INVARIANT MEASURE OF HIGH-DIMENSIONAL SDES

YITONG QIU

Contents

1. Introduction

2. Preliminaries

2.1. **The definitions of Uniform Large Deviation Principle (ULDP).** Consider a Polish space $(\mathcal{E}, \rho)$ alongside a topological space $\mathcal{E}_0$ used for indexing, a function $I : \mathcal{E} \longrightarrow [0, +\infty]$ is called a rate function when it always has compact level sets, i.e. for every finite M, the set $\{\varphi : I(\varphi) \leq M\}$ is compact.

Given any $\varphi \in \mathcal{E}$ and a subset $K \subset \mathcal{E}$, denote the distance from φ to K as

$$\operatorname{dist}(\varphi, K) := \inf_{\psi \in K} \rho(\varphi, \psi).$$

For an element $x \in \mathcal{E}_0$, we use $I_x(\cdot)$ to show the rate function depends on x. Let $\mathscr{K}$ represent the collection of all compact subsets within $\mathcal{E}_0$, and let $\Phi_x(s) \doteq \{\varphi : I_x(\varphi) \leq s\}$ denote the set of elements satisfying the rate condition with respect to x under value s.

Definition 2.1. *A family of rate functions $\{I_x\}_{x \in \mathcal{E}_0}$ on $\mathcal{E}$ has compact level sets on compact subsets of $\mathcal{E}_0$ if, for any compact K and for every $s \in [0, \infty)$, the union $\cup_{x \in K} \Phi_x(s)$ forms a compact subset of $\mathcal{E}$.*

Definition 2.2 (Freidlin-Wentzell ULDP). *Let $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ be a family of $\mathcal{E}$-valued random variables indexed by $x \in \mathcal{E}_0$. It is said to satisfy a Uniform Large Deviation Principle (ULDP) with respect to the rate function I_x, $x \in \mathcal{E}_0$, uniformly over $\mathscr{K}$, if the following conditions hold:*

(1) *LDP Lower Bound: For any $K \in \mathscr{K}$, $\delta > 0$, $\gamma > 0$, and $s_0 > 0$, there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$, $x \in K$, and $\varphi \in \Phi_x(s_0)$,*

$$\mathbb{P}\left(\rho\left(X^{\varepsilon,x}, \varphi\right) < \delta\right) \geq \exp\left(-\varepsilon^{-1}\left(I_x(\varphi) + \gamma\right)\right)$$

(2) *LDP Upper Bound: For any $K \in \mathscr{K}$, $\delta > 0$, $\gamma > 0$, and $s_0 > 0$, there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$, $s \leq s_0$, and $x \in K$,*

$$\mathbb{P}\left(\operatorname{dist}\left(X^{\varepsilon,x}, \Phi_x(s)\right) \geq \delta\right) \leq \exp\left(-\varepsilon^{-1}(s - \gamma)\right)$$

The ULDP is a strengthened version of the Large Deviation Principle (LDP), requiring that the lower and upper bounds hold uniformly over $\mathscr{K}$. The lower bound guarantees that the probability of the random variable staying close to a target path is not too small, while the upper bound limits the probability of it deviating too far from the target path.

Another similar definition is the uniform Laplace principle, which describes the expected behavior of random variables in the limit case.

Definition 2.3 (Uniform Laplace Principle). *Let $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ be a family of $\mathcal{E}$-valued random variables indexed by $x \in \mathcal{E}_0$. It is said to satisfy the uniform Laplace principle on $\mathscr{K}$ with respect to the rate function I_x, $x \in \mathcal{E}_0$, if for any compact set $K \in \mathscr{K}$ and any bounded continuous function $h : \mathcal{E} \longrightarrow \mathbb{R}$, the following condition holds:*

$$\lim_{\varepsilon \to 0} \sup_{x \in K} \left| \varepsilon \log \mathbb{E} \exp\left(-\varepsilon^{-1} h\left(X^{\varepsilon,x}\right)\right) + \inf_{\varphi \in \mathcal{E}} \left\{h(\varphi) + I_x(\varphi)\right\} \right| = 0$$

2.2. **ULDP for solutions of Stochastic Differential Equations (SDEs).** Let's first clarify the notations. Let $\left(\mathbb{R}^d, \langle\cdot,\cdot\rangle, |\cdot|\right)$ represent the d-dimensional Euclidean space equipped with the inner product $\langle\cdot,\cdot\rangle$, which induces the Euclidean norm $|\cdot|$. The notation $\|\cdot\|$ denotes the Hilbert-Schmidt norm defined as $\|\sigma\|^2 := \sum_{i=1}^d \sum_{j=1}^m \sigma_{ij}^2$ for any $d\times m$ matrix $\sigma = (\sigma_{ij}) \in \mathbb{R}^d \otimes \mathbb{R}^m$. Here, $\sigma^\top$ represents the transpose of the matrix σ.

Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying right-continuity and completeness. $\{B(t)\}_{t\geq 0}$ is an m-dimensional Brownian motion defined on this probability space. Fix $T\in(0,\infty)$. Consider the following stochastic differential equations:

$$dX^\varepsilon(t) = b\left(X^\varepsilon(t)\right)dt + \sqrt{\varepsilon}\sigma\left(X^\varepsilon(t)\right)dB(t), t\in[0,T], \quad x(0) = x \in \mathbb{R}^d \tag{2.1}$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ are continuous and we use the notation $X^{\varepsilon,x}$ to indicate the solution of (2.1) starting from x.

Let's first introduce two classical assumptions.

Assumption 1. *Let $\varepsilon_0 \in (0,1)$. For arbitrary $R > 0$, if $|x| \vee |y| \leq R$, there exists $L_R > 0$ such that the following locally monotonicity condition*

$$2\langle x-y, b(x)-b(y)\rangle + \|\sigma(x)-\sigma(y)\|^2 \leq L_R|x-y|^2 \tag{2.2}$$

holds for $|x-y| \leq \varepsilon_0$.

Assumption 2. *There exists a Lyapunov function $V \in C^2(\mathbb{R}^d;\mathbb{R}_+)$ along with constants $\theta > 0$ and $\eta > 0$ such that*

$$\lim_{|x|\to+\infty} V(x) = +\infty$$

$$J(x) := \langle b(x), \nabla V(x)\rangle + \frac{\theta}{2}\mathrm{Trace}(\sigma^*(x)\nabla^2 V(x)\sigma(x)) + \frac{|\sigma^*(x)\cdot\nabla V(x)|^2}{\eta V(x)} \leq C(1+V(x)),$$

and

$$\mathrm{Trace}(\sigma^*(x)\nabla^2 V(x)\sigma(x)) \geq -M - CV(x).$$

Here, ∇V and $\nabla^2 V$ denote the gradient vector and Hessian matrix of the function V, respectively. $C, M > 0$ are some fixed constants.

These two assumptions are considered classical because they jointly ensure the existence and uniqueness of a global solution to the (2.1). Assumption 1 guarantees that the solution is locally uniquely existent, while Assumption 2 ensures that this local solution can be uniquely extended to become the unique global solution.

Proposition 2.4. *For any $0 < \varepsilon < 1$, under Assumptions 1 and 2, there exists a unique solution to (2.1) defined on $[0,+\infty)$.*

For each $h \in L^2([0,T],\mathbb{R}^m)$, consider the following ODE (also known as skeleton equation):

$$dX_x^h(t) = b\left(X_x^h(t)\right)dt + \sigma\left(X_x^h(t)\right)h(t)dt \tag{2.3}$$

with the initial value $X_x^h(0) = x$. We have the similar result:

Proposition 2.5. *Under Assumptions 1 and 2, there always exists a unique solution to (2.3).*

We can now finally introduce the ULDP for the solutions of the SDE (2.1). We first need to define the required rate function. For any $x \in \mathbb{R}^d$ and $f \in C\left([0,T],\mathbb{R}^d\right)$, we define

$$I_x(f) = \inf_{\left\{h\in L^2([0,T];\mathbb{R}^m): f = X_x^h\right\}} \left\{\frac{1}{2}\int_0^T |h(s)|^2 ds\right\}, \tag{2.4}$$

and we adopt the convention that $\inf\{\emptyset\} = \infty$, here $X_x^h \in C\left([0,T],\mathbb{R}^d\right)$ is the solution to (2.3).

We now formulate the main result on ULDP for solutions of Stochastic Differential Equations.

Theorem 2.6. *For $\varepsilon > 0$, let $X^{\varepsilon,x}$ be the solution to (2.1). Suppose Assumptions 1 and 2 are satisfied, and rate functions I_x are defined by (2.4) on $C\left([0,T],\mathbb{R}^d\right)$. Then the following conclusions hold:*

(1) The family $\left\{I_x, x\in\mathbb{R}^d\right\}$ of rate functions has compact level sets on compacts.
(2) The solutions $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ satisfies a ULDP on the space $C\left([0,T],\mathbb{R}^d\right)$ with the rate function I_x, uniformly over the initial value x in bounded subsets of $\mathbb{R}^d$.

Remark 2.7. *The conclusions of Freidlin and Wentzell on ULDP are based on strict assumptions such as **bounded coefficients** or **linear growth drift**. They are corollaries of Theorem 2.6 with the Lyapunov function $V(x) = |x|^2 + 1$.*

2.3. **A Sufficient Condition for ULDP.** In order to derive a more general ULDP criterion, we need a sufficient condition as a discriminant criterion.

Let $\mathcal{E} = C\left([0,T], \mathbb{R}^d\right).\rho(\cdot,\cdot)$ stands for the uniform metric in the space $\mathcal{E}$ and $\mathcal{E}_0 = \mathbb{R}^d$. The first result shows that uniform Laplace principle implies the ULDP.

Theorem 2.8. *Let I_x be a family of rate functions on $\mathcal{E}$ indexed by x in $\mathcal{E}_0$ and assume that this family has compact level sets on compacts. Let $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ be a sequence of $\mathcal{E}$-valued random variables and suppose that $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ satisfies a uniform Laplace principle with rate function I_x uniformly over any element in $\mathscr{K}$. Then $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ satisfies ULDP with rate function I_x uniformly over any element in $\mathscr{K}$.*

The following result provides a criterion for the uniform Laplace principle. Let

$$S^N := \left\{h \in L^2\left([0,T], \mathbb{R}^m\right) : |h|^2_{L^2([0,T],\mathbb{R}^m)} \le N\right\}$$

and

$$\tilde{S}^N := \left\{\phi : \phi \text{ is an } \mathbb{R}^m\text{-valued } \mathcal{F}_t\text{-predictable process, such that } \phi(\omega) \in S^N, \mathbb{P}\text{-almost surely}\right\}.$$

Then we consider the concept of weak convergence within the space $L^2\left([0,T], \mathbb{R}^m\right)$ specifically on the set S^N, this set S^N is endowed with a topology that makes it a Polish space.

For any $\varepsilon > 0$, let $\Gamma^\varepsilon : \mathcal{E}_0 \times C\left([0,T], \mathbb{R}^m\right) \to \mathcal{E}$ be a measurable mapping. Define $X^{\varepsilon,x} := \Gamma^\varepsilon(x, B(\cdot))$. Then we have the following result.

Theorem 2.9. *Suppose there exists a measurable mapping $\Gamma^0 : \mathcal{E}_0 \times C\left([0,T], \mathbb{R}^m\right) \to \mathcal{E}$, and let*

$$I_x(f) = \inf_{\left\{h \in L^2([0,T];\mathbb{R}^m) : f = \Gamma^0\left(x, \int_0^\cdot h(s)ds\right)\right\}} \left\{\frac{1}{2}\int_0^T |h(s)|^2 ds\right\}. \tag{2.5}$$

Assume that for all $f \in \mathcal{E}$, the mapping $x \mapsto I_x(f)$ is a lower semi-continuous (l.s.c.) map from $\mathcal{E}_0$ to $[0,\infty]$, and the following conditions hold:

(1) For each $N < \infty$ and $K \in \mathscr{K}$, the set

$$\Lambda_{N,K} \doteq \left\{\Gamma^0\left(x, \int_0^\cdot h(s)ds\right) : h \in S^N, x \in K\right\}$$

is a compact subset of $\mathcal{E}$;

(2) For each $N < \infty$ and any family $\{h^\varepsilon\} \subset \tilde{S}^N$ and $\{x^\varepsilon\} \subset \mathcal{E}_0$ satisfying $x^\varepsilon \to x$ and h^ε converges in distribution to some element h as $\varepsilon \to 0$, we have that $\Gamma^\varepsilon\left(x^\varepsilon, B(\cdot) + \frac{1}{\sqrt{\varepsilon}}\int_0^\cdot h^\varepsilon(s)ds\right)$ converges in distribution to $\Gamma^0\left(x, \int_0^\cdot h(s)ds\right)$ as $\varepsilon \to 0$.

Then for all $x \in \mathcal{E}_0$, I_x is a rate function on $\mathcal{E}$, the family of rate functions $\{I_x, x \in \mathcal{E}_0\}$ has compact level sets on compact sets, and $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ satisfies a uniform Laplace principle with respect to the rate function I_x, uniformly over $\mathscr{K}$.

Next we present a sufficient condition for verifying the assumptions in Theorem 2.9.

Theorem 2.10. *Suppose that there exists a measurable map $\Gamma^0 : \mathcal{E}_0 \times C\left([0,T], \mathbb{R}^m\right) \to \mathcal{E}$ such that*

(1) for every $N < +\infty, x_n \to x$ and any family $\{h_n, n \in \mathbb{N}\} \subset S^N$ converging weakly to some element h as $n \to \infty, \Gamma^0\left(x_n, \int_0^\cdot h_n(s)ds\right)$ converges to $\Gamma^0\left(x, \int_0^\cdot h(s)ds\right)$ in the space $C\left([0,T], \mathbb{R}^d\right)$;

(2) for every $N < +\infty$, $\{x^\varepsilon, \varepsilon > 0\} \subset \{x : |x| \le N\}$ and any family $\{h^\varepsilon, \varepsilon > 0\} \subset \tilde{S}^N$ and any $\delta > 0$,

$$\lim_{\varepsilon \to 0} \mathbb{P}\left(\rho\left(Y^{\varepsilon,x^\varepsilon}, Z^{\varepsilon,x^\varepsilon}\right) > \delta\right) = 0$$

where $Y^{\varepsilon,x^\varepsilon} = \Gamma^\varepsilon\left(x^\varepsilon, B(\cdot) + \frac{1}{\sqrt{\varepsilon}}\int_0^\cdot h^\varepsilon(s)ds\right)$ and $Z^{\varepsilon,x^\varepsilon} = \Gamma^0\left(x^\varepsilon, \int_0^\cdot h^\varepsilon(s)ds\right)$.

Then for all $x \in \mathcal{E}_0, I_x$ defined by (2.5) is a rate function on $\mathcal{E}$, the family $\{I_x, x \in \mathcal{E}_0\}$ of rate functions has compact level sets on compacts and $\{X^{\varepsilon,x}\}_{\varepsilon>0}$ satisfies a uniform Laplace principle with the rate function I_x uniformly over $\mathscr{K}$.

3. The concentration of limiting measures and metastability

From now on, we will denote by $\mathbf{C}_T$ the space of continuous functions from $[0,T]$ to $\mathbb{R}^n$. The space $\mathbf{C}_T^x(\mathbf{AC}_T^x)$ will refer to the set of (absolutely) continuous functions on $[0,T]$ that start at x and take values in $\mathbb{R}^n$.

We now consider the small noise perturbation, where the system is driven by m independent Brownian motions:

$$\begin{cases} dX_\varepsilon^{(i)}(t) = b_i(X_\varepsilon^{(0)}(t), \dots, X_\varepsilon^{(n-1)}(t))dt, & i = 0, 1, \dots, s, \\ dX_\varepsilon^{(i)}(t) = b_i(X_\varepsilon^{(0)}(t), \dots, X_\varepsilon^{(n-1)}(t))dt + \sqrt{\varepsilon}\sum_{k=1}^m \sigma_{i,k}(X_\varepsilon^{(0)}(t), \dots, X_\varepsilon^{(n-1)}(t))dB_k, & i = s+1, s+2, \dots, n-1. \end{cases} \tag{3.1}$$

The system of stochastic differential equations can be expressed compactly in the following matrix form:

$$d\mathbf{X}_\varepsilon(t) = \mathbf{b}(\mathbf{X}_\varepsilon(t))dt + \sqrt{\varepsilon}\tilde{\sigma}(\mathbf{X}_\varepsilon(t))d\mathbf{B}(t), \tag{3.2}$$

where the terms are defined as follows:

- $\mathbf{X}_\varepsilon(t) = \left(X_\varepsilon^{(0)}(t), \dots, X_\varepsilon^{(n-1)}(t)\right)^{\text{tr}}$ is the n-dimensional state vector. Here, $X_\varepsilon^{(i)}$ for $i \geq 1$ represents the i-th derivative of the base state $X_\varepsilon^{(0)}$.
- $\mathbf{b}\left(\mathbf{X}_\varepsilon(t)\right) = \left(b_0\left(\mathbf{X}_\varepsilon\right), \dots, b_{n-1}\left(\mathbf{X}_\varepsilon\right)\right)^{\text{tr}}$ is the drift vector. Its scalar components b_i are real-valued, locally Lipschitz continuous functions on $\mathbb{R}^n$.
- $\mathbf{B}(t) = (B_1(t), \dots, B_m(t))^{\text{tr}}$ is a vector of m independent standard Brownian motions, and $d\mathbf{B}(t)$ represents its vector of increments.
- $\tilde{\sigma}\left(\mathbf{X}_\varepsilon(t)\right)$ is the $n \times m$ diffusion matrix, which has a block structure:

$$\tilde{\sigma}(\mathbf{X}_\varepsilon(t)) = \begin{pmatrix} \mathbf{0}_{(s+1)\times m} \\ \Sigma(\mathbf{X}_\varepsilon(t)) \end{pmatrix}$$

  Here, $\mathbf{0}_{(s+1)\times m}$ is a zero matrix of size $(s+1) \times m$. The lower block, $\Sigma\left(\mathbf{X}_\varepsilon(t)\right)$, is an $(n-s-1) \times m$ matrix, its entries are $\sigma_{i,k}, i = s+1, s+2, \dots, n-1, k = 1, 2, \dots m$. To ensure the noise is active in the subspace it acts upon, the matrix $\Sigma\left(\mathbf{X}_\varepsilon\right)^{\text{tr}}\Sigma\left(\mathbf{X}_\varepsilon\right)$ is assumed to be non-singular (and then positive-definite) for all relevant states $\mathbf{X}_\varepsilon$.

In this case, the rate function for the large deviation of X_ε can be written as

$$I_x^T(\psi) = \begin{cases} \frac{1}{2}\int_0^T h(t)^{\text{tr}}\Sigma\left(\Sigma^{\text{tr}}\Sigma\right)^{-2}\Sigma^{\text{tr}}h(t)dt, & \text{if } \psi = \left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right) \in \mathbf{AC}_T^x, \\ +\infty, & \text{otherwise,} \end{cases} \tag{3.3}$$

where ψ is a path consisting of a function φ and its first to $(n-1)$-th derivatives, i.e., $\varphi^{(i)}$ denotes the i-th derivative of φ, and

$$h(t)^{\text{tr}} = \left(\varphi^{(s+2)}(t) - b_{s+1}\left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(t), \dots, \varphi^{(n)}(t) - b_{n-1}\left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(t)\right)$$

is a column vector of dimension $n-1-s$, then the corresponding quasipotential can be expressed as

$$\begin{aligned} \mathbb{V}(x,y) = \inf\Big\{ I_x^T(\psi) : &\psi(0) = \left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(0) = x, \\ &\psi(T) = \left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(T) = y,\ T > 0 \Big\}, \quad x, y \in \mathbb{R}^n. \end{aligned} \tag{3.4}$$

We first show that the quasipotential of system (3.1) (3.2) is well defined.

Lemma 3.1. *Suppose that the rate function I_x^T is of the form of (3.3). Then for any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$ and $T > 0$, there exists $\psi^x \in C_T$ with $\psi^x(0) = x$ and $\psi^x(T) = y$ such that $I_x^T\left(\psi^x\right) < \infty$. It follows that for any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$, $\mathbb{V}(x,y) < \infty$.*

Proof. Fix $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ and $T > 0$. Let

$$f_1(t) = \sum_{k=1}^n \frac{1}{(k-1)!}x_k t^{k-1} \quad \text{and} \quad f_2(t) = \sum_{k=1}^n \frac{1}{(k-1)!}y_k(t-T)^{k-1}. \tag{3.5}$$

and we choose a cut-off function $\alpha \in C^\infty(\mathbb{R})$:

$$\alpha(t) = \begin{cases} 1 & t \in [0, \frac{T}{3}] \\ > 0 & t \in (\frac{T}{3}, \frac{2T}{3}) \\ 0 & t \in [\frac{2T}{3}, T]. \end{cases} \tag{3.6}$$

Then $\varphi = \alpha f_1 + (1-\alpha)f_2$ satisfies $\left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(0) = x$, $\left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(T) = y$. By the assumptions that the coefficients $b_i, i = 1, 2, \dots, n-1$ and $\sigma_{i,k}, i = s+1, s+2, \dots, n-1, k = 1, 2, \dots m$ are real-valued, locally Lipschitz continuous functions on $\mathbb{R}^n$ and $\Sigma^{\text{tr}}\Sigma$ is always invertible for all $x \in \mathbb{R}^n$, it is easy to see that $I_x^T\left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right) < \infty$. $\square$

Next, we present the equivalent conditions for the continuity of quasipotential, which will serve as the foundation for our subsequent analysis.

Lemma 3.2. *The following two statements are equivalent:*

(1) The quasipotential $\mathbb{V}$ is upper-semicontinuous on $\mathbb{R}^n \times \mathbb{R}^n$, i.e., for any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$\varlimsup_{(\tilde{x},\tilde{y})\to(x,y)} \mathbb{V}(\tilde{x},\tilde{y}) \leq \mathbb{V}(x,y).$$

(2) *For any $x, y \in \mathbb{R}^n$ with $\|x-y\|_\infty < \delta$ (where $\delta > 0$ is sufficiently small), there exists a path $\psi(t)$ and a time $T = T(\delta) > 0$ such that the path satisfies the boundary conditions:*

$$\begin{cases} \psi(0) = \left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(0) = x, \\ \psi(T) = \left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)(T) = y, \end{cases}$$

and the corresponding rate function is bounded:

$$I_x^T(\psi) = \frac{1}{2}\int_0^T h(t)^{\mathrm{tr}}\Sigma(\psi(t))\left(\Sigma(\psi(t))\Sigma(\psi(t))^{\mathrm{tr}}\right)^{-2}\Sigma(\psi(t))^{\mathrm{tr}}h(t)\,dt < D_\delta, \quad \textit{with } \lim_{\delta\to 0} D_\delta = 0.$$

Proof. We first assume that the quasipotential $\mathbb{V}(x,y)$ is a continuous function on $\mathbb{R}^n \times \mathbb{R}^n$. According to upper-semicontinuous:

$$\lim_{y\to x}\mathbb{V}(x,y) \le \mathbb{V}(x,x) = 0.$$

According to the definition of quasipotential, this formula corresponds to (2).

On the other hand, we assume that (2) holds. Since $\mathbb{V}(x,y) < \infty$, for any $\epsilon > 0$, there exist $T_0 > 0$ and $\psi_0 = \left(\varphi_0, \varphi_0^{(1)}, \dots, \varphi_0^{(n-1)}\right) \in \mathbf{AC}_{T_0}^x$ satisfying

(1) $\psi_0(0) = x, \psi_0(T_0) = y$;
(2) The rate function on this path satisfying

$$\frac{1}{2}\int_0^{T_0} h_0(t)^T\left(\Sigma\Sigma^T\right)^{-1}h_0(t)dt \le \mathbb{V}(x,y) + \epsilon. \tag{3.7}$$

Then from (2), if $\|\tilde{x}-x\|$ are small enough, then there exists $T_1 > 0$ and $\psi_x = \left(\varphi_x, \varphi_x^{(1)}, \dots, \varphi_x^{(n-1)}\right)$ such that $\psi_x(0) = \tilde{x}, \psi_x(T_1) = x$ and $I_x^{T_1}(\psi_x) < \epsilon$. Similarly, if $\|\tilde{y}-y\|$ are small enough, we can find $T_2 > 0$ and $\psi_y = \left(\varphi_y, \varphi_y^{(1)}, \dots, \varphi_y^{(n-1)}\right)$ such that $\psi_y(0) = y, \psi_y(T_2) = \tilde{y}$ and $I_x^{T_2}(\psi_y) < \epsilon$. We consider

$$\varphi^{(n)}(t) = \begin{cases} \varphi_x^{(n)}(t), & t \in [0, T_1), \\ \varphi_0^{(n)}(t-T_1), & t \in [T_1, T_0+T_1), \\ \varphi_y^{(n)}(t-T_0-T_1), & t \in [T_0+T_1, T_0+T_1+T_2]. \end{cases} \tag{3.8}$$

and

$$\begin{aligned} \varphi^{(n-1)}(t) &= \tilde{x}_n + \int_0^t \varphi^{(n)}(s)ds, \\ \varphi^{(n-2)}(t) &= \tilde{x}_{n-1} + \int_0^t \varphi^{(n-1)}(s)ds, \\ &\dots \\ \varphi(t) &= \tilde{x}_1 + \int_0^t \varphi^{(1)}(s)ds. \end{aligned} \tag{3.9}$$

Let $\psi = \left(\varphi, \varphi^{(1)}, \dots, \varphi^{(n-1)}\right)$, it's easy to know

$$\psi(0) = \tilde{x} \quad \psi(T_1) = x \quad \psi(T_1+T_0) = y \quad \psi(T_0+T_1+T_2) = \tilde{y}, \tag{3.10}$$

and

$$\mathbb{V}(\tilde{x},\tilde{y}) \le I_x^{T_0+T_1+T_2}(\psi) < \mathbb{V}(x,y) + 3\epsilon. \tag{3.11}$$

Therefore

$$\lim_{(\tilde{x},\tilde{y})\to(x,y)}\mathbb{V}(\tilde{x},\tilde{y}) \le \mathbb{V}(x,y). \tag{3.12}$$

and we ultimately establish the lemma. □

To analyze the continuity of $\mathbb{V}(x,y)$ on $\mathbb{R}^n \times \mathbb{R}^n$, the core lies in verifying that $\mathbb{V}(x,y) \to 0$ as $y \to x$. Let us define a family of functions that will be instrumental in the following discussion.

Let $n \in \mathbb{N}^+, T > 0$, and $m \le n$. We define the family of functions as:

$$f_T(s;m) = c \cdot P_m^{**}(s;T) = c \cdot P_m\left(\frac{2s}{T}-1\right), \quad s \in [0,T], \tag{3.13}$$

where $c \in \mathbb{R}\backslash\{0\}$ is an arbitrary non-zero constant, $P_m(x)$ denotes the standard Legendre polynomial of degree m and $P_m^{**}(s;T)$ is the corresponding shifted form. Therefore, $P_m^{**}(s;T)$ is orthogonal to all polynomials of degree less than m on the interval $[0,T]$, i.e.

$$\int_0^T P_m^{**}(s;T)q(s)ds = 0.$$

We now recall the following classical result concerning the roots of Legendre polynomials.

Lemma 3.3. *Let $P_m(x)$ be the Legendre polynomial of degree m (with $m \geq 1$). Define a sequence of functions $g_k(x)$ as follows:*

$$g_0(x) := P_m(x), \quad g_k(x) := \int_{-1}^{x} g_{k-1}(t)\,dt, \text{for } k \geq 1.$$

Then for each integer k satisfying $1 \leq k \leq m$:

1. *The function $g_k(x)$ vanishes at the endpoints $x = -1$ and $x = 1$, i.e., $g_k(-1) = 0$ and $g_k(1) = 0$.*
2. *The function $g_k(x)$ has at most $m - k$ distinct real roots in the open interval $(-1, 1)$.*

Proof. We will prove this by mathematical induction on k. We first prove the statement for $k = 1$. Consider the function $g_1(x) = \int_{-1}^{x} P_m(t)\,dt$. At $x = -1$, by the definition of the integral, $g_1(-1) = \int_{-1}^{-1} P_m(t)\,dt = 0$. At $x = 1$, $g_1(1) = \int_{-1}^{1} P_m(t)\,dt$. Since $m \geq 1$, we can write the constant function 1 as the zeroth-order Legendre polynomial $P_0(t)$. By the orthogonality of Legendre polynomials:

$$g_1(1) = \int_{-1}^{1} P_m(t)P_0(t)\,dt = 0 \quad (\text{because } m \neq 0)$$

Therefore, $g_1(x)$ has roots at both $x = -1$ and $x = 1$.

Let N_k denote the number of distinct real roots of the function $g_k(x)$ in the open interval $(-1, 1)$. We need to prove that $N_1 \leq m - 1$.

The function $g_0(x) = P_m(x)$ is the Legendre polynomial of degree m. A well-known property is that when $m \geq 1$, $P_m(x)$ has exactly m distinct real roots in the open interval $(-1, 1)$. Thus, $N_0 = m$. Suppose $g_1(x)$ has N_1 distinct real roots in $(-1, 1)$. Then, the function $g_1(x)$ has at least $N_1 + 2$ distinct real roots on the closed interval $[-1, 1]$ (including the endpoints -1 and 1). By Rolle's Theorem, between any two roots of $g_1(x)$, its derivative $g_1'(x)$ must have at least one root. Therefore, $g_1'(x)$ has at least $(N_1 + 2) - 1 = N_1 + 1$ distinct real roots in the open interval $(-1, 1)$. Since $g_1'(x) = g_0(x) = P_m(x)$, it follows that $P_m(x)$ has at least $N_1 + 1$ roots in $(-1, 1)$. That is, $N_0 \geq N_1 + 1$. Substituting $N_0 = m$, we obtain $m \geq N_1 + 1$, which implies $N_1 \leq m - 1$.

Therefore, the conclusion holds for $k = 1$.

Assume the theorem holds for some integer j where $1 \leq j < m$. We now prove the theorem for $k = j + 1$. Consider $g_{j+1}(x) = \int_{-1}^{x} g_j(t)\,dt$. At $x = -1$, $g_{j+1}(-1) = \int_{-1}^{-1} g_j(t)\,dt = 0$. At $x = 1$, $g_{j+1}(1) = \int_{-1}^{1} g_j(t)\,dt$. To show $g_{j+1}(1) = 0$, we first state Cauchy formula for repeated integration here, i.e.:

$$g_k(t) = \frac{1}{(k-1)!}\int_{-1}^{t} (t-\tau)^{k-1} P_m(\tau)\,d\tau. \tag{3.14}$$

This formula can also be proven by induction. It is easy to verify that (3.14) holds for $k = 0$ and $k = 1$. Now assume that (3.14) holds for a given integer $k \geq 0$, we then prove it for $k + 1$:

$$g_{k+1}(1) = \int_{-1}^{1} \left(\frac{1}{(k-1)!}\int_{-1}^{t} (t-\tau)^{k-1} P_m(\tau)\,d\tau \right) dt$$

Interchanging the order of integration (by Fubini's theorem),

$$g_{k+1}(1) = \frac{1}{(k-1)!}\int_{-1}^{1} P_m(\tau) \left(\int_{\tau}^{1} (t-\tau)^{k-1}\,dt \right) d\tau$$

The inner integral evaluates to:

$$\int_{\tau}^{1} (t-\tau)^{k-1}\,dt = \left[\frac{(t-\tau)^k}{k} \right]_{\tau}^{1} = \frac{(1-\tau)^k}{k}$$

Substituting back,

$$g_{k+1}(1) = \frac{1}{k!}\int_{-1}^{1} (1-\tau)^k P_m(\tau)\,d\tau$$

Therefore (3.14) holds for all k. Consider g_{j+1}, since $j < m$, the polynomial $(1-\tau)^j$ has degree $j < m$. By the orthogonality of $P_m(\tau)$ against all polynomials of degree less than m, this integral is zero. Therefore, $g_{j+1}(1) = 0$. So $g_{j+1}(x)$ vanishes at both endpoints.

By induction hypothesis $N_j \leq m - j$. We need to prove that the conclusion also holds for $j+1$, i.e., $N_{j+1} \leq m-(j+1)$. From the first part, we know that $g_{j+1}(-1) = 0$ and $g_{j+1}(1) = 0$. Suppose $g_{j+1}(x)$ has N_{j+1} distinct real roots in $(-1, 1)$. Then, the function $g_{j+1}(x)$ has at least $N_{j+1} + 2$ distinct real roots on the closed interval $[-1, 1]$ (including the endpoints). Applying Rolle's Theorem again, its derivative $g'_{j+1}(x)$ must have at least $(N_{j+1} + 2) - 1 = N_{j+1} + 1$ distinct real roots in the open interval $(-1, 1)$. Since $g'_{j+1}(x) = g_j(x)$, it follows that $g_j(x)$ has at least $N_{j+1} + 1$ distinct real roots in $(-1, 1)$. That is, $N_j \geq N_{j+1} + 1$. Then, we obtain:

$$N_{j+1} \leq m - j - 1$$

The inductive step is complete. □

For convenience, we define the following notation:

Definition 3.4. *Let $n \in \mathbb{N}$ be a positive integer and let $f : [0, \infty) \to \mathbb{R}$ be a function such that all the integrals in the following expression are well-defined. We define the n-th iterated integral operator, denoted by I_n as follows:*

$$I_n(f)(t) = \int_0^t \int_0^{s_1} \int_0^{s_2} \cdots \int_0^{s_{n-1}} f(s_n)\, \mathrm{d}s_n\, \mathrm{d}s_{n-1} \cdots\, \mathrm{d}s_1$$

The expression defines $I_n(f)$ as a function of the upper integration limit t.

By Cauchy formula for repeated integration (3.14), if f is continuous, then its n-th iterated integral can be written as a single integral:

$$I_n(f)(t) = \frac{1}{(n-1)!} \int_0^t (t-s)^{n-1} f(s) dt$$

First, let's prove several useful lemmas.

Lemma 3.5. *The family of functions $f_T(s; m)$ (3.13) satisfies the following three conditions:*

(1) For all $1 \leq i \leq m$:

$$I_i\left(f_T(s; m)\right)(T) = \frac{1}{(i-1)!} \int_0^T (T-s)^{i-1} f_T(s) ds = 0.$$

(2) For $i = m+1$:

$$I_{m+1}\left(f_T(s; m)\right)(T) = \frac{c}{m!} \int_0^T (T-s)^m P_m^{**}(s; T) ds = c \cdot (-1)^m \frac{m!}{(2m)!(2m+1)} T^{m+1} \neq 0.$$

(3) For all $m+1 < i \leq n$, there exist constants $C > 0$ and $\alpha_i = i$, such that:

$$\left| I_i\left(f_T(s; m)\right)(T) \right| \leq C \cdot T^{\alpha_i}.$$

Proof. First, consider $i \leq m$, $(T-s)^{i-1}$ is a polynomial in s of degree $i-1$, and since $i - 1 < m$, it lies in the space of polynomials of degree less than m. Because $P_m^{**}(s; T)$ is orthogonal on $[0, T]$ to all such polynomials, we have:

$$\int_0^T (T-s)^{i-1} P_m^{**}(s; T) ds = 0.$$

Therefore,

$$I_i\left(f_T(s; m)\right)(T) = \frac{c}{(i-1)!} \int_0^T (T-s)^{i-1} P_m^{**}(s; T) ds = 0.$$

Then, consider $i = m+1$, with

$$I_{m+1}\left(f_T(s; m)\right)(T) = \frac{c}{m!} \int_0^T (T-s)^m P_m^{**}(s; T) ds.$$

We expand $(T-s)^m$ in the orthogonal basis $\left\{P_0^{**}, P_1^{**}, \ldots, P_m^{**}\right\}$, i.e. $(T-s)^m = \sum_{j=0}^m a_j P_j^{**}(s; T)$, where the coefficients a_j are determined by the inner product:

$$a_j = \frac{\left\langle (T-\cdot)^m, P_j^{**}(\cdot; T) \right\rangle}{\left\langle P_j^{**}(\cdot; T), P_j^{**}(\cdot; T) \right\rangle}.$$

Substituting this expansion into the integral and by orthogonality we get:

$$\int_0^T (T-s)^m P_m^{**}(s; T) ds = \int_0^T \left(\sum_{j=0}^m a_j P_j^{**}(s; T) \right) P_m^{**}(s; T) ds = a_m \int_0^T \left(P_m^{**}(s; T)\right)^2 ds.$$

Notice that $\int_0^T (P_m^{**}(s;T))^2\,ds = \frac{T}{2m+1}$, then we need to determine a_m. The leading coefficient of the standard Legendre polynomial $P_m(x)$ is known to be $\frac{(2m)!}{2^m(m!)^2}$. Thus, the leading term of $P_m^{**}(s;T)$ is $\frac{(2m)!}{2^m(m!)^2}\left(\frac{2s}{T}\right)^m = \frac{(2m)!}{(m!)^2T^m}s^m$. Compare the coefficients of s^m on both sides of the equation $(T-s)^m = \sum_{j=0}^m a_jP_j^{**}(s;T)$ and solve a_m, we obtain:

$$a_m = (-1)^m \cdot \frac{(m!)^2T^m}{(2m)!}$$

Substituting the above results into the expression for $I_{m+1}\left(f_T(s;m)\right)(T)$:

$$I_{m+1}\left(f_T(s;m)\right)(T) = \frac{c}{m!}\cdot a_m \cdot \int_0^T (P_m^{**}(s;T))^2\,ds = \frac{c}{m!}\cdot\left((-1)^m\cdot\frac{(m!)^2T^m}{(2m)!}\right)\cdot\frac{T}{2m+1}.$$

Simplifying:

$$I_{m+1}\left(f_T(s;m)\right)(T) = c\cdot(-1)^m\cdot\frac{m!}{(2m)!(2m+1)}\cdot T^{m+1}.$$

For $m+1 < i \le n$, let $s = Tu, t = Ty$, so $u \in [0,y], y \in [0,1]$. The integral becomes:

$$I_i\left(f_T(s;m)\right) = \frac{cT^i}{(i-1)!}\int_0^1 (1-u)^{i-1}P_m(2u-1)du.$$

The integrand $(1-u)^{i-1}P_m(2u-1)$ is a polynomial in u, and its absolute integral over $[0,1]$ is bounded by some constant $K_{i,m}$. Thus:

$$\left|I_i\left(f_T(s;m)\right)\right| \le \left|\frac{cK_{i,m}}{(i-1)!}\right|T^i$$

Set $C = \left|\frac{cK_{i,m}}{(i-1)!}\right|$, $\alpha_i = i$. Then Condition (3) holds. □

Lemma 3.6. *We have the following maximum value estimates:*

(1) For $1 \le i \le m$, the function $t \mapsto \left|I_i\left(f_T\right)(t)\right|$ attains its maximum at some interior point $t \in (0,T)$, and

$$\max_{t\in[0,T]}\left|I_i\left(f_T\right)(t)\right| = O\left(T^i\right) \qquad \text{as } T \to 0^+$$

where the implicit constant depends only on i and m.

(2) For $i = m+1$, the maximum is attained at $t = T$, and

$$|I_{m+1}(f_T)(T)| = |c|\cdot\frac{m!}{(2m)!(2m+1)}T^{m+1}.$$

(3) For $i > m+1$, the maximum is attained at $t = T$, and

$$|I_i(f_T)(T)| \le \left|\frac{c}{(i-1)!}\right|\cdot|K_{i,m}|\cdot T^i,$$

where $K_{i,m} := \int_0^1 (1-u)^{i-1}P_m(2u-1)\mathrm{d}u$ is a finite constant depending solely on i and m.

Proof. For $1 \le i \le m$, from Condition (1) in Lemma 3.5, we have $I_i(f_T)(T) = 0$. Since $f_T(s)$ is not identically zero, its integral $I_i(f_T)(t)$ is also not identically zero. By Rolle's theorem, a continuously differentiable function that takes equal values at the endpoints of a closed interval and is not constant must have a critical point in the open interval. This implies that the maximum of $|I_i(f_T)(t)|$ must be attained at some extremal point inside the interval $(0,T)$.

By the change of variables $t = \tau T$, $s = vT$ (with $\tau, v \in [0,1]$), we obtain:

$$I_i(f_T)(t) = T^i\cdot\left[\frac{c}{(i-1)!}\int_0^\tau (\tau - v)^{i-1}P_m(2v-1)dv\right].$$

The expression in the brackets is a bounded function of τ, and therefore the maximum value of $|I_i(f_T)(t)|$ is of order $O(T^i)$.

When $i = m+1$, we need to show that this function is monotonic. The monotonicity is determined by the sign of its derivative:

$$\frac{d}{dt}I_{m+1}(f_T)(t) = I_m(f_T)(t).$$

According to the properties of integrals of Legendre polynomials (Lemma 3.3), the function $I_k(f_T)(t)$ has at most $m-k$ roots in the open interval $(0,T)$. Therefore, when $k = m$, $I_m(f_T)(t)$ has 0 roots in $(0,T)$. Moreover, since $I_m(f_T)(0) = I_m(f_T)(T) = 0$, and $I_m(f_T)(t)$ has no roots in $(0,T)$, it follows that $I_m(f_T)(t)$ does not change sign in $(0,T)$—that is, it is either non-negative or non-positive throughout the interval. Since the derivative $I_m(f_T)(t)$ is of constant sign, the function

$I_{m+1}(f_T)(t)$ is monotonic on $[0, T]$. A monotonic function starting from $I_{m+1}(f_T)(0) = 0$ must achieve the maximum of its absolute value at the other endpoint, $t = T$. This maximum value is:

$$\left|I_{m+1}(f_T)(T)\right| = \left|c \cdot (-1)^m \frac{m!}{(2m)!(2m+1)} T^{m+1}\right| = |c| \cdot \frac{m!}{(2m)!(2m+1)} T^{m+1}.$$

When $i > m+1$, we use a similar argument as in the previous step. The derivative of the function is

$$\frac{d}{dt} I_i(f_T)(t) = I_{i-1}(f_T)(t), \quad \text{where } i-1 \geq m+1.$$

We have already shown that $I_{m+1}(f_T)(t)$ is monotonic and satisfies $I_{m+1}(f_T)(0) = 0$, so it is of constant sign on $(0, T]$. Therefore, the derivative of its integral, $I_{m+2}(f_T)(t)$, namely $I_{m+1}(f_T)(t)$, has constant sign, which implies that $I_{m+2}(f_T)(t)$ is also monotonic. This reasoning can be extended inductively to all $i \geq m+1$. Hence, for all $i > m+1$, the function $I_i(f_T)(t)$ is monotonic on $[0, T]$. Consequently, the maximum of its absolute value is attained at the endpoint $t = T$, and equals $\left|I_i(f_T)(T)\right|$. From Lemma 3.5, we obtain:

$$|I_i(f_T)(T)| \leq \left|\frac{cK_{i,m}}{(i-1)!}\right| T^i.$$

□

We seek to construct a smooth function $\varphi : [0, T_{\text{final}}] \to \mathbb{R}$ and a terminal time $T_{\text{final}} > 0$ (both to be determined) such that the state vector

$$\psi(t) = \left(\varphi(t), \varphi^{(1)}(t), \ldots, \varphi^{(n-1)}(t)\right) \in \mathbb{R}^n$$

satisfies the boundary conditions $\psi(0) = x$ and $\psi(T_{\text{final}}) = y$, where $x, y \in \mathbb{R}^n$ are given and $\|x - y\|$ is sufficiently small.

To achieve this, introduce the control input $u(t) = \varphi^{(n)}(t)$. The state dynamics are governed by the system

$$\begin{cases} \varphi_1' = \varphi_2, \\ \varphi_2' = \varphi_3, \\ \vdots \\ \varphi_{n-1}' = \varphi_n, \\ \varphi_n' = u(t), \end{cases}$$

where $\varphi_i = \varphi^{(i-1)}$ for $i = 1, \ldots, n$. Equivalently, the solution can be expressed via successive integration:

$$\begin{cases} \varphi_n(t) = x_n + \displaystyle\int_0^t u(s)ds, \\ \varphi_{n-1}(t) = x_{n-1} + \displaystyle\int_0^t \varphi_n(s)ds, \\ \vdots \\ \varphi_1(t) = x_1 + \displaystyle\int_0^t \varphi_2(s)ds. \end{cases}$$

For each $i = 1, \ldots, n$, this yields the explicit representation

$$\varphi_i(t) = \sum_{k=0}^{n-i} \frac{x_{i+k}}{k!} t^k + \underbrace{\int_0^t \int_0^{s_1} \cdots \int_0^{s_{n-i}} u(\tau) d\tau ds_{n-i} \cdots ds_1}_{(n-i+1)\text{-fold integral of } u} = \sum_{k=0}^{n-i} \frac{x_{i+k}}{k!} t^k + I_{n-i+1}(u(x))(t).$$

The construction proceeds inductively through n stages, steering the state from $\psi(0) = x$ to $\psi(T_{\text{final}}) = y$ by sequentially adjusting derivatives from highest to lowest order.

Stage 1: Steer φ_n from x_n to y_n. Set the control $u_1(t) = \text{sgn}\left(y_n - x_n\right)$ and duration $T_1 = \left|y_n - x_n\right|$. At time T_1,

$$\varphi_n(T_1) = x_n + \int_0^{T_1} u_1(s)ds = y_n$$

Stage k ($2 \leq k \leq n$): Steer φ_{n-k+1} while preserving previously adjusted coordinates. Assume after time $T^{(k-1)} := T_1 + \cdots + T_{k-1}$, we have $\varphi_i\left(T^{(k-1)}\right) = y_i$ for all $i \in \{n-k+2, \ldots, n\}$. Define the residual error:

$$\Delta_k := y_{n-k+1} - \varphi_{n-k+1}\left(T^{(k-1)}\right)$$

On the interval $\left[T^{(k-1)}, T^{(k-1)}+T_k\right]$, apply the polynomial control

$$u_k(t)=c_k \cdot P_{k-1}\left(\frac{2\left(t-T^{(k-1)}\right)}{T_k}-1\right)$$

where P_{k-1} is the Legendre polynomial of degree $k-1$, and $c_k= \pm 1, T_k>0$ are parameters to be chosen.

The change in φ_{n-k+1} over this stage is given by the k-fold integral

$$\Delta \varphi_{n-k+1}=\int_{T^{(k-1)}}^{T^{(k)}} \cdots \int_{T^{(k-1)}}^{s_2} u_k\left(s_1\right) d s_1 \cdots d s_k=c_k \cdot(-1)^{k-1} \frac{(k-1)!}{(2 k-1)!} T_k^k$$

per Lemma 3.5. Setting $\Delta \varphi_{n-k+1}=\Delta_k$ and choosing $c_k=\operatorname{sgn}\left(\Delta_k\right)$ yields

$$T_k=\left(\frac{\left|\Delta_k\right| \cdot(2 k-1)!}{(k-1)!}\right)^{1 / k}$$

For any previously adjusted coordinate $\varphi_i((i>n-k+1))$, the net change is

$$\Delta \varphi_i=\int_{T^{(k-1)}}^{T^{(k)}} \cdots \int_{T^{(k-1)}}^{s_{n-i+1}} u_k\left(s_1\right) d s_1 \cdots d s_{n-i+1}$$

Since $n-i+1 \leq k-1$ (as $i \geq n-k+2$), Lemma 3.5 implies $\Delta \varphi_i=0$. Thus, all coordinates φ_i for $i \geq n-k+2$ remain fixed at their target values y_i.

By sequentially carrying out these n stages, we construct a piecewise function $u(t)$:

$$u(t)= \begin{cases}u_1(t) & t \in\left[0, T_1\right) \\ u_2(t) & t \in\left[T_1, T_1+T_2\right) \\ \vdots & \vdots \\ u_n(t) & t \in\left[T^{(n-1)}, T^{(n-1)}+T_n\right)\end{cases}$$

where $T^{(k)}=\sum_{i=1}^k T_i$. At the final time $T_{\text{final}}=T^{(n)}=T_1+\cdots+T_n$, we have:

$$\left(\varphi, \varphi^{(1)}, \ldots, \varphi^{(n-1)}\right)\left(T_{\text{final}}\right)=\left(y_1, y_2, \ldots, y_n\right).$$

Lemma 3.7. *Suppose $x, y \in \mathbb{R}^n$ satisfy $\|x-y\|_{\infty} \leq \delta$ for some $\delta>0$. Then there exist constants $C>0$ depending only on n such that the total transition time satisfies*

$$T_{\text{final}}=\sum_{k=1}^n T_k \leq C \delta^{1 / n}$$

Consequently, $\lim_{\delta \rightarrow 0} T_{\text{final}}=0$.

Proof. We proceed by induction on k to establish $T_k \leq C_k \delta^{1 / k}$ for constants $C_k>0$. From Stage 1, $T_1=\left|y_n-x_n\right| \leq \delta$. Thus $T_1 \leq C_1 \delta^{1 / 1}$ with $C_1=1$.

Assume for all $j<k$ that $T_j \leq C_j \delta^{1 / j}$. According to the construction scheme, we have $T_k \propto\left|\Delta_{n-k+1}\right|^{1 / k}$. Consider the residual error for coordinate $m=n-k+1$ before Stage k :

$$\Delta_{n-k+1}:=y_{n-k+1}-\varphi_{n-k+1}\left(T^{(k-1)}\right)$$

By the system dynamics and the induction hypothesis, By construction,

$$\Delta_{n-k+1}=\left(y_{n-k+1}-x_{n-k+1}\right)-\sum_{j=1}^{k-1} I_k\left(u_j\right)\left(T_j\right),$$

where $\sum_{j=1}^{k-1} I_k\left(u_j\right)\left(T_j\right)$ represents the total error accumulated during the first $k-1$ steps. The first term, $\left(y_{n-k+1}-x_{n-k+1}\right)$, is clearly $O(\delta)$.

By Lemma 3.5, we know that $\left|I_k\left(u_j\right)\left(T_j\right)\right| \leq C \cdot T_j^k$. Substituting the inductive hypothesis $T_j=O\left(\delta^{1 / j}\right)$ yields

$$\left|I_k\left(u_j\right)\left(T_j\right)\right|=O\left(T_j^k\right)=O\left(\left(\delta^{1 / j}\right)^k\right)=O\left(\delta^{k / j}\right) .$$

Thus, the order of the total perturbation is $\sum_{j=1}^{k-1} O\left(\delta^{k / j}\right)$. Substituting this into the expression for Δ_{n-k+1} gives

$$\Delta_{n-k+1}=O(\delta)-O\left(\delta^{k /(k-1)}\right)=O(\delta),$$

since $\delta^{k/(k-1)}$ is of higher order than δ as $\delta \to 0$. Consequently,

$$T_k = O\left(|\Delta_{n-k+1}|^{1/k}\right) = O\left(\delta^{1/k}\right).$$

The total time is therefore

$$T_{\text{final}} = \sum_{k=1}^{n} T_k = O(\delta) + O\left(\delta^{1/2}\right) + \cdots + O\left(\delta^{1/n}\right) = O\left(\delta^{1/n}\right).$$

Clearly, $\lim_{\delta\to 0} T_{\text{final}} = 0$.

□

Lemma 3.8. *Let $\varphi(t)$ be the path connecting states x and y constructed as previously described, where $\|x-y\|_\infty \le \delta$ and δ is sufficiently small. Define $\psi(t) = \left(\varphi(t), \varphi^{(1)}(t), \ldots, \varphi^{(n-1)}(t)\right)$ as the state vector of the path at time t. Then there exists a constant $C > 0$ independent of δ such that the maximum norm of the path's state vector satisfies the following bound:*

$$\max_{t\in[0,T_{final}]} \|\psi(t)\|_\infty \le \|x\|_\infty + C \cdot \delta^{1/n}.$$

Proof. Our goal is to establish an upper bound for $\max_{t,i}\left|\varphi_i(t)\right|$. At any time $t \in [0, T_{\text{final}}]$, t must belong to a specific construction stage $k \in \{1, \ldots, n\}$. Thus, we can express t as $t = T^{(k-1)} + \tau$, where $T^{(k-1)} = \sum_{j=1}^{k-1} T_j$ denotes the total duration of the first $k-1$ stages, and $\tau \in [0, T_k]$. At this moment, the value of any coordinate $\varphi_i(t)$ decomposes into the sum of its initial value at the start of phase k and the dynamical variation within this phase:

$$\varphi_i(t) = \varphi_i\left(T^{(k-1)}\right) + I_{n-i+1}(u_k)(\tau).$$

Our proof strategy involves bounding these two terms separately.

The initial value $\varphi_i\left(T^{(k-1)}\right)$ at the beginning of stage k is obtained by accumulating perturbations from the first $k-1$ stages starting from the initial value x_i:

$$\varphi_i\left(T^{(k-1)}\right) = x_i + \sum_{j=1}^{k-1} I_{n-i+1}(u_j)(T_j).$$

Based on Lemma 3.5 and our analysis of stage durations $T_j = O\left(\delta^{1/j}\right)$, the magnitude of each perturbation term satisfies

$$\left|I_{n-i+1}(u_j)(T_j)\right| = O\left(T_j^{n-i+1}\right) = O\left(\delta^{(n-i+1)/j}\right).$$

Therefore, the deviation between the initial value of stage k and the original initial value x_i is negligible:

$$\left|\varphi_i\left(T^{(k-1)}\right) - x_i\right| \le \sum_{j=1}^{k-1} O\left(\delta^{(n-i+1)/j}\right) = O\left(\delta^{\alpha}\right),$$

where α is the minimum value among all exponents $(n-i+1)/j$ (which is positive, and we will determine its exact value later). This implies that the initial value of each stage remains well-controlled near the original initial value:

$$\left|\varphi_i\left(T^{(k-1)}\right)\right| \le \|x\|_\infty + O\left(\delta^{\alpha}\right).$$

Then, we use Lemma 3.6 to bound the maximum value of the dynamic variation term $|I_{n-i+1}(u_k)(\tau)|$ for $\tau \in [0, T_k]$. In stage k, the function u_k is based on Legendre polynomials of degree $m = k-1$. Let $p = n-i+1$ denote the integration order.

(1) Case 1: $p \ge k$ (corresponding to $i \le n-k+1$). Here the integration order p is greater than or equal to the polynomial degree plus one ($m+1 = k$). By Lemma 3.6, the maximum of $\left|I_p(u_k)(\tau)\right|$ occurs at the endpoint $\tau = T_k$. This implies monotonic coordinate evolution with no internal explosion exceeding the endpoint value. The endpoint deviation is already accounted for as a small perturbation while estimating the initial value.

(2) Case 2: $p \le k-1$ (corresponding to $i \ge n-k+2$) . Here the integration order p is less than or equal to the polynomial degree $m = k-1$. By Lemma 3.6, the maximum occurs internally within $(0, T_k)$, causing the excess. Its magnitude is:

$$\max_{\tau\in[0,T_k]} \left|I_p(u_k)(\tau)\right| = O\left(T_k^p\right) = O\left(\left(\delta^{1/k}\right)^p\right) = O\left(\delta^{p/k}\right) = O\left(\delta^{(n-i+1)/k}\right)$$

As $\delta \to 0$, the asymptotic order of the total deviation is dominated by the slowest-converging term (i.e., the term with the smallest exponent of δ). For $k \in \{2, \ldots, n\}$ and $i \in \{n-k+2, \ldots, n\}$, the minimal exponent $\frac{n-i+1}{k}$ is found to be $1/n$. Consequently,

$$\left|\varphi_i(t)\right| \le \left|\varphi_i\left(T^{(k-1)}\right)\right| + \max_{\tau\in[0,T_k]} |I_{n-i+1}(u_k)(\tau)| \le \left(\|x\|_\infty + O\left(\delta^{\alpha}\right)\right) + O\left(\delta^{1/n}\right).$$

Similarly, we can find $\alpha = \frac{1}{n-1}$. Then $1/n$ is the smallest exponent among all possible deviation terms. Therefore, there exists a constant $C > 0$ such that:

$$|\varphi_i(t)| \leq \|x\|_\infty + C \cdot \delta^{1/n}$$

for all i and t. Consequently, taking the maximum over both sides yields the desired bound:

$$\max_{t\in[0,T_{\text{final}}]} \|S(t)\|_\infty = \max_{t,i} |\varphi_i(t)| \leq \|x\|_\infty + C \cdot \delta^{1/n},$$

which completes the proof of the Lemma. □

Theorem 3.9. *For any $x, y \in \mathbb{R}^n$ with $\|x-y\|_2 < \delta$ (where $\delta > 0$ is sufficiently small), there exists a path $\psi(t)$ and a time $T = T(\delta) > 0$ such that the path satisfies the boundary conditions:*

$$\begin{cases} \psi(0) = \left(\varphi, \varphi^{(1)}, \ldots, \varphi^{(n-1)}\right)(0) = x, \\ \psi(T) = \left(\varphi, \varphi^{(1)}, \ldots, \varphi^{(n-1)}\right)(T) = y, \end{cases}$$

and the corresponding rate function is bounded:

$$I_x^T(\psi) = \frac{1}{2}\int_0^T h(t)^T \left(\Sigma(\psi(t))\Sigma(\psi(t))^T\right)^{-1} h(t)\,dt < D_\delta, \quad \text{with } \lim_{\delta\to 0} D_\delta = 0.$$

Proof. We select the path $\psi(t)$ constructed using Legendre polynomials, which exactly connects the starting point x and the ending point y. According to Lemmas B.5 and B.6, this path possesses two crucial properties:

(1) The total duration T of the path depends on δ. As $\delta \to 0$, $T = O(\delta^{1/n})$, hence $T \to 0$.
(2) For sufficiently small δ, the path $\psi(t)$ remains uniformly bounded over the entire time interval $[0, T]$. This implies the existence of a compact set $K \subset \mathbb{R}^n$ independent of δ such that $\psi(t) \in K$ for all $t \in [0, T]$.

To analyze this integral, we use adjugate matrix to express the integrand as a quotient:

$$h(t)^T \left(\Sigma\Sigma^T\right)^{-1} h(t) = \frac{h(t)^T \operatorname{adj}\left(\Sigma\Sigma^T\right) h(t)}{\det\left(\Sigma\Sigma^T\right)}.$$

By assumption, the functions $b(x)$ and $\sigma_{i,k}(x)$ are locally Lipschitz continuous (and hence continuous). As continuous functions on a compact set K, they must be uniformly bounded. Thus, there exist positive constants C_b and C_σ such that for all $t \in [0, T]$:

$$\|b(\psi(t))\|_2 \leq C_b \quad \text{and} \quad \|\Sigma(\psi(t))\| \leq C_\sigma \ \text{ for any induced matrix norm}$$

The vector $h(t)$ has components $\varphi^{(k)}(t) - b_{k-1}(\psi(t))$. The construction of our path ensures that all derivatives $\varphi^{(k)}(t)$ are bounded on $[0, T]$. Since both $\varphi^{(k)}(t)$ and $b_{k-1}(\psi(t))$ are bounded, their difference $h(t)$ is uniformly bounded. Therefore, there exists a constant $C_h > 0$ satisfying $\|h(t)\|_2 \leq C_h$ for all t.

The entries of $\Sigma(\psi(t))\Sigma(\psi(t))^T$ are quadratic polynomials in $\sigma_{i,k}(\psi(t))$, hence continuous and bounded along the path. By assumption, the inverse exists, implying $\det\left(\Sigma\Sigma^T\right) \neq 0$ everywhere. As a continuous non-vanishing function on the compact set K, $\left|\det\left(\Sigma(\psi(t))\Sigma(\psi(t))^T\right)\right|$ attains a positive lower bound $C_{\det} > 0$. The adjugate matrix $\operatorname{adj}\left(\Sigma\Sigma^T\right)$ has entries that are bounded combinations of $\Sigma\Sigma^T$ elements, so it is bounded. Consequently, the inverse $\left(\Sigma\Sigma^T\right)^{-1} = \frac{\operatorname{adj}\left(\Sigma\Sigma^T\right)}{\det\left(\Sigma\Sigma^T\right)}$ has bounded norm. Specifically, there exists $C_{\text{inv}} > 0$ such that

$$\left\|\left(\Sigma(\psi(t))\Sigma(\psi(t))^T\right)^{-1}\right\| \leq C_{\text{inv}} \quad \forall t \in [0, T].$$

Consequently, we can bound the integrand as follows:

$$\left|h(t)^T \left(\Sigma\Sigma^T\right)^{-1} h(t)\right| \leq \|h(t)\|^2 \cdot \left\|\left(\Sigma\Sigma^T\right)^{-1}\right\| \leq C_h^2 \cdot C_{\text{inv}}.$$

Let $M = C_h^2 \cdot C_{\text{inv}}$, where $M > 0$ is a constant independent of δ. The entire integral is then bounded by:

$$I_x^T(\psi) = \frac{1}{2}\int_0^T h(t)^T \left(\Sigma\Sigma^T\right)^{-1} h(t)dt \leq \frac{1}{2}\int_0^T M dt = \frac{1}{2}MT.$$

Since $T = O\left(\delta^{1/n}\right)$, there exists a constant $C_T > 0$ such that $T \leq C_T\delta^{1/n}$. Substituting this yields:

$$I_x^T(\psi) \leq \frac{1}{2}M\left(C_T\delta^{1/n}\right).$$

Setting $D_\delta = \frac{1}{2}MC_T\delta^{1/n}$ completes the proof. □