Safely Learning Controlled Stochastic Dynamics

Luc Brogat-Motte[1,2]

Joint work with

Alessandro Rudi[3], Riccardo Bonalli[1]

[1] Laboratoire des Signaux et Systèmes, CNRS, CentraleSupélec, Université Paris-Saclay, France
[2] Istituto Italiano di Tecnologia, Italy
[3] SDA Bocconi, Bocconi University, Italy

NEURAL INFORMATION PROCESSING SYSTEMS

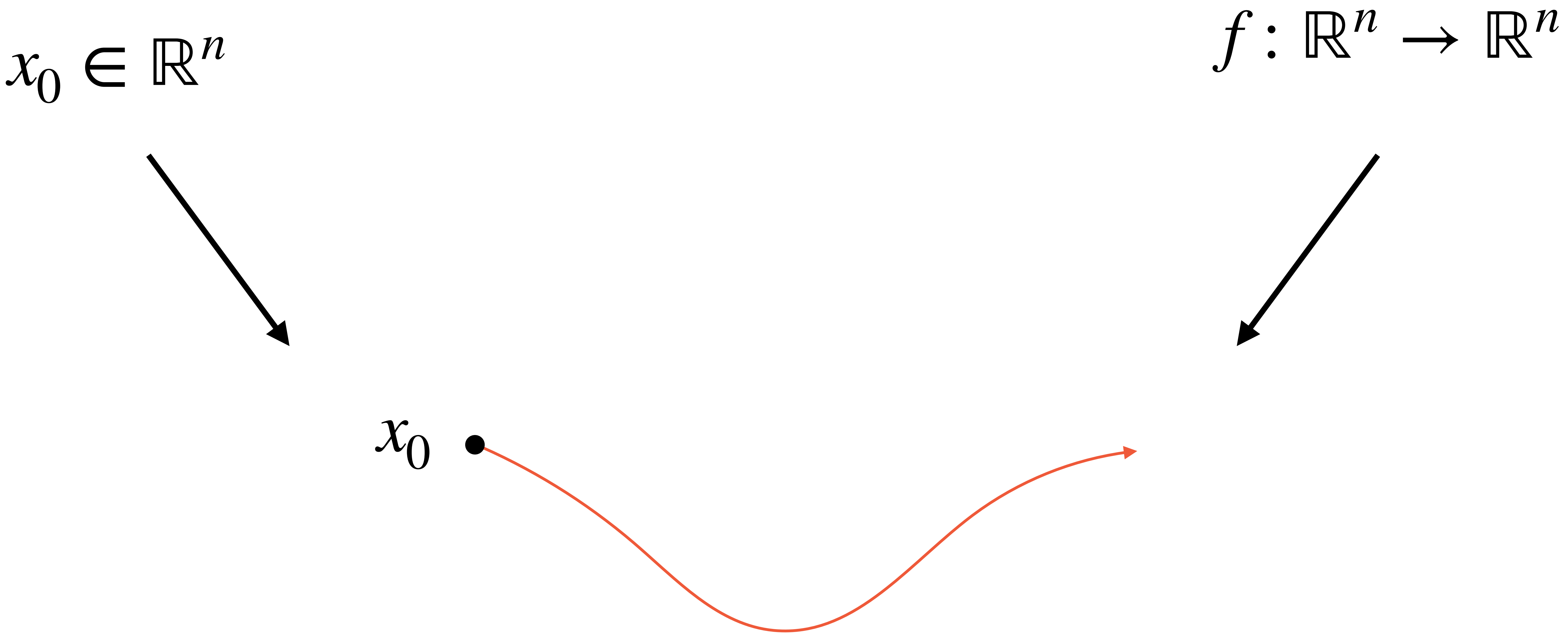
Controlled Dynamical Systems

Continuous-time Dynamics

$x_t \in \mathbb{R}^n$: system's state at time t

$dx_t = f(x_t)\,dt$: system's drift

Continuous-time Dynamics



Controlled Dynamics

$u_t \in \mathbb{R}^d$: control input at time t

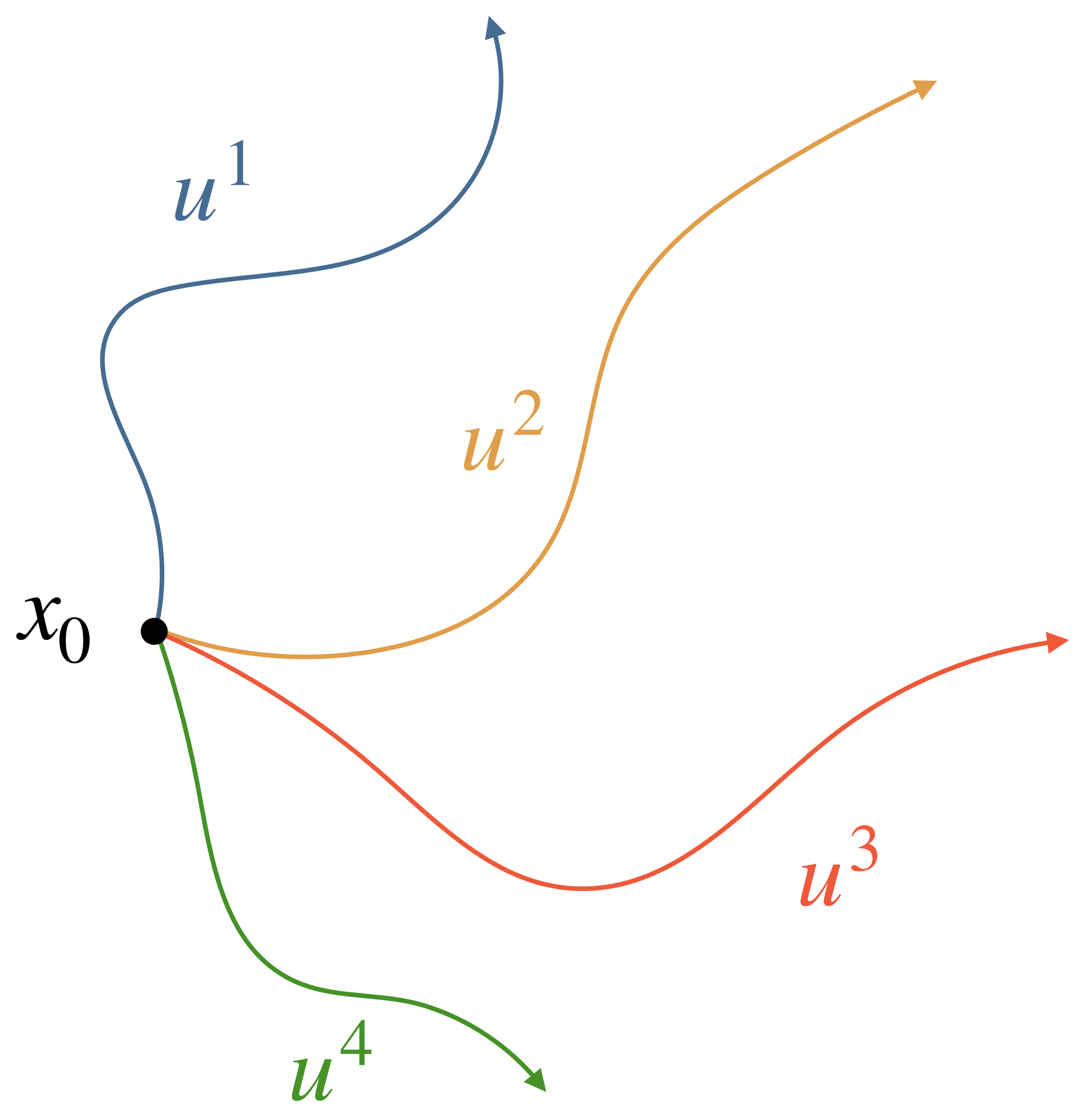
$$dx_t = f(x_t, u_t)\, dt$$

Controlled Dynamics

$$f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$$

$$x_0 \in \mathbb{R}^n$$

$$u : t \mapsto u_t$$

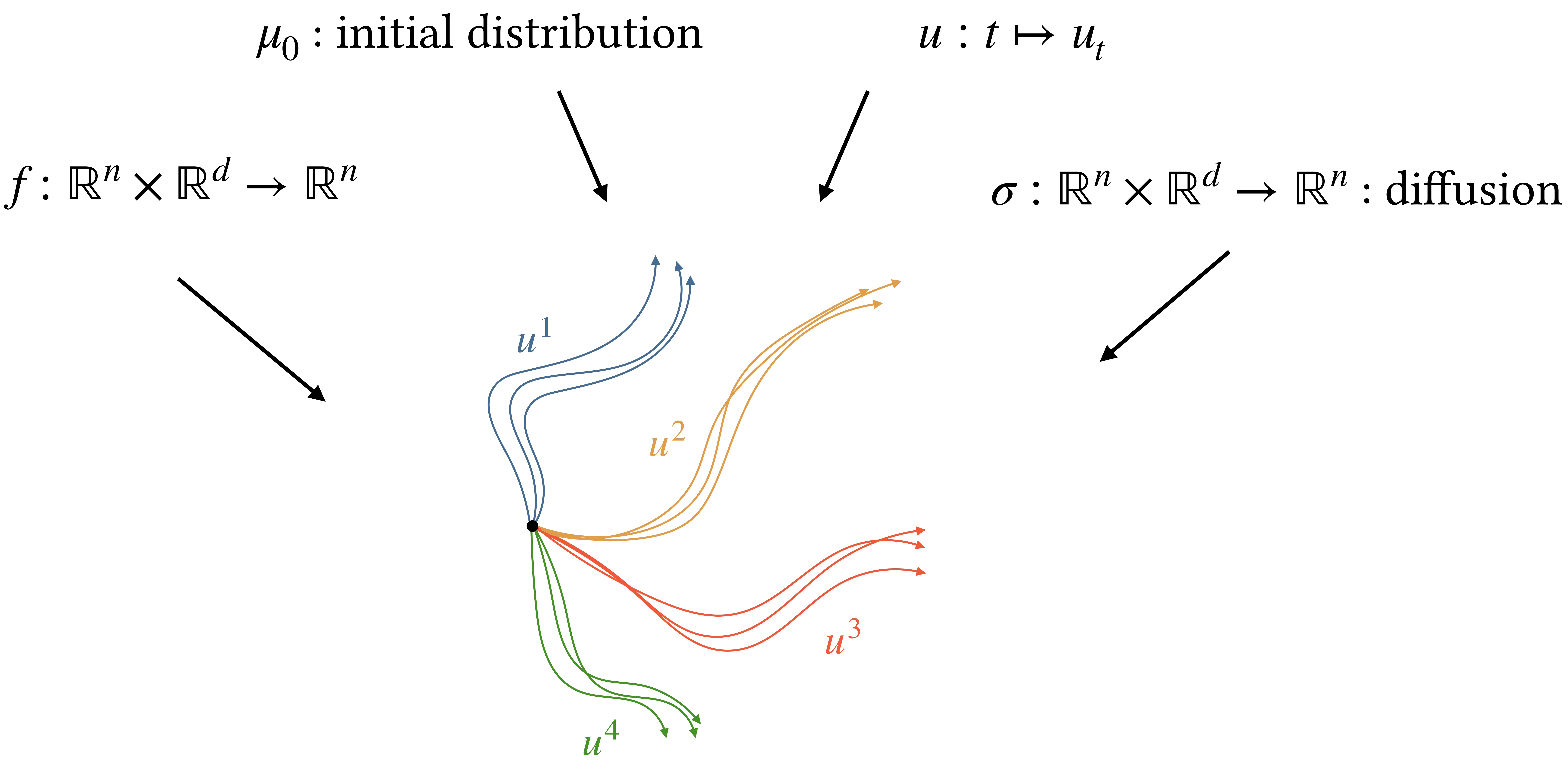


Controlled Stochastic Dynamics

W_t : Brownian motion

$$dX_t = f(X_t, u_t)\, dt + \sigma(X_t, u_t) dW_t$$

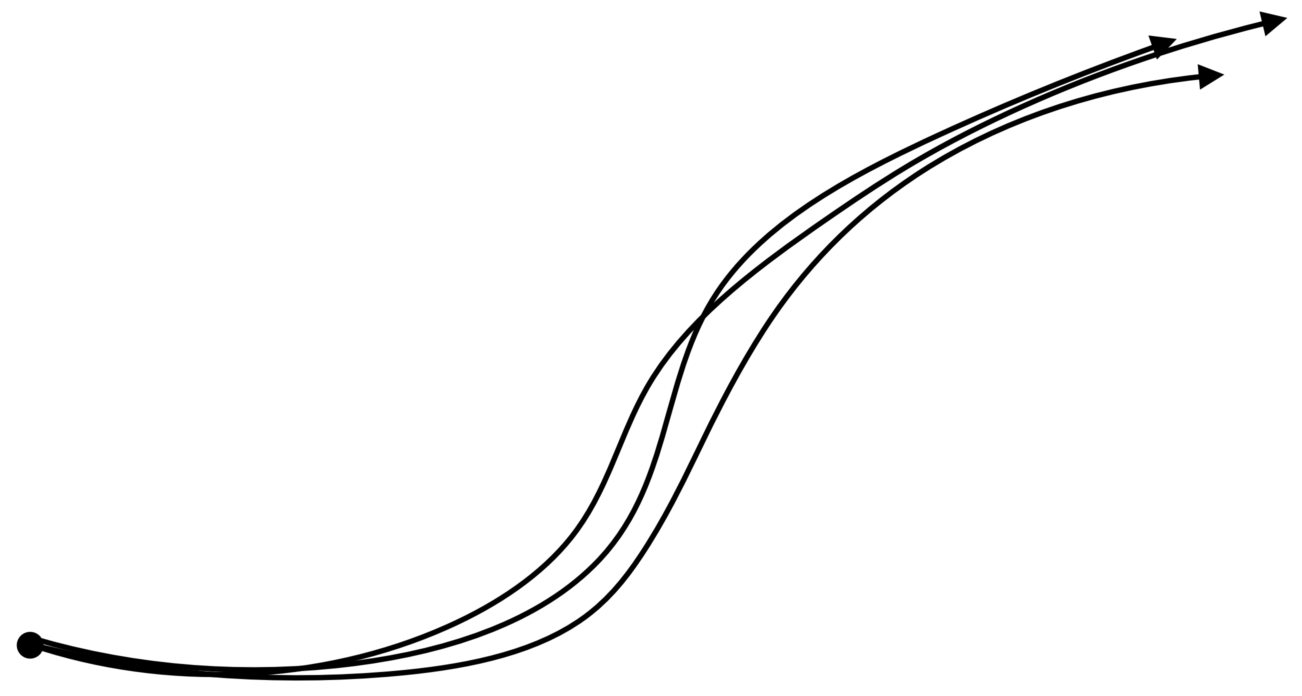
Controlled Stochastic Dynamics



When the model is known

$$(f, \sigma) \longrightarrow u$$

Knowing (f, σ) allows to design controls for desired trajectories



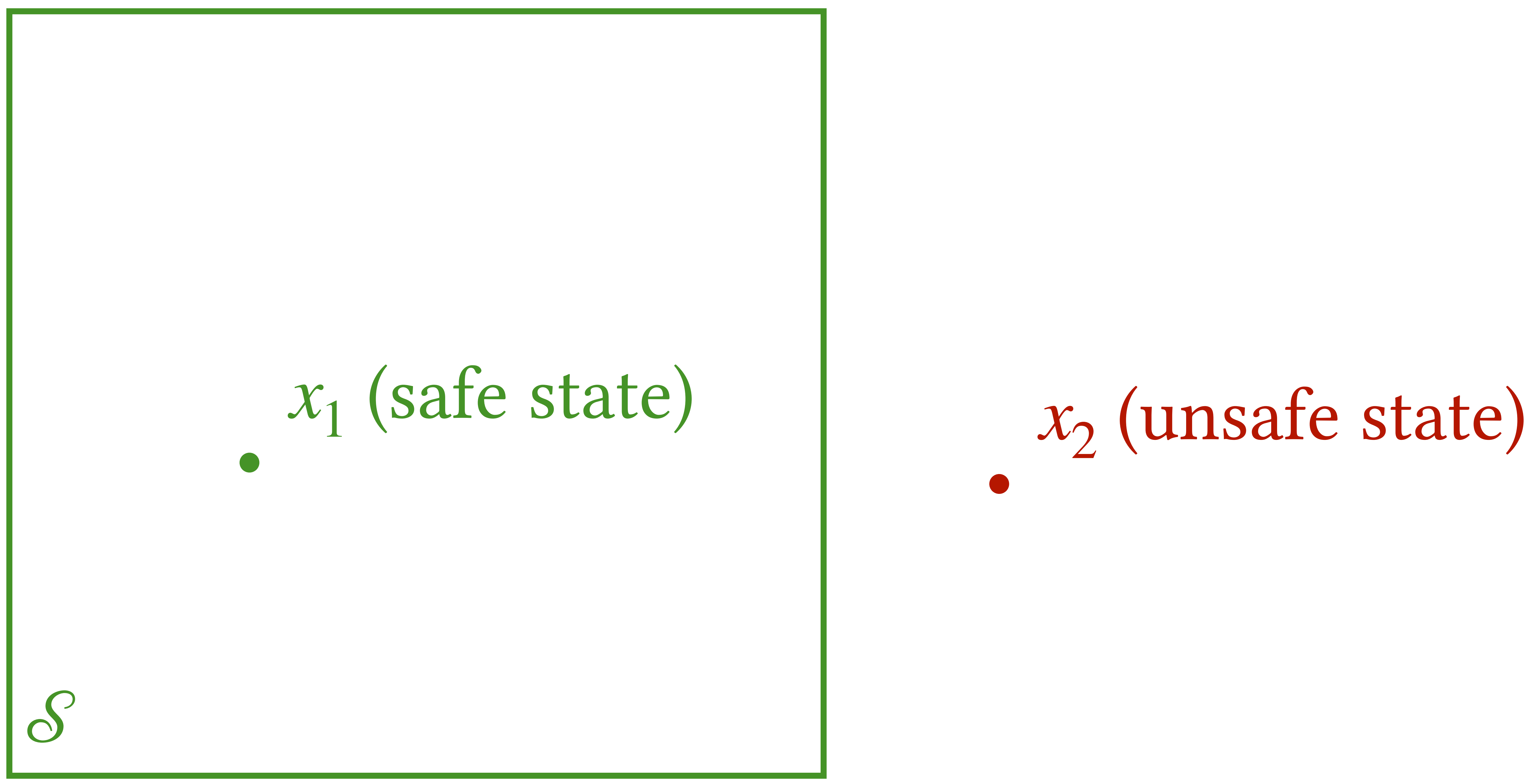
Often the model is unknown

$$u \xrightarrow{?} (f, \sigma)$$

Learning (f, σ) requires collecting trajectories under various controls $u^1, u^2, \ldots$

Safe and unsafe states

In real systems, some states may lead to failure or damage.



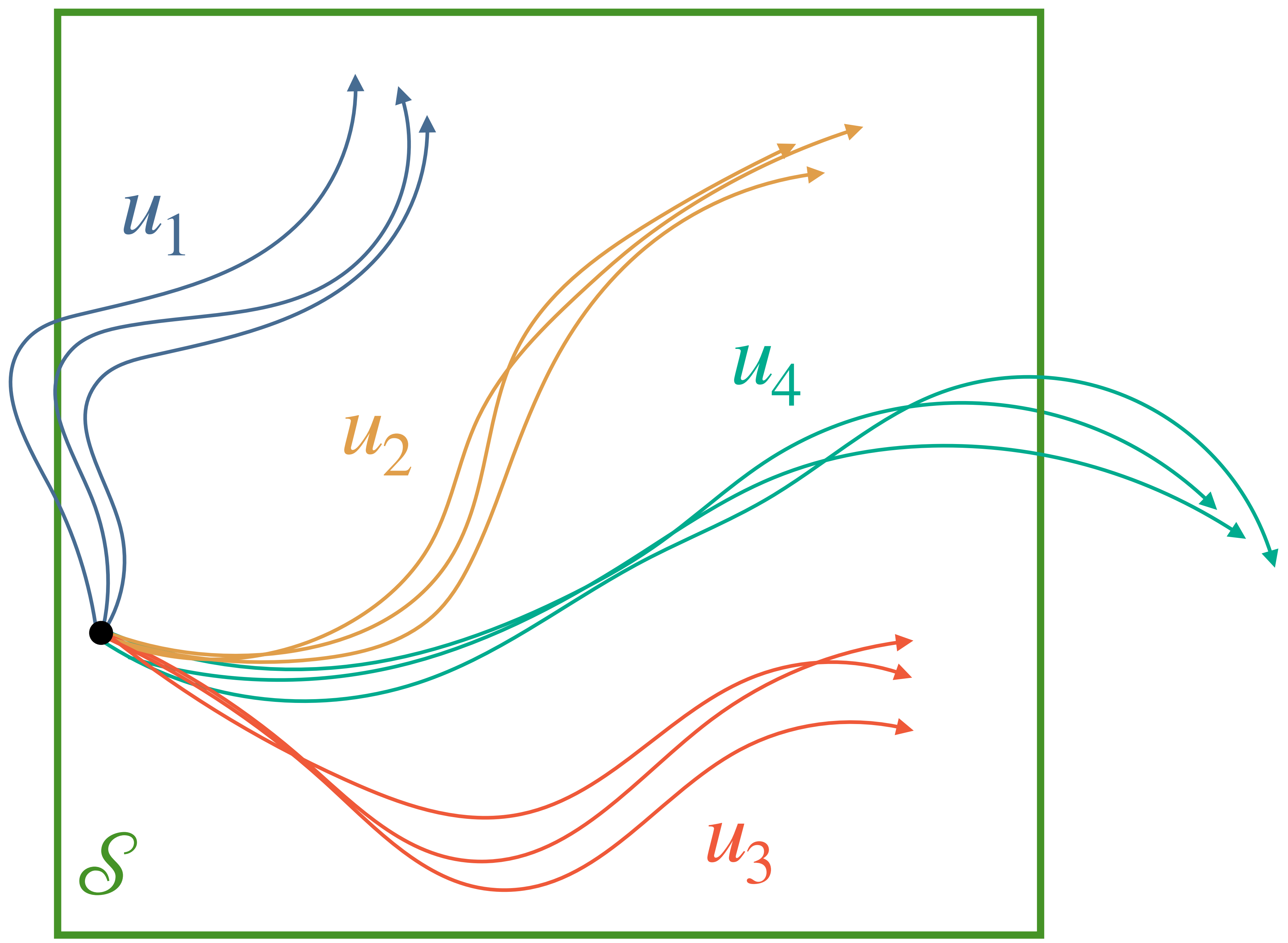
Control Parametrization

$$u^\theta : [0,T] \times \mathbb{R}^n \to \mathbb{R}^d$$

$$\theta \in D \subset \mathbb{R}^m$$

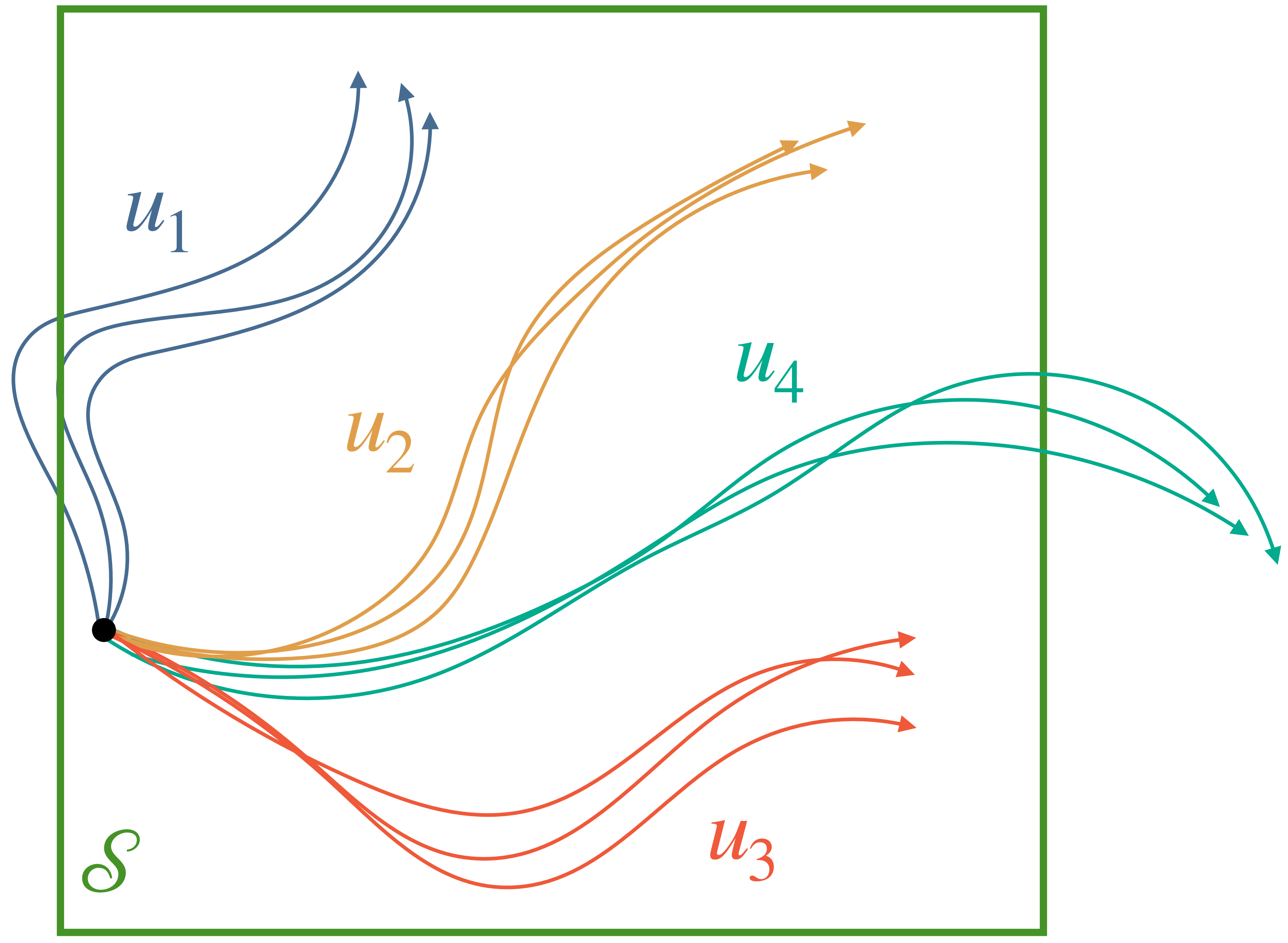
Safe controls and horizons

$$\inf_{t \in [0,T]} \mathbb{P}(X_t^\theta \in \mathcal{S}) \geq 1 - \varepsilon \quad \longrightarrow \quad \text{control } \theta \text{ is safe up to horizon } T$$

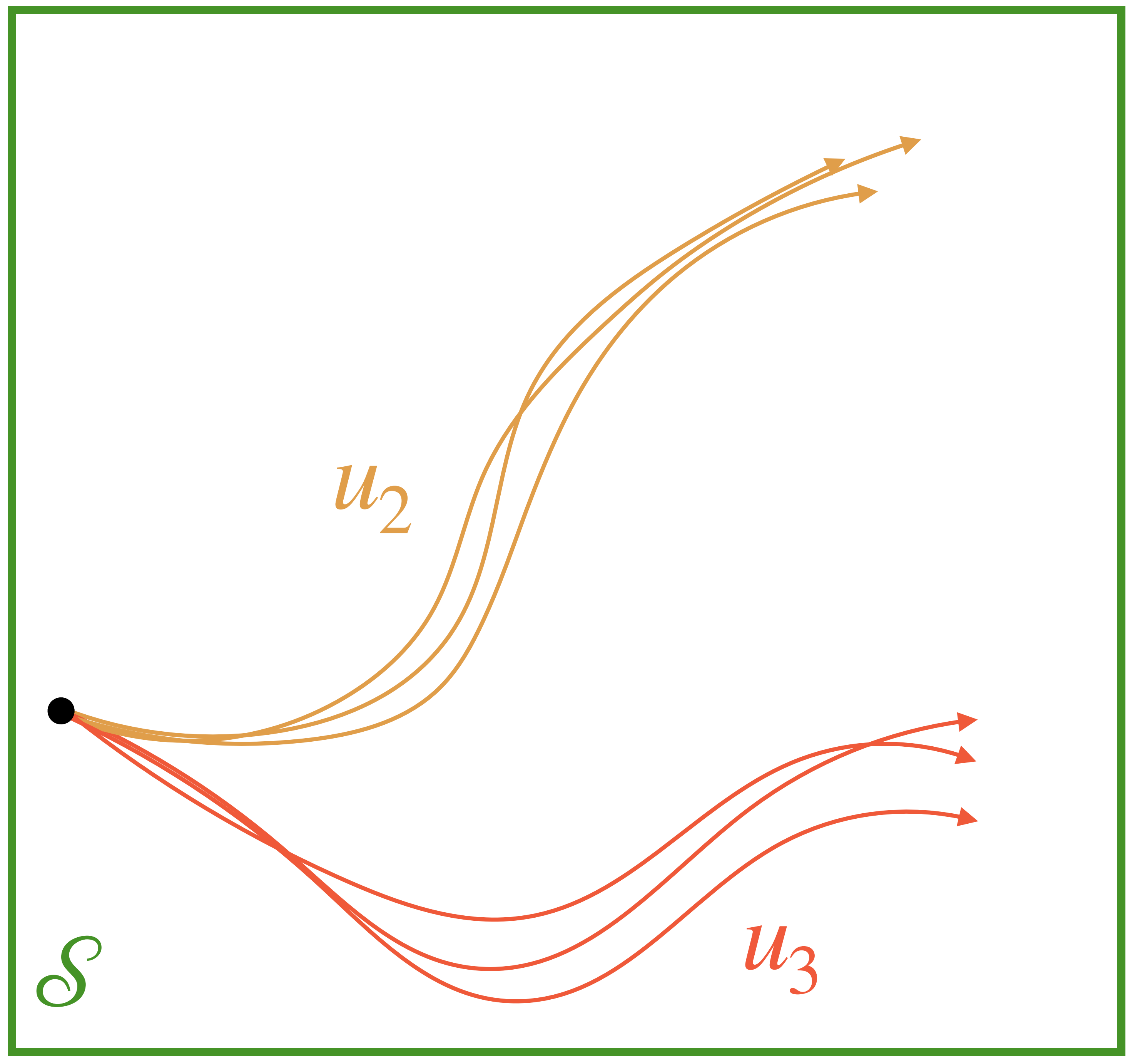


Safe Data Collection

Safe data collection = only safe control executed



Unsafe collection



Safe collection

Learning Problem

Learning the system's density map

$$\theta, t, x \mapsto p_\theta(t,x) = \text{density of } \; X_t^{u_\theta}$$

while ensuring *safe data collection.*

Challenge

Learning $\theta, t, x \mapsto p_\theta(t, x)$ requires data collection,

and safe data collection depends on $\theta, t, x \mapsto p_\theta(t, x)$

Proposed Method

Start from a single a-priori known safe control

Iterative method.

1. Safe data collection thanks to safe control
2. Model update thanks to collected data
3. Sample a new safe control thanks to updated system model

1. Safe Data Collection

Execute a safe controls u_θ and collect $Q \in \mathbb{N}^*$ trajectories

$$\left(X^{u_\theta}(w_i, t_\ell)\right)_{i,l}$$

sampled at times $t_1, \ldots, t_M$, with $M \in \mathbb{N}^*$.

2. Model Update

1. Dynamics model: $\theta, t, x \mapsto \hat{p}(\theta, t, x)$
2. Predictive uncertainty at (θ, t): $\hat{\sigma}(\theta, t)$

2. Model Update

1. Safety probability at t under θ: $\hat{s}(\theta, t) = \int_{\mathcal{S}} \hat{p}_\theta(t, x)\, dx$

2. Set $\hat{\Gamma}$ of certified safe control:

$$(\theta, T) \in \hat{\Gamma} \iff \mathrm{LCB}\Big(\hat{s}(\theta, [0,T])\Big) \geq 1 - \varepsilon$$

3. Sample New Safe Control

Select the most uncertain point inside the certified safe region $\hat{\Gamma}$

$$(\theta_{\text{new}}, t_{\text{new}}) = \arg \max_{(\theta,T)\in\hat{\Gamma},\, t\leq T} \hat{\sigma}(\theta, t)\,.$$

Assumptions

1. Single a-priori known safe control

$$\mathbb{P}(X_t^\theta \in \mathcal{S}) \geq 1 - \varepsilon.$$

2. Sobolev-regular dynamics

$$p \in H^\nu(\mathbb{R}^{n+m+1}),\ \nu > \max(n, m+1)/2,$$

uniform in x and (θ, t).

Certified high-probability safety

At all iterations, the maintained set $\hat{\Gamma}$ contains only safe controls, therefore:

1. Safety holds during training
2. $\hat{\Gamma}$ can be used as a certified of safe controls during deployment

Estimation and sample complexity

For all $(\theta, T) \in \hat{\Gamma}, t \leq T$,

$$\|\hat{p}_\theta(t, \cdot\,) - p_\theta(t, \cdot\,)\|_{L^\infty} \leq c\eta,$$

with accuracy η achieved (up to log factors) using

$$Q \gtrsim N^{\frac{2\nu + n}{2\nu - n}}, \qquad N = \mathcal{O}\left(\eta^{-\frac{2}{1-\alpha}}\right), \quad \alpha > \frac{m+1}{m+1+2\nu}.$$

Smoother dynamics (larger ν) yield *faster convergence* and *lower sample complexity.*

Numerical Experiments

$$dX_t = V_t\,dt, \qquad dV_t = u(t, X_t, V_t)\,dt + a(X_t)\,dW_t,$$

with

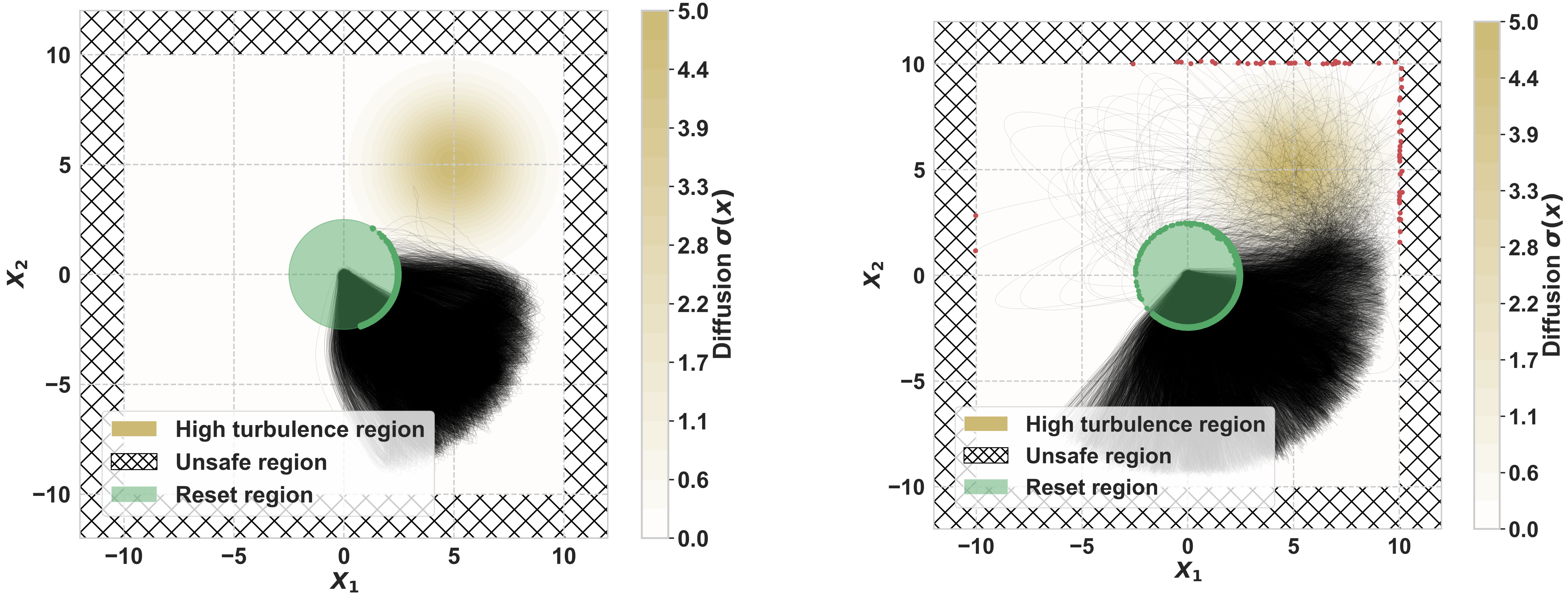
- diffusion $a(X) = Ae^{-\|X - X_c\|^2/2\sigma^2}$,
- safe region $(-10,10)^2$

Numerical Experiments

Control parameterized by two acceleration angles (θ_1, θ_2) with fixed magnitude, followed by a feedback phase steering to the initial region.

Numerical Experiments

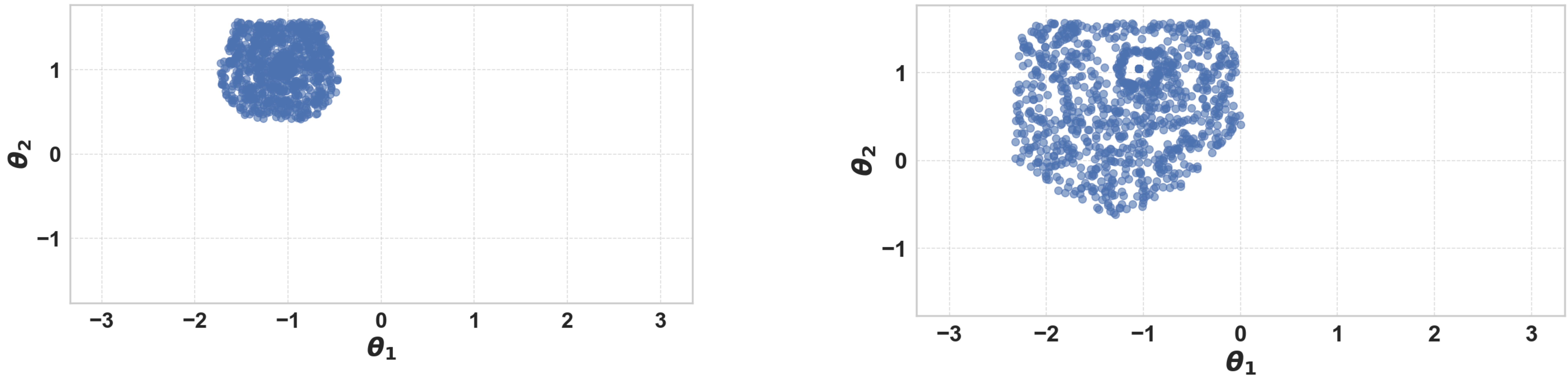
We run 1000 iterations, starting from a single known safe control, under safety thresholds $\varepsilon \in \{0.1, 0.3\}$



Sample trajectories

Numerical Experiments

We run 1000 iterations, starting from a single known safe control, under safety thresholds $\varepsilon \in \{0.1, 0.3\}$



Control coverage

Key takeaways

Safe learning method for controlled stochastic system dynamics

- Certified safety under Sobolev regularity
- Open-source implementation
  - https://github.com/lmotte/dynamics-safe-learn
- Empirical validation:
  - safe exploration
  - accurate dynamics estimation
  - efficient computation