T$ such that $f(x^\star) = 0$. Moreover, on $\mathcal{T}_T$:*

1. *(*Uniform coercivity*) $J(x) \succeq \mu_{\mathrm{eff}}(T)\, I$ for all $x \in \mathcal{T}_T$, with $\mu_{\mathrm{eff}}(T) > 0$ from* (21) *(which now includes the trajectory observation information $\lambda^+_{\min}(\mathcal{G}_{\mathrm{obs}})$ defined in* (18)*).*
2. *(*Uniform third derivative bound*) $\|D^2 f(x)\|_{\mathrm{op}} \le L_3(T)$ for all $x \in \mathcal{T}_T$, with $L_3(T)$ from* (22).
3. *(*Uniform fourth derivative bound*) $\|D^3 f(x)\|_{\mathrm{op}} \le L_4(T)$ for all $x \in \mathcal{T}_T$, with $L_4(T)$ from* (23).

**Lemma 2** (Lipschitz continuity of the Jacobian). *Under Assumption 3(ii),*

$$\|J(x) - J(y)\|_{\mathrm{op}} \le L_3(T)\, \|x - y\|, \qquad \forall x, y \in \mathcal{T}_T.$$

*Proof.* Use the fundamental theorem of calculus: $J(x) - J(y) = \int_0^1 D^2 f(y + t(x - y))[x - y]\, dt$, then take norms and apply $\|D^2 f\| \le L_3(T)$. □

**Lemma 3** (Inverse bound). *Under Assumption 3(i), $J(x)$ is invertible for all $x \in \mathcal{T}_T$ and*

$$\|J(x)^{-1}\|_{\mathrm{op}} \le \frac{1}{\mu_{\mathrm{eff}}(T)}.$$

*Proof.* Immediate from $J(x) \succeq \mu_{\mathrm{eff}}(T) I$. □

## IV.b Newton basin theorem

Newton's method for (24) is

$$x^{(m+1)} = x^{(m)} - J(x^{(m)})^{-1} f(x^{(m)}). \tag{25}$$

**Theorem 4** (Newton basin and quadratic contraction)**.** *Assume Assumption 3 and $\overline{B}(x^\star, r) \subset \mathcal{T}_T$. Let*

$$r_N(T) := \min\Big\{r,\ \frac{\mu_{\text{eff}}(T)}{2L_3(T)}\Big\}.$$

*If $\|x^{(0)} - x^\star\| \le r_N(T)$, then Newton iterates are well-defined, remain in $\overline{B}(x^\star, r_N(T))$, and converge to $x^\star$. Moreover, with $e_m := x^{(m)} - x^\star$,*

$$\|e_{m+1}\| \le \frac{L_3(T)}{2\mu_{\text{eff}}(T)} \|e_m\|^2.$$

*Proof.* Let $e := x - x^\star$ and $x \in \overline{B}(x^\star, r_N(T))$. Using $f(x^\star) = 0$,

$$J(x)e - f(x) = \int_0^1 (J(x) - J(x^\star + te))e\, dt,$$

hence by Lemma 2, $\|J(x)e - f(x)\| \le \frac{L_3(T)}{2}\|e\|^2$. Then

$$\|\mathcal{N}(x) - x^\star\| = \|J(x)^{-1}(J(x)e - f(x))\| \le \frac{1}{\mu_{\text{eff}}(T)} \cdot \frac{L_3(T)}{2}\|e\|^2.$$

Invariance follows from $\|e\| \le \mu_{\text{eff}}(T)/(2L_3(T))$. □

## IV.c Halley basin theorem

Halley's method (order 3) can be written as [17, 18]

$$x^{(m+1)} := x^{(m)} - \Big(I - \frac{1}{2} J(x^{(m)})^{-1} D^2 f(x^{(m)})[p_m]\Big)^{-1} p_m, \qquad p_m := J(x^{(m)})^{-1} f(x^{(m)}). \tag{26}$$

**Theorem 5** (Halley basin and cubic contraction)**.** *Assume Assumption 3 and $\overline{B}(x^\star, r) \subset \mathcal{T}_T$. Define*

$$\kappa(T) := \frac{\mu_{\text{eff}}(T) L_4(T)}{L_3(T)^2}, \qquad r_H(T) := \min\Big\{r,\ \frac{1}{4\sqrt{1+\kappa(T)}}\Big(\frac{\mu_{\text{eff}}(T)}{L_3(T)}\Big)^{1/2}\Big\}.$$

*If $\|x^{(0)} - x^\star\| \le r_H(T)$, then Halley iterates are well-defined, remain in $\overline{B}(x^\star, r_H(T))$, and converge to $x^\star$ with*

$$\|e_{m+1}\| \le C_H(T)\|e_m\|^3, \qquad C_H(T) := \frac{2L_3(T)^2}{\mu_{\text{eff}}(T)^2} + \frac{L_4(T)}{3\mu_{\text{eff}}(T)}.$$

*Proof.* Standard Banach-space Halley analysis using cancellation in the third-order Taylor expansion and bounds $\|J^{-1}\| \le 1/\mu_{\text{eff}}(T)$, $\|D^2 f\| \le L_3(T)$, $\|D^3 f\| \le L_4(T)$; see [17, 18]. □

## IV.d Householder order-$p$ theorem

For general order-$(p+1)$ Householder methods, one obtains basin scaling [19, 17, 18].

**Theorem 6** (Householder basin scaling)**.** *Fix $p \geq 1$. Assume $J(x) \succeq \mu_{\rm eff}(T)I$ on $\mathcal{T}_T$ and that $\|D^{\ell-1}f(x)\|_{\rm op} \leq L_\ell(T)$ for $\ell = 2, \ldots, p+1$ on $\mathcal{T}_T$. Then there exist constants $c_p, C_p > 0$ depending only on $p$ such that the order-$(p+1)$ Householder iterates converge from any $x^{(0)}$ with*

$$\|x^{(0)} - x^\star\| \leq r_p(T) := \min\Big\{r,\ c_p\Big(\frac{\mu_{\rm eff}(T)}{L_{p+1}(T)}\Big)^{1/p}\Big\},$$

*and satisfy*

$$\|x^{(m+1)} - x^\star\| \leq C_p\Big(\frac{L_{p+1}(T)}{\mu_{\rm eff}(T)}\Big)\ \|x^{(m)} - x^\star\|^{p+1}.$$

*Proof.* Classical majorant analysis for Householder methods in Banach spaces; see [19, 17, 18]. □

### IV.e Three-body instantiation of basin radii

In the planar three-body regime of Section V, the general radii $r_N(T)$, $r_H(T)$, and $r_p(T)$ can be expressed in terms of explicit physical constants. Indeed, Lemma 4 yields an explicit bound $L_1$ on $\|Df\|$ away from collisions, and Lemma 6 and (39) together with Theorem 7 provide an explicit lower bound on the observation information through the finite-time Lyapunov exponent $\Lambda(t;\theta)$.

**Corollary 1** (Three-body basin scaling in explicit constants)**.** *Assume the three-body trajectory remains in the collision-free set $\Omega$ on $[0,T]$ and that the tube $\mathcal{T}_T$ is chosen so that Assumption 3 holds. Then the Newton basin radius satisfies*

$$r_N(T)\ \gtrsim\ \frac{\mu_{\rm eff}(T)}{L_3(T)},$$

*with $\mu_{\rm eff}(T)$ given by (21) (hence containing $\lambda^+_{\min}(\mathcal{G}_{\rm obs})$) and $L_3(T)$ given by Theorem 2. Moreover, using the FTLE-controlled information bound (39) from Section V, $\lambda^+_{\min}(\mathcal{G}_{\rm obs})$ can be lower bounded explicitly by $\alpha_R\sum_{k\in\mathcal{K}}\exp(-2t_k\Lambda(t_k;\theta))$. The corresponding Halley and Householder radii obey the analogous explicit scalings*

$$r_H(T)\ \gtrsim\ \Big(\frac{\mu_{\rm eff}(T)}{L_3(T)}\Big)^{1/2}, \qquad r_p(T)\ \gtrsim\ \Big(\frac{\mu_{\rm eff}(T)}{L_{p+1}(T)}\Big)^{1/p},$$

*with constants depending only on the order.*

*Proof.* This is a direct substitution of the three-body explicit information lower bound from Section V into Theorems 4, 5, and 6. □

## V Three-Body Specialisation and Sample-Efficiency Benchmarking

The general results of Sections II, III, and IV apply to finite-horizon inference for smooth Hamiltonian systems. We now specialise the analysis to the planar three-body problem, a canonical chaotic Hamiltonian system. Its importance here is twofold: (i) it exhibits strong sensitivity to initial conditions on moderate horizons, and (ii) its Hamiltonian structure yields additional algebraic control on sensitivity matrices via symplecticity. This makes it a natural benchmark for comparing trajectory inference procedures—including learning-based predictors—at the level of provable sampling requirements.

Throughout, we treat the unknown parameter as the initial state $\theta := x_0 \in \mathbb{R}^{12}$ and consider collision-free trajectories on a finite horizon $[0,T]$. All constants are made explicit in terms of the masses, $G$, a minimal separation margin, and bounds on the observation covariance.

## V.a Planar three-body Hamiltonian and admissible set

Let $m_1, m_2, m_3 > 0$ and let $q_i(t) \in \mathbb{R}^2$, $p_i(t) \in \mathbb{R}^2$ denote positions and momenta. The Hamiltonian on $\mathbb{R}^{12}$ is

$$H(q,p) = \sum_{i=1}^{3} \frac{\|p_i\|^2}{2m_i} - \sum_{1\le i<j\le 3} \frac{Gm_im_j}{\|q_i - q_j\|}. \tag{27}$$

Write $x = (q_1, q_2, q_3, p_1, p_2, p_3) \in \mathbb{R}^{12}$ and denote by $J \in \mathbb{R}^{12\times 12}$ the canonical symplectic matrix. The continuous dynamics are $\dot{x} = J\nabla H(x)$, and we write the flow by $x(t) = \Phi_t(\theta)$.

Fix constants $d_0 > 0$ and $B_p > 0$, and define the admissible set

$$\Omega := \Big\{ x = (q,p) \in \mathbb{R}^{12} \ : \ \min_{i<j} \|q_i - q_j\| \ge d_0,\ \max_i \|p_i\| \le B_p \Big\}. \tag{28}$$

We assume the reference trajectory remains in $\Omega$ for all $t \in [0, T]$. This excludes collisions and yields uniform derivative bounds.

## V.b Explicit derivative bounds for the three-body vector field

Let $f(x) := J\nabla H(x)$ be the Hamiltonian vector field. Since the kinetic term is quadratic and the potential is a sum of Newtonian terms, all derivatives of $f$ can be bounded explicitly on $\Omega$.

Define

$$m_{\min} := \min_i m_i, \qquad m_{\max} := \max_i m_i, \qquad M_{12} := \max_{i<j} (m_im_j).$$

Let $\|\cdot\|$ denote the operator norm induced by the Euclidean norm.

**Lemma 4** (Explicit bounds on $Df$ and $D^2f$ on $\Omega$)**.** *On $\Omega$, the Jacobian $Df(x)$ satisfies*

$$\|Df(x)\| \le L_1 := \max\Big\{ m_{\min}^{-1},\ C_1 \frac{G\,M_{12}}{d_0^3} \Big\}, \qquad x \in \Omega, \tag{29}$$

*and the second derivative satisfies*

$$\|D^2 f(x)\| \le L_2 := C_2 \frac{G\,M_{12}}{d_0^4}, \qquad x \in \Omega, \tag{30}$$

*where $C_1, C_2 > 0$ are absolute constants depending only on the embedding dimension (here 2) and the number of bodies (here 3).*

*Proof.* Write $f(x) = (\dot{q}, \dot{p})$ with $\dot{q}_i = p_i/m_i$ and $\dot{p}_i = -\nabla_{q_i} U(q)$ where $U(q) = \sum_{i<j} Gm_im_j/\|q_i - q_j\|$. The kinetic contribution yields a constant block in $Df$ of size $\|\partial \dot{q}/\partial p\| = \max_i m_i^{-1} = m_{\min}^{-1}$.

For the potential term, note that for $r \in \mathbb{R}^2 \setminus \{0\}$, the map $r \mapsto r/\|r\|^3$ has Jacobian with norm bounded by $c/\|r\|^3$ for an absolute constant $c > 0$. Each pairwise interaction contributes such a term with coefficient $Gm_im_j$. Summing over the three pairs and using $\|q_i - q_j\| \ge d_0$ on $\Omega$ gives (29) with an absolute constant $C_1$ absorbing both $c$ and the finite number of pairs.

A similar calculation one derivative higher yields that the Hessian of $r \mapsto r/\|r\|^3$ is bounded by $c'/\|r\|^4$ for an absolute constant $c' > 0$. Summing again over pairs yields (30) with $C_2$ absorbing $c'$ and the finite number of pairs. □

## V.c Symplectic sensitivity and finite-time Lyapunov control

For each $t$, the Jacobian $D\Phi_t(\theta) \in \mathbb{R}^{12\times 12}$ is symplectic:

$$\big(D\Phi_t(\theta)\big)^\top J\, D\Phi_t(\theta) = J. \tag{31}$$

Consequently, the singular values of $D\Phi_t(\theta)$ occur in reciprocal pairs [20, 21]. In particular,

$$\sigma_{\min}\big(D\Phi_t(\theta)\big) = \frac{1}{\sigma_{\max}\big(D\Phi_t(\theta)\big)}. \tag{32}$$

Define the maximal finite-time Lyapunov exponent along the reference trajectory by

$$\Lambda(t;\theta) := \frac{1}{t}\log\sigma_{\max}\big(D\Phi_t(\theta)\big), \qquad t > 0, \tag{33}$$

so that

$$\sigma_{\min}\big(D\Phi_t(\theta)\big) = \exp\big(-t\,\Lambda(t;\theta)\big). \tag{34}$$

**Lemma 5** (Deterministic upper bound on $\Lambda(t;\theta)$)**.** *Assume $\Phi_s(\theta) \in \Omega$ for all $s \in [0, T]$ and let $L_1 := \sup_{x\in\Omega}\|Df(x)\|$ (bounded explicitly in Lemma 4). Then for all $t \in (0, T]$,*

$$0 \le \Lambda(t;\theta) \le L_1. \tag{35}$$

*Proof.* The variational equation gives $\frac{d}{dt}\|D\Phi_t\| \le \|Df(\Phi_t)\|\|D\Phi_t\|$, hence $\|D\Phi_t\| \le e^{\int_0^t \|Df(\Phi_s)\|ds} \le e^{L_1 t}$ by Grönwall, and therefore $\Lambda(t;\theta) = \frac{1}{t}\log\sigma_{\max}(D\Phi_t) \le \frac{1}{t}\log\|D\Phi_t\| \le L_1$. Nonnegativity follows since $\sigma_{\max}(D\Phi_0) = 1$ and $\sigma_{\max}$ is continuous. □

## V.d Observation model, identifiable subspace, and Fisher information

We observe only planar positions with Gaussian noise:

$$y_k = \Pi_q\,\Phi_{t_k}(\theta) + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, R), \qquad k \in \mathcal{K}, \tag{36}$$

where $\Pi_q : \mathbb{R}^{12} \to \mathbb{R}^6$ extracts $(q_1, q_2, q_3)$ and $R \in \mathbb{R}^{6\times 6}$ is symmetric positive definite.

Since only positions are observed, the Fisher information is generally singular on $\mathbb{R}^{12}$. All eigenvalue bounds below are interpreted on the *identifiable subspace* determined by the sampling schedule, equivalently on the quotient by the nullspace of the linearised observation operator.

**Definition 1** (Identifiable subspace and $\lambda^+_{\min}$)**.** *Define the stacked sensitivity matrix*

$$A(\theta) := \begin{bmatrix} \Pi_q D\Phi_{t_{k_1}}(\theta) \\ \vdots \\ \Pi_q D\Phi_{t_{k_m}}(\theta) \end{bmatrix} \in \mathbb{R}^{6m\times 12}, \qquad \mathcal{K} = \{k_1, \dots, k_m\}.$$

*Let $\mathcal{V}(\theta) := \mathrm{range}(A(\theta)^\top) \subseteq \mathbb{R}^{12}$ be the identifiable subspace. For a symmetric positive semidefinite matrix $M$, we write $\lambda^+_{\min}(M)$ for its smallest strictly positive eigenvalue, i.e. the minimum of $v^\top M v$ over $v \in \mathcal{V}(\theta)$ with $\|v\| = 1$.*

Let $\lambda_R^{\max} := \lambda_{\max}(R)$ and set $\alpha_R := 1/\lambda_R^{\max}$.

**Lemma 6** (Explicit Fisher information)**.** *The Fisher information matrix for $\theta$ under (36) is*

$$\mathcal{I}(\theta) = \sum_{k\in\mathcal{K}} \big(D\Phi_{t_k}(\theta)\big)^\top \Pi_q^\top R^{-1}\Pi_q D\Phi_{t_k}(\theta) = A(\theta)^\top (I_m \otimes R^{-1}) A(\theta). \tag{37}$$

*Moreover,*

$$\lambda^+_{\min}\big(\mathcal{I}(\theta)\big) \;\ge\; \alpha_R \sum_{k\in\mathcal{K}} \sigma_{\min}\big(D\Phi_{t_k}(\theta)\big)^2. \tag{38}$$

DOI: 10.5281/zenodo.21281735

*Proof.* The expression (37) follows from standard differentiation of Gaussian likelihoods with mean $\Pi_q \Phi_{t_k}(\theta)$ and covariance $R$. For the lower bound, note $\Pi_q^\top R^{-1}\Pi_q \succeq \alpha_R \Pi_q^\top \Pi_q$. For any $v \in \mathcal{V}(\theta)$, at least one block $\Pi_q D\Phi_{t_k}(\theta)v$ is nonzero by definition of $\mathcal{V}(\theta)$, and

$$v^\top \mathcal{I}(\theta) v = \sum_{k\in\mathcal{K}} \|\Pi_q D\Phi_{t_k}(\theta)v\|_{R^{-1}}^2 \geq \alpha_R \sum_{k\in\mathcal{K}} \|\Pi_q D\Phi_{t_k}(\theta)v\|^2 \geq \alpha_R \sum_{k\in\mathcal{K}} \sigma_{\min}(D\Phi_{t_k}(\theta))^2 \, \|\Pi_q v\|^2,$$

and on $\mathcal{V}(\theta)$ the restriction yields the stated bound in terms of $\lambda_{\min}^+$. □

Using (34) in (38) yields the sharper, trajectory-dependent estimate

$$\lambda_{\min}^+\big(\mathcal{I}(\theta)\big) \;\geq\; \alpha_R \sum_{k\in\mathcal{K}} \exp\big(-2t_k \Lambda(t_k;\theta)\big). \tag{39}$$

By Lemma 5, this implies the conservative bound $\lambda_{\min}^+(\mathcal{I}(\theta)) \geq \alpha_R \sum_{k\in\mathcal{K}} e^{-2L_1 t_k}$.

## V.e A self-contained Cramér–Rao inequality

We state a Cramér–Rao inequality for unbiased estimators of $\theta$ and include a complete proof.

**Theorem 7** (Cramér–Rao bound on the identifiable subspace)**.** *Let $\tilde\theta = \tilde\theta(\{y_k\}_{k\in\mathcal{K}})$ be an unbiased estimator of $\theta$, i.e. $\mathbb{E}_\theta[\tilde\theta] = \theta$, with finite covariance. Then on the identifiable subspace $\mathcal{V}(\theta)$,*

$$\mathrm{Cov}_\theta(\tilde\theta)\big|_{\mathcal{V}(\theta)} \;\succeq\; \big(\mathcal{I}(\theta)\big)^\dagger\big|_{\mathcal{V}(\theta)}, \tag{40}$$

*where $^\dagger$ denotes the Moore–Penrose pseudoinverse. Consequently,*

$$\mathbb{E}_\theta\|\tilde\theta - \theta\|^2 \;\geq\; \mathrm{tr}\big(\mathcal{I}(\theta)^\dagger\big) \;\geq\; \frac{12}{\lambda_{\min}^+\big(\mathcal{I}(\theta)\big)}. \tag{41}$$

*Proof.* Let $p_\theta(y)$ be the joint density of the observations under parameter $\theta$, and define the score $s_\theta(y) := \nabla_\theta \log p_\theta(y)$. Under Gaussian regularity, $\mathbb{E}_\theta[s_\theta] = 0$ and $\mathbb{E}_\theta[s_\theta s_\theta^\top] = \mathcal{I}(\theta)$.

Unbiasedness implies $\mathbb{E}_\theta[\tilde\theta] = \theta$. Differentiate both sides with respect to $\theta$ to obtain

$$I_{12} = \nabla_\theta \mathbb{E}_\theta[\tilde\theta] = \int \tilde\theta(y)\, \nabla_\theta p_\theta(y)\, dy = \int \tilde\theta(y)\, p_\theta(y)\, s_\theta(y)\, dy = \mathbb{E}_\theta[\tilde\theta\, s_\theta^\top].$$

Subtracting $\theta\, \mathbb{E}_\theta[s_\theta]^\top = 0$ yields

$$I_{12} = \mathbb{E}_\theta[(\tilde\theta - \theta)\, s_\theta^\top].$$

Let $A := \mathrm{Cov}_\theta(\tilde\theta)$ and $B := \mathcal{I}(\theta)$. For any $u \in \mathbb{R}^{12}$,

$$u^\top u = u^\top \mathbb{E}_\theta[(\tilde\theta - \theta)\, s_\theta^\top] u = \mathbb{E}_\theta\big[u^\top(\tilde\theta - \theta) \cdot u^\top s_\theta\big].$$

By Cauchy–Schwarz,

$$\|u\|^4 \leq (u^\top A u)\,(u^\top B u).$$

If $u \in \mathcal{V}(\theta)$ then $u^\top B u > 0$. Rearranging gives $u^\top A u \geq \|u\|^4/(u^\top B u)$ on $\mathcal{V}(\theta)$, which is equivalent to $A|_{\mathcal{V}(\theta)} \succeq B^\dagger|_{\mathcal{V}(\theta)}$, yielding (40). Taking traces and using $\mathrm{tr}(B^\dagger) \geq 12/\lambda_{\min}^+(B)$ establishes (41). □

Combining (41) with (39) yields the explicit, trajectory-dependent lower bound

$$\mathbb{E}_\theta\|\tilde\theta - \theta\|^2 \;\geq\; \frac{12}{\alpha_R \sum_{k\in\mathcal{K}} \exp(-2t_k \Lambda(t_k;\theta))}. \tag{42}$$

### V.f Explicit sampling corollary for uniform schedules

Assume uniform sampling $t_k = kh_s$ with $k = 0, \dots, N$ and $(N+1)h_s \leq T$. Using the worst-case bound $\Lambda(t_k;\theta) \leq L_1$, we obtain

$$\sum_{k=0}^{N} \exp(-2t_k\Lambda(t_k;\theta)) \;\geq\; \sum_{k=0}^{N} e^{-2L_1 k h_s} = \frac{1 - e^{-2L_1(N+1)h_s}}{1 - e^{-2L_1 h_s}}.$$

If $2L_1 h_s \leq 1$, then $1 - e^{-2L_1 h_s} \leq 2L_1 h_s$ and $1 - e^{-2L_1(N+1)h_s} \geq 1 - e^{-2L_1 T}$, yielding

$$\mathbb{E}_\theta \|\tilde{\theta} - \theta\|^2 \;\geq\; \frac{12}{\alpha_R}\,\frac{1 - e^{-2L_1 h_s}}{1 - e^{-2L_1(N+1)h_s}} \;\geq\; \frac{6L_1 h_s}{\alpha_R}\,\frac{1}{1 - e^{-2L_1 T}}\,\frac{1}{N+1}. \tag{43}$$

Here $L_1$ is explicit from (29), hence the bound depends only on $m_{\min}$, $G$, $M_{12}$, $d_0$, $\lambda_{\max}(R)$, and the sampling schedule.

### V.g Learning-based predictors as three-body estimators

Learning-based predictors for the three-body system include: (i) PINNs that parameterise trajectories and fit them by enforcing dynamics residuals and data mismatch [9, 22]; (ii) neural ODE approaches that parameterise the vector field and infer initial conditions by backpropagation through the solver [10]; and (iii) Hamiltonian/symplectic architectures that parameterise $H$ (or an equivalent structure-preserving representation) [11, 23]. When applied in the same statistical setting (36), each such method produces an estimator $\tilde{\theta}(y)$ (or an equivalent finite-dimensional parameter estimate). Therefore, all such predictors are subject to the explicit information bound (42) and its uniform-sampling consequence (43): no estimator can circumvent the sampling threshold dictated by the three-body sensitivity spectrum encoded in $\Lambda(t;\theta)$ and the collision-free constants in (29).

## VI Discussion and Conclusion

We have developed a quantitative basin-geometry theory for high-order stationary-point iterations applied to finite-horizon trajectory inference in chaotic Hamiltonian systems. The central technical step was to express the variational objective as an action functional induced by a (symplectic) discrete-time Hamiltonian flow, and then to derive explicit bounds on its second, third, and fourth derivatives on a collision-free tube of trajectories. These bounds yield coercivity on the identifiable subspace and, in turn, explicit basin-of-attraction radii for Newton, Halley, and general Householder iterations.

**Forward instability versus inverse conditioning.** A recurring source of confusion in chaotic mechanics is the conflation of forward-time instability with ill-posedness of inverse trajectory problems. Our results separate these effects: forward chaos manifests through the growth of derivative bounds (equivalently, the stretching spectrum of the flow), while inverse conditioning is controlled by the competition between (i) coercivity of the induced action functional on the identifiable subspace and (ii) nonlinear distortion measured by higher-order derivatives. This separation is made explicit in the basin radii, which scale as ratios of curvature constants to higher-order derivative bounds, and in the sampling-dependent information lower bounds, which depend on finite-time Lyapunov growth along the reference trajectory.

**Why higher-order methods enlarge admissible initialization.** Newton's method is governed by a third-derivative distortion constant, whereas Halley and higher-order Householder methods additionally exploit cancellation of lower-order error terms. The resulting admissible radii scale as $r_N(T) \asymp \mu_{\text{eff}}(T)/L_3(T)$ for Newton and $r_H(T) \asymp (\mu_{\text{eff}}(T)/L_3(T))^{1/2}$ for Halley, with the general Householder radius $r_p(T) \asymp (\mu_{\text{eff}}(T)/L_{p+1}(T))^{1/p}$. In chaotic regimes, $L_{p+1}(T)$ grows with horizon length, so these formulas make precise both (a) the benefit of higher order at fixed $T$ and (b) the inevitable basin shrinkage as $T$ increases.

**Three-body specialisation and sampling efficiency.** For the planar three-body problem, symplecticity implies the reciprocal-pair structure of singular values of $D\Phi_t$, yielding $\sigma_{\min}(D\Phi_t) = \sigma_{\max}(D\Phi_t)^{-1}$ and therefore an information lower bound in terms of finite-time Lyapunov exponents. This gives a sharp and physically interpretable obstruction: no estimator can surpass the sample-efficiency threshold dictated by the sensitivity spectrum of the flow and the observation covariance. In particular, the analysis identifies how sampling schedules must compensate for instability to maintain non-degenerate information on the identifiable subspace.

**Limitations and extensions.** Our analysis assumes collision-free trajectories on a compact admissible tube, ensuring uniform derivative control. Extending the theory to allow close encounters would require a refined local analysis with state-dependent constants. Second, we treat fixed observation operators and Gaussian noise; more general partial observations and heavy-tailed noise can be incorporated by modifying the observation term and repeating the coercivity analysis on the corresponding identifiable subspace. Third, while we analysed sampling density in Appendix A, an important direction is *adaptive* sampling: selecting measurement times to maximise $\lambda^{+}_{\min}(\mathcal{G}_{\text{obs}})$ subject to constraints. Finally, although the three-body problem is a canonical chaotic benchmark, the framework applies verbatim to other smooth Hamiltonian systems and to other structure-preserving discretizations, with constants computed from the corresponding vector-field derivatives and sensitivity growth.

**Conclusion.** The main message is that basin geometry in chaotic Hamiltonian inference is governed by explicit, computable constants: coercivity on the identifiable subspace competes against higher-order distortion induced by flow sensitivity. High-order Householder iterations provide provably larger admissible initialization radii than Newton, but no method can avoid the fundamental sampling barrier imposed by finite-time Lyapunov growth. These results supply a rigorous, mechanism-level explanation for observed instability in practical trajectory reconstruction, and provide a principled basis for designing solvers and sampling schedules in chaotic mechanical systems.

# A Appendix A: Sampling Density and Gramian Lower Bounds

## A.a Setting and notation

Let $n \in \mathbb{N}$ and consider a discrete-time dynamical system

$$x_{k+1} = F(x_k), \qquad k = 0, 1, \dots, N-1, \tag{44}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is $C^2$ on a neighbourhood of a reference trajectory $\bar{x}_{0:N} = (\bar{x}_0, \dots, \bar{x}_N)$.

Fix a linear observation operator $\mathcal{O} \in \mathbb{R}^{r\times n}$ and a positive definite covariance $R \in \mathbb{R}^{r\times r}$. We observe at indices $\mathcal{K} = \{k_1 < \cdots < k_m\} \subseteq \{0, 1, \dots, N\}$ with Gaussian noise:

$$y_{k_i} = \mathcal{O}x_{k_i} + \eta_{k_i}, \qquad \eta_{k_i} \sim \mathcal{N}(0, R) \quad \text{independent.} \tag{45}$$

Define the state transition Jacobians along $\bar{x}_{0:N}$ by

$$\Phi_{j:k} := \begin{cases} DF(\bar{x}_{k-1})\, DF(\bar{x}_{k-2}) \cdots DF(\bar{x}_j), & j < k, \\ I, & j = k, \end{cases} \tag{46}$$

so that the linearized perturbation obeys $\delta x_k = \Phi_{0:k}\, \delta x_0$.

## A.b Observation Gramian

The (discrete) observation Gramian on the initial condition is

$$\mathcal{G}_{\rm obs} := \sum_{i=1}^{m} \Phi_{0:k_i}^{\top}\, \mathcal{O}^{\top} R^{-1} \mathcal{O}\, \Phi_{0:k_i} \ \in\ \mathbb{R}^{n\times n}. \tag{47}$$

It is symmetric positive semidefinite. Its smallest strictly positive eigenvalue $\lambda_{\min}^{+}(\mathcal{G}_{\rm obs})$ governs curvature on the *identifiable* subspace range$(\mathcal{G}_{\rm obs})$.

For any $v \in \mathbb{R}^n$,

$$v^{\top}\mathcal{G}_{\rm obs}v = \sum_{i=1}^{m} \|\mathcal{O}\Phi_{0:k_i}v\|_{R^{-1}}^2. \tag{48}$$

## A.c A density-based lower bound

We provide a lower bound that is explicit in the sampling schedule. Assume there exist constants $0 < \underline{\sigma} \le \overline{\sigma} < \infty$ such that for all $k \in \{0, \dots, N\}$,

$$\underline{\sigma}\,\|v\| \le \|\Phi_{0:k}v\| \le \overline{\sigma}\,\|v\| \qquad \forall v \in \mathbb{R}^n. \tag{49}$$

Assume also that the observation operator has a uniform lower bound on a fixed subspace $\mathcal{U} \subseteq \mathbb{R}^n$, i.e.,

$$\|\mathcal{O}u\| \ge \kappa\,\|u\| \qquad \forall u \in \mathcal{U}, \tag{50}$$

for some $\kappa > 0$.

Define the *propagated* subspaces $\mathcal{U}_k := \Phi_{0:k}^{-1}\mathcal{U}$ and suppose that for the chosen sampling set $\mathcal{K}$ there exists a subspace $\mathcal{V} \subseteq \mathbb{R}^n$ and an angle parameter $\rho \in (0, 1]$ such that for every $v \in \mathcal{V}$ and every $i = 1, \dots, m$,

$$\|P_{\mathcal{U}}\Phi_{0:k_i}v\| \ge \rho\,\|\Phi_{0:k_i}v\|, \tag{51}$$

where $P_{\mathcal{U}}$ denotes orthogonal projection onto $\mathcal{U}$.

**Lemma 7** (Gramian lower bound on a controlled subspace)**.** *Let* (49)–(51) *hold and set* $\alpha_R := \lambda_{\min}(R^{-1})$*. Then for all* $v \in \mathcal{V}$*,*

$$v^{\top}\mathcal{G}_{\rm obs}v \ \ge\ m\,\alpha_R\,(\kappa\rho)^2\,\underline{\sigma}^2\,\|v\|^2. \tag{52}$$

*Consequently,* $\lambda_{\min}\big(\mathcal{G}_{\rm obs}\big|_{\mathcal{V}}\big) \ge m\,\alpha_R\,(\kappa\rho)^2\,\underline{\sigma}^2$.

*Proof.* Fix $v \in \mathcal{V}$. Using (48) and $\alpha_R I \preceq R^{-1}$,

$$v^{\top}\mathcal{G}_{\rm obs}v = \sum_{i=1}^{m} \|\mathcal{O}\Phi_{0:k_i}v\|_{R^{-1}}^2 \ge \alpha_R \sum_{i=1}^{m} \|\mathcal{O}\Phi_{0:k_i}v\|^2.$$

By (51) and (50),

$$\|\mathcal{O}\Phi_{0:k_i}v\| \ge \|\mathcal{O}P_{\mathcal{U}}\Phi_{0:k_i}v\| \ge \kappa\,\|P_{\mathcal{U}}\Phi_{0:k_i}v\| \ge \kappa\rho\,\|\Phi_{0:k_i}v\|.$$

Finally, (49) gives $\|\Phi_{0:k_i}v\| \ge \underline{\sigma}\|v\|$ for all $i$. Summing yields (52). $\square$

### A.d Interpretation

Lemma 7 isolates three mechanisms: (i) measurement strength $(\alpha_R, \kappa)$, (ii) alignment of propagated directions with the observed subspace $(\rho)$, and (iii) expansion/contraction of the linearized flow $(\underline{\sigma})$. Once the alignment condition (51) is ensured along the sampled indices, the sampling density enters through $m = |\mathcal{K}|$.

This appendix is self-contained: all quantities and assumptions used to bound $\mathcal{G}_{\mathrm{obs}}$ are defined above, and no external references are required.

## B Appendix B: Explicit Derivative Expansions for the Discrete Action

### B.a Discrete action functional

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be $C^4$ on a neighbourhood of a reference tube. Fix symmetric positive definite matrices $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{r\times r}$ and a linear observation operator $\mathcal{O} \in \mathbb{R}^{r\times n}$. Let $\mathcal{K} \subseteq \{0, 1, \ldots, N\}$.

For $x_{0:N} = (x_0, \ldots, x_N) \in (\mathbb{R}^n)^{N+1}$ define the discrete action

$$\mathcal{J}(x_{0:N}) = \frac{1}{2}\|x_0 - \mu_0\|^2 + \frac{1}{2}\sum_{k=0}^{N-1} \|x_{k+1} - F(x_k)\|^2_{Q^{-1}} + \frac{1}{2}\sum_{k\in\mathcal{K}} \|y_k - \mathcal{O}x_k\|^2_{R^{-1}} + \frac{\mu}{2}\sum_{k=0}^{N} \|x_k\|^2, \quad (53)$$

with $\mu > 0$.

Write the model residuals and observation residuals as

$$r_k := x_{k+1} - F(x_k) \in \mathbb{R}^n, \qquad e_k := y_k - \mathcal{O}x_k \in \mathbb{R}^r. \quad (54)$$

### B.b Gradient

Let $A_k := DF(x_k) \in \mathbb{R}^{n\times n}$. A direct differentiation of (53) yields the block gradient:

$$\nabla_{x_0}\mathcal{J} = (x_0 - \mu_0) + \mu x_0 - A_0^\top Q^{-1} r_0, \quad (55)$$

$$\nabla_{x_j}\mathcal{J} = \mu x_j + Q^{-1} r_{j-1} - A_j^\top Q^{-1} r_j - \mathbf{1}_{\{j\in\mathcal{K}\}}\mathcal{O}^\top R^{-1} e_j, \qquad 1 \le j \le N-1, \quad (56)$$

$$\nabla_{x_N}\mathcal{J} = \mu x_N + Q^{-1} r_{N-1} - \mathbf{1}_{\{N\in\mathcal{K}\}}\mathcal{O}^\top R^{-1} e_N. \quad (57)$$

### B.c Hessian: block tridiagonal structure

Let $H_k(\cdot,\cdot) := D^2F(x_k)[\cdot,\cdot]$ be the bilinear second derivative. The Hessian $\nabla^2\mathcal{J}(x_{0:N})$ is block tridiagonal. For $1 \le j \le N-1$, the diagonal block is

$$\nabla^2_{x_j x_j}\mathcal{J} = \mu I + Q^{-1} + A_j^\top Q^{-1} A_j - \mathcal{T}_j + \mathbf{1}_{\{j\in\mathcal{K}\}}\mathcal{O}^\top R^{-1}\mathcal{O}, \quad (58)$$

where $\mathcal{T}_j$ is the linear operator defined by

$$\mathcal{T}_j[v] := \big(D\big(A_j^\top Q^{-1} r_j\big)[v]\big) = H_j(v,\cdot)^\top (Q^{-1} r_j). \quad (59)$$

The off-diagonal blocks are

$$\nabla^2_{x_j x_{j-1}}\mathcal{J} = -A_{j-1}^\top Q^{-1}, \qquad \nabla^2_{x_j x_{j+1}}\mathcal{J} = -Q^{-1}, \quad (60)$$

with symmetry giving the transposed counterparts.

### B.d Third and fourth derivative bounds

Assume there exist constants $M_1, M_2, M_3, M_4$ such that on the tube

$$\|DF(x)\| \leq M_1, \quad \|D^2F(x)\| \leq M_2, \quad \|D^3F(x)\| \leq M_3, \quad \|D^4F(x)\| \leq M_4. \tag{61}$$

Then $\nabla^3\mathcal{J}$ and $\nabla^4\mathcal{J}$ exist and are supported on at most four consecutive time indices (block sparsity). Moreover, there exist constants $L_3, L_4$ depending only on

$$(M_1, M_2, M_3, M_4),\ \|Q^{-1}\|,\ \|R^{-1}\|,\ \|\mathcal{O}\|,\ \mu,\ N,$$

such that

$$\|\nabla^3\mathcal{J}(x_{0:N})\| \leq L_3, \qquad \|\nabla^4\mathcal{J}(x_{0:N})\| \leq L_4. \tag{62}$$

A conservative explicit choice is

$$L_3 \leq c_3\,\|Q^{-1}\|\,(M_2 + M_1M_2 + M_3)\ +\ c_3'\,\|R^{-1}\|\,\|\mathcal{O}\|^3, \tag{63}$$
$$L_4 \leq c_4\,\|Q^{-1}\|\,(M_3 + M_2^2 + M_1M_3 + M_4)\ +\ c_4'\,\|R^{-1}\|\,\|\mathcal{O}\|^4, \tag{64}$$

for absolute constants $c_3, c_3', c_4, c_4'$.

### B.e Self-containment

All quantities used above are defined within this appendix: the action functional (53), residuals (54), and derivative bounds (61). No reference to experiments or to any external section is required.

## References

[1] Henri Poincaré. *Les Méthodes Nouvelles de la Mécanique Céleste*. Gauthier-Villars, 1892.

[2] Carles Simó. New families of solutions in n-body problems. *European Congress of Mathematics*, pages 101–115, 2001.

[3] Angshul Majumdar. Hinduism as an open normative system: An axiomatic and model-theoretic analysis. *Natural Sciences and Applied Technology*, 3(1):134 – 149, 2026. Cited by: 0.

[4] Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, 15:9–20, 1980.

[5] Kody Law, Andrew Stuart, and Konstantinos Zygalakis. *Data Assimilation: A Mathematical Introduction*. Springer, 2015.

[6] Andrew M. Stuart. Inverse problems: A bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[7] Peter Deuflhard. *Newton Methods for Nonlinear Problems*. Springer, 2011.

[8] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration*. Springer, 2006.

[9] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[10] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, 2018.

[11] Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, pages 15353–15363, 2019.

[12] V. I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer, 2 edition, 1989.

[13] Jerrold E. Marsden and Matthew West. Discrete mechanics and variational integrators. *Acta Numerica*, 10:357–514, 2001.

[14] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, 2 edition, 2006.

[15] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A*, 38A(2):97–110, 1986.

[16] Kody Law, Andrew M. Stuart, and Konstantinos C. Zygalakis. *Data Assimilation: A Mathematical Introduction*. Springer, 2015.

[17] Joseph F. Traub. *Iterative Methods for the Solution of Equations*. Prentice-Hall, 1964.

[18] James M. Ortega and Werner C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, 1970.

[19] Alston S. Householder. *The Numerical Treatment of a Single Nonlinear Equation*. McGraw-Hill, 1970.

[20] Maurice A. de Gosson. *Symplectic Geometry and Quantum Mechanics*. Birkhäuser, 2006.

[21] C. C. Paige and C. F. Van Loan. A schur decomposition for hamiltonian matrices. *Linear Algebra and its Applications*, 41:11–32, 1981.

[22] Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021.

[23] Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. Symplectic ode-net: Learning hamiltonian dynamics with control. In *International Conference on Learning Representations*, 2020.

**Conflict of interest:** The Authors have no conflicts of interest to declare that they are relevant to the content of this article.