Safe model-based learning for robot control

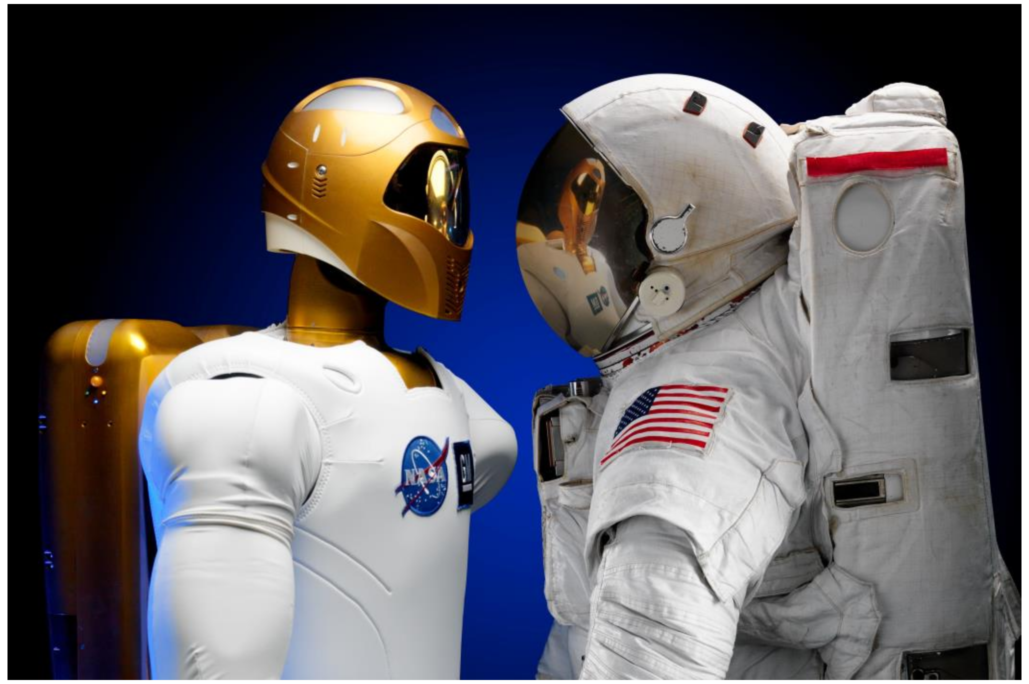
Felix Berkenkamp, Andreas Krause, Angela P. Schoellig

@CDC Workshop on Learning for Control

16th December 2018

Institute for Aerospace Studies
UNIVERSITY OF TORONTO

The future of automation



The future of automation

Large prior uncertainties, active decision making

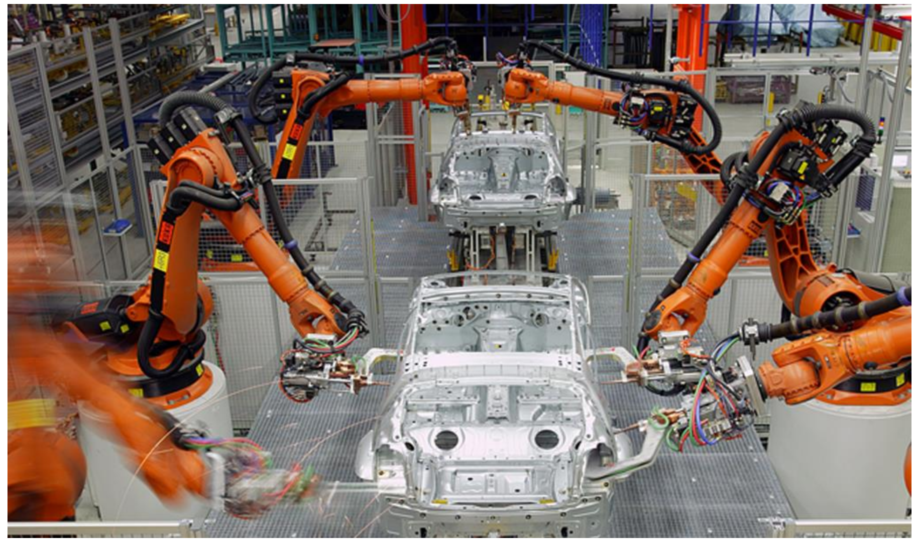
Need **safe** and **high-performance** behavior

Control approach

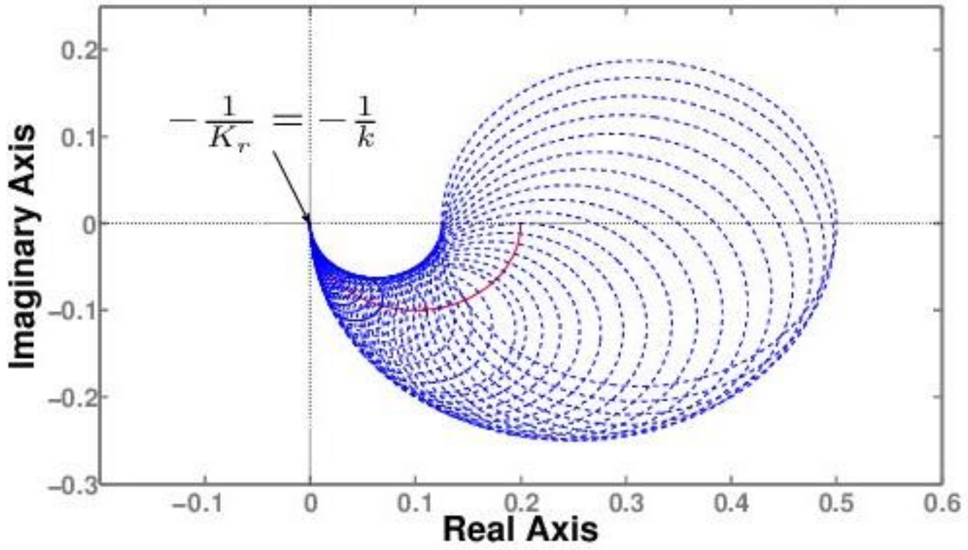
System model → System identification → Controller design

data collection

Controlled environments



Robustness towards errors



Safety constraints

$$\min_{u_{1:T}} \sum_{k=0}^{T} r(x_t, u_t)$$

$$x_{k+1} = f(x_k, u_k)$$

$$x_k \in \mathcal{X}_k,\ \ u_k \in \mathcal{U}_k$$

Two approaches

Control
(Systems)

+ Models
+ Feedback
+ Safety
+ Worst-case
- Learning
- Data




Systems must **learn and adapt**

Performance limited by system understanding

Reinforcement learning approach

~~System model~~ → Data samples → Controller optimization

Agent

Reward

State

Action

Environment

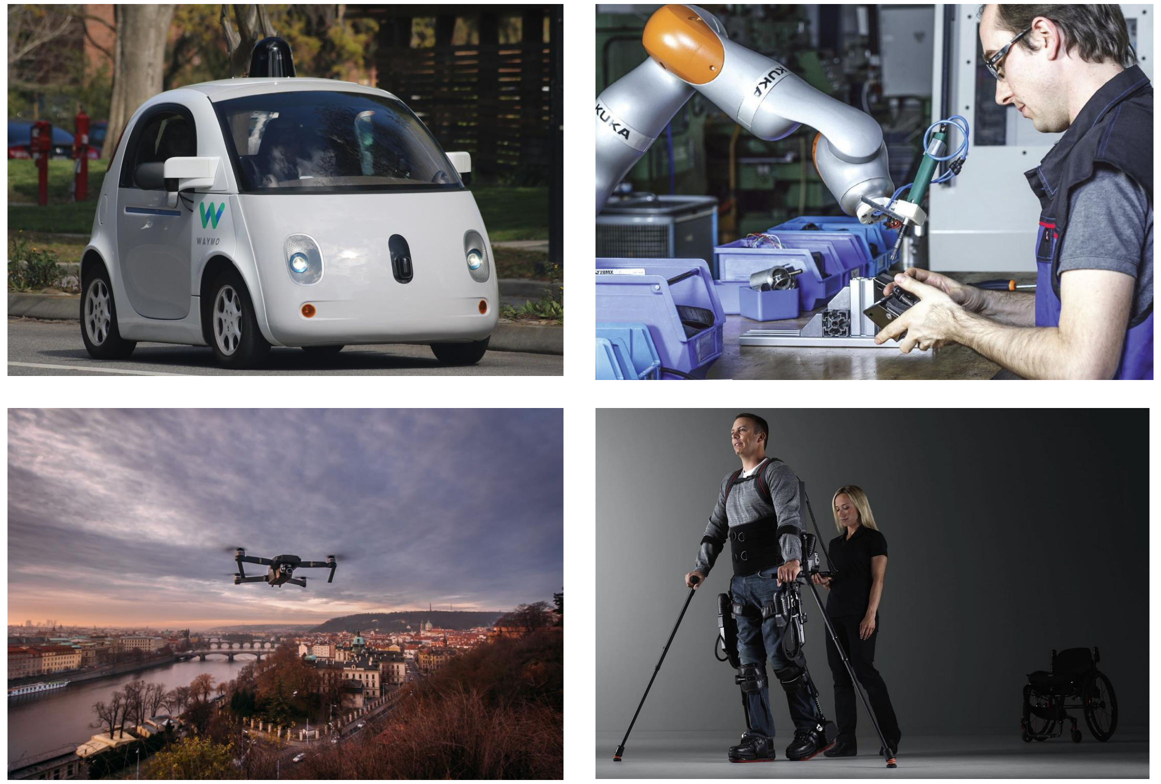
Collecting **relevant data** for the task (in **controlled environments**)

Performance typically in **expectation**

Two approaches

Control
(Systems)

+ Models
+ Feedback
+ Safety
+ Worst-case
- Learning
- Data



Systems must **learn and adapt**
safety, data efficiency

Machine Learning
(Data)

+ Learning
+ Data collection
+ Explore / exploit
+ Average case
- Worst-case
- Safety

Performance limited by system understanding

Model-based reinforcement learning

Safety limited by lack of system understanding

Prerequisites for safe reinforcement learning

Understand model errors and learning dynamics

Define safety, analyze a model for safety

Algorithm to safely acquire data and optimize task

Safe Model-based Reinforcement Learning

Overview

Understand model errors and learning dynamics

Define safety, analyze a model for safety

Algorithm to safely acquire data and optimize task

Safe Model-based Reinforcement Learning

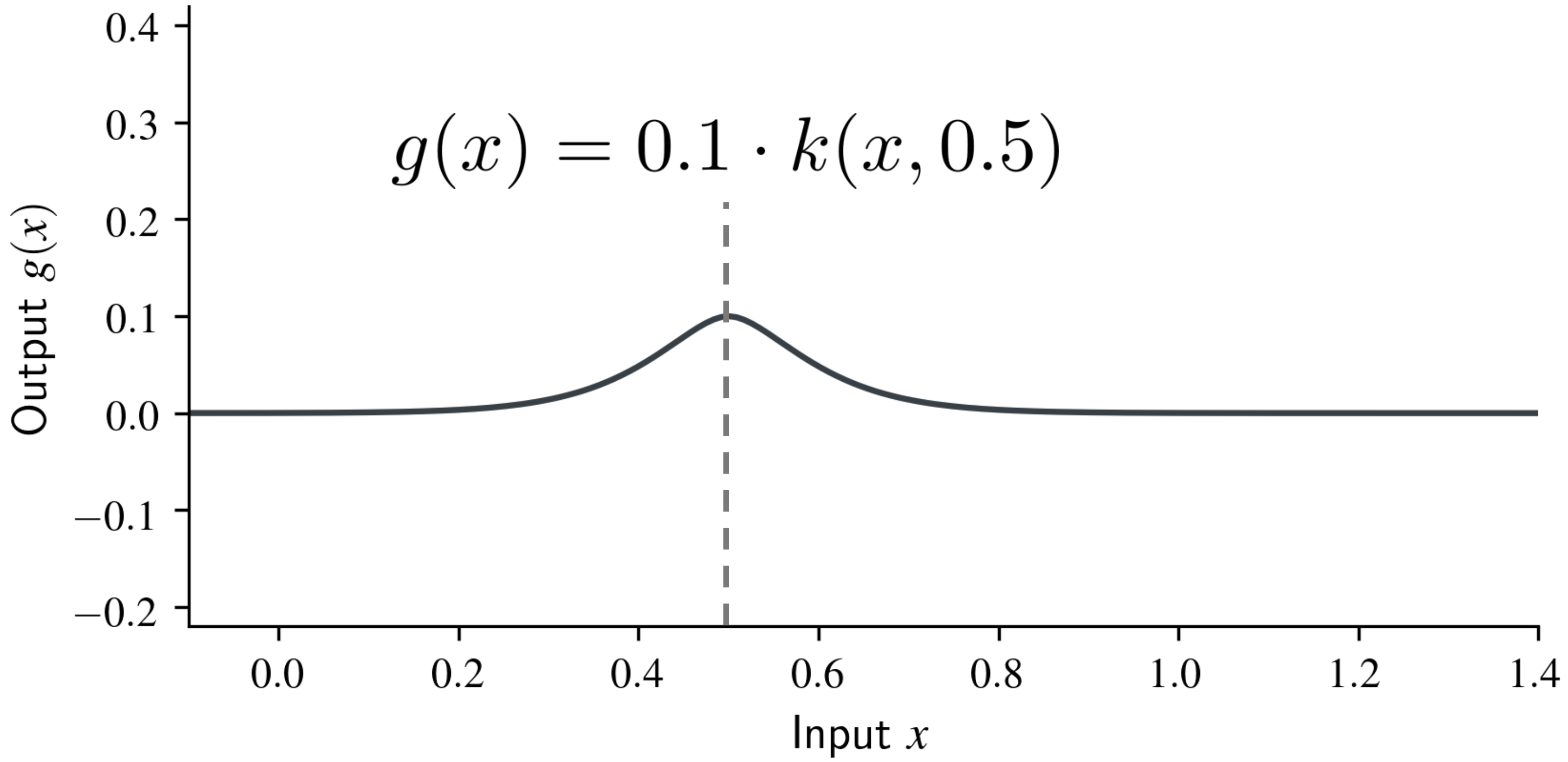
Learning a model

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori\ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$

Need to quantify model error

Model error must decrease with data

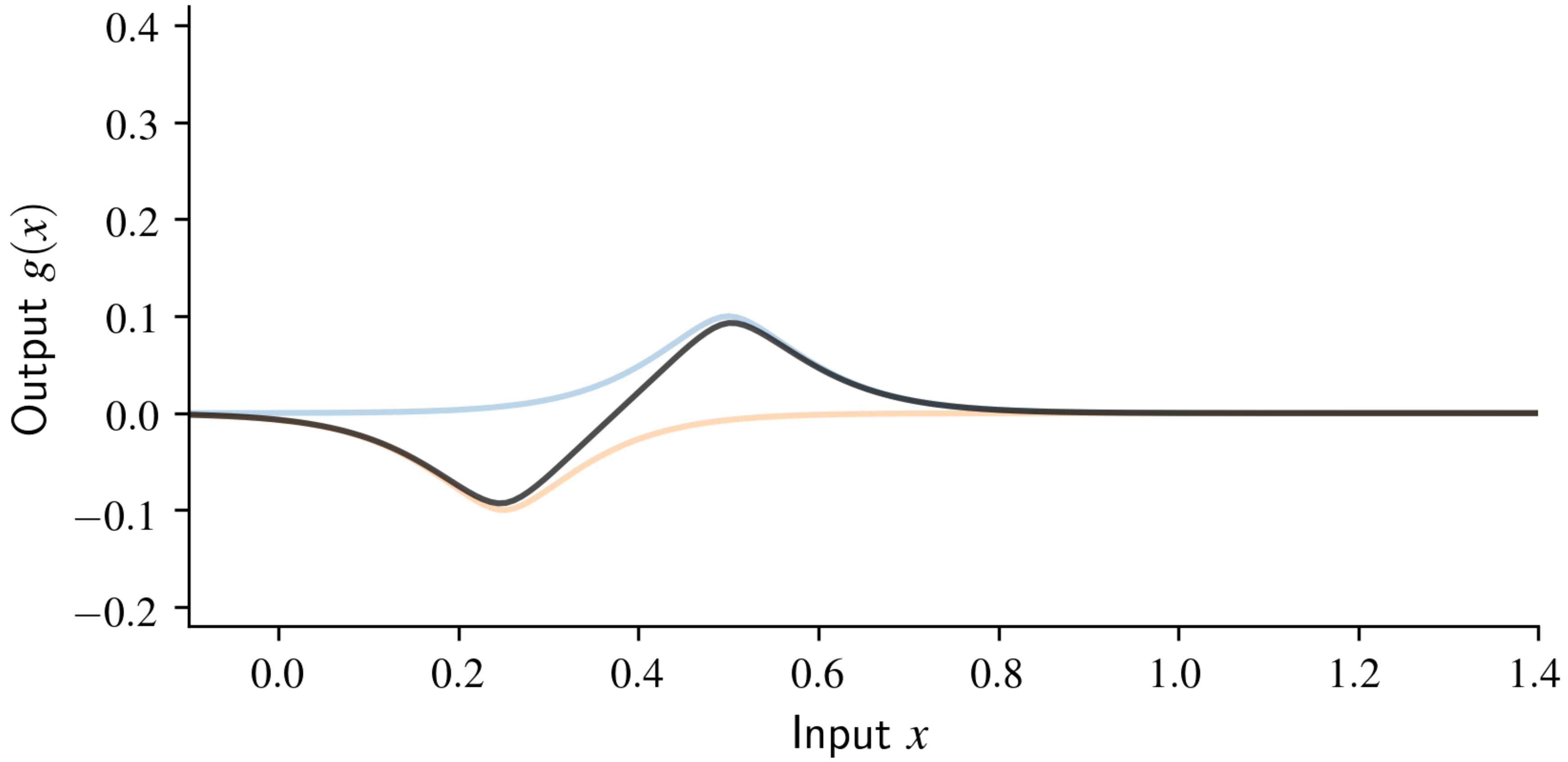


$$g(x) = \sum_{n \geq 0} \alpha_n \, k(x, x_n)$$

Learning a model

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori\ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$



Need to quantify model error

Model error must decrease with data

$$g(x) = \sum_{n \geq 0} \alpha_n \, k(x, x_n)$$

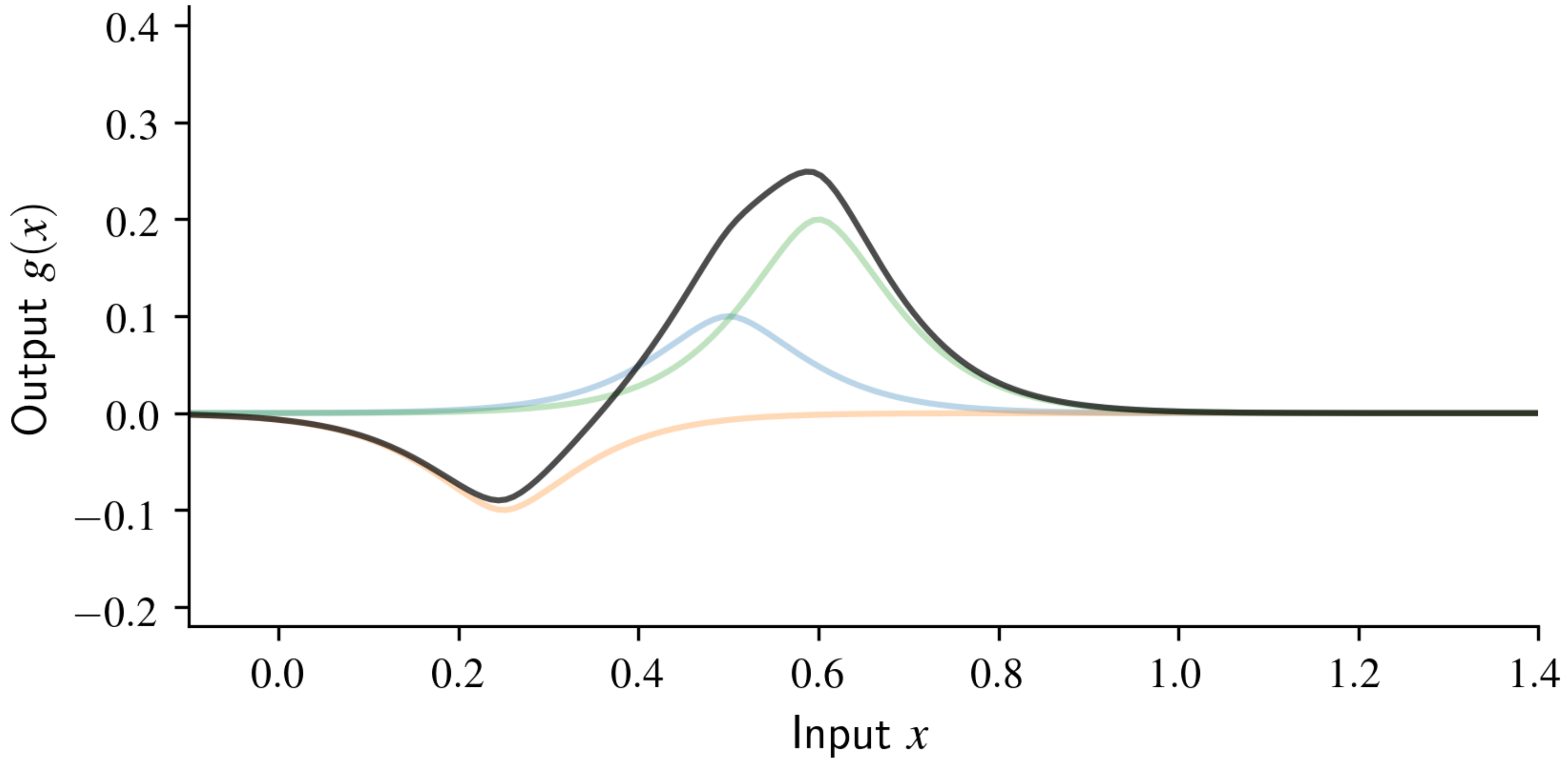
Learning a model

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori\ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$

Need to quantify model error

Model error must decrease with data



$$g(x) = \sum_{n \geq 0} \alpha_n \, k(x, x_n)$$

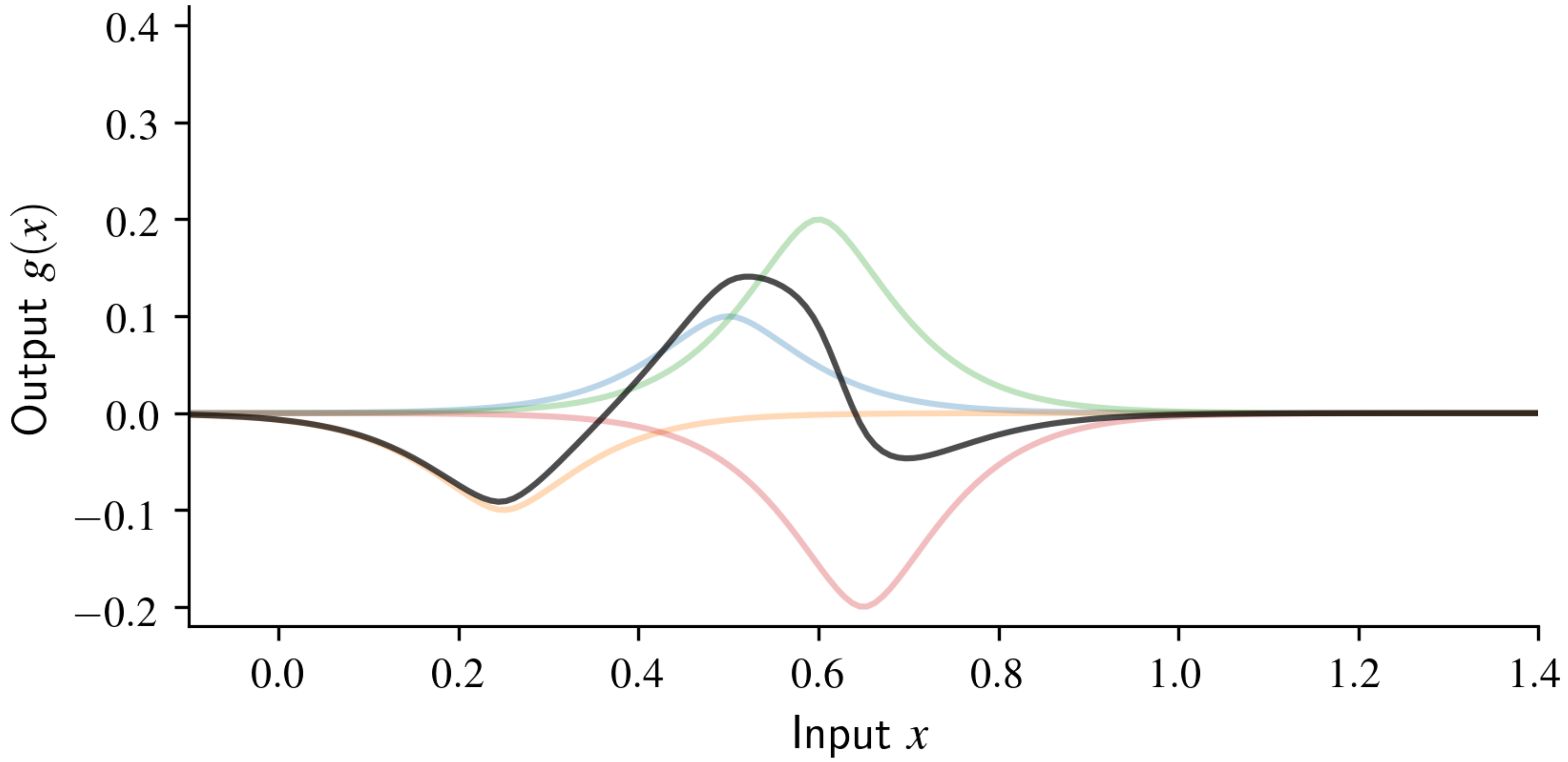
Learning a model

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori\ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$

Need to quantify model error

Model error must decrease with data



$$g(x) = \sum_{n \geq 0} \alpha_n \, k(x, x_n)$$

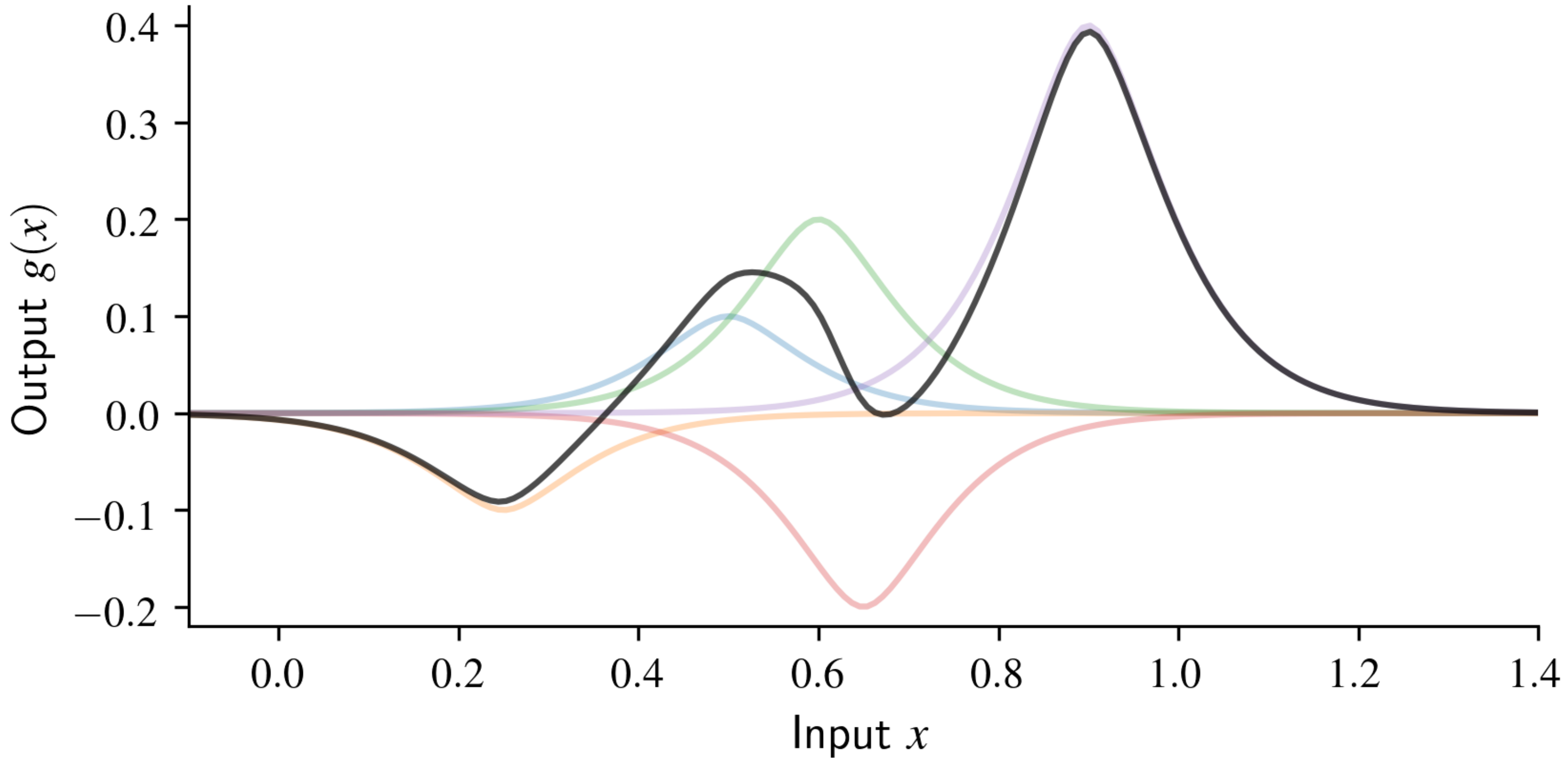
Learning a model

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori\ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$

Need to quantify model error

Model error must decrease with data



$$g(x) = \sum_{n \geq 0} \alpha_n \, k(x, x_n)$$

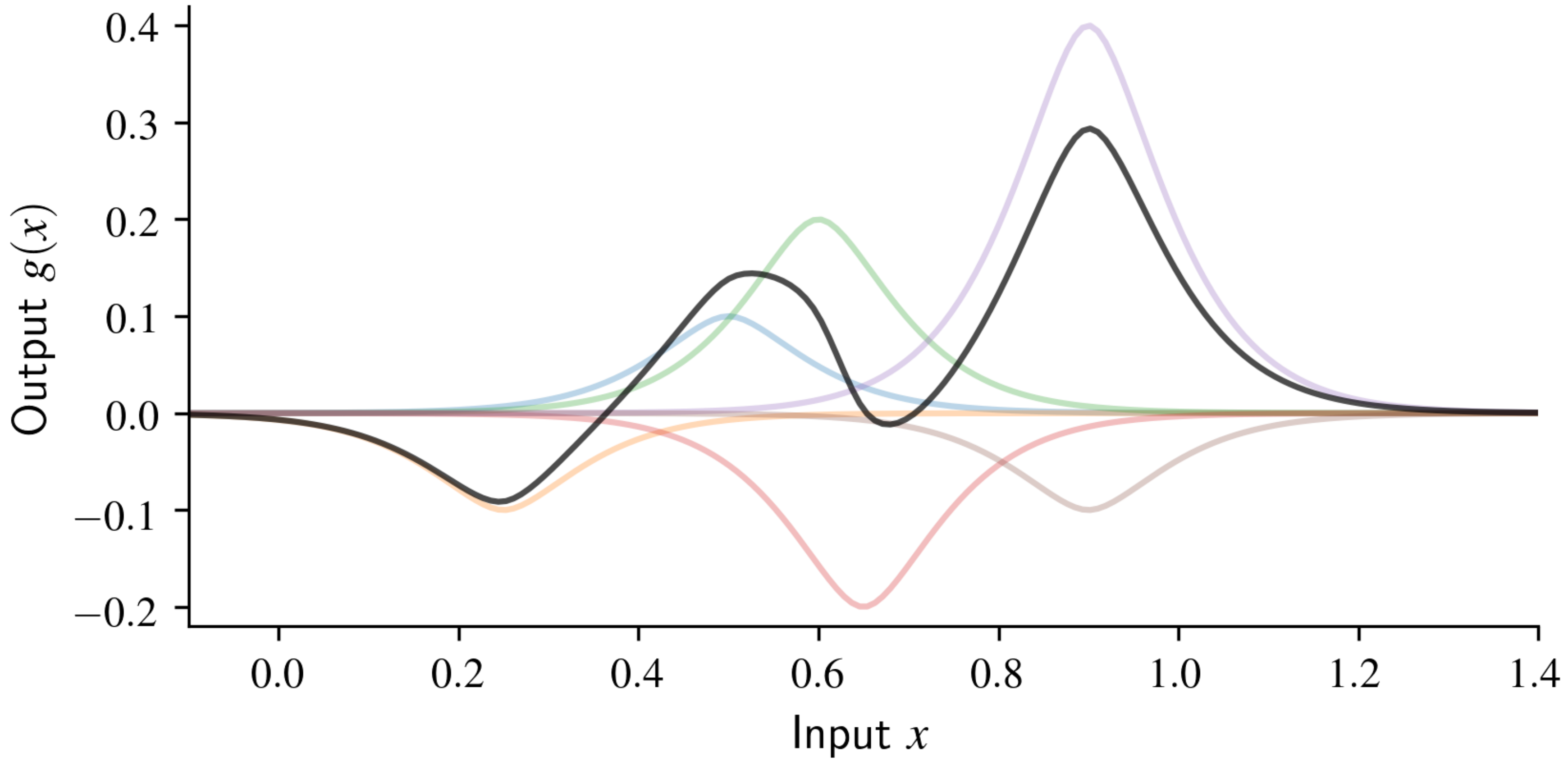
Learning a model

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori \text{ model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$

Need to quantify model error

Model error must decrease with data

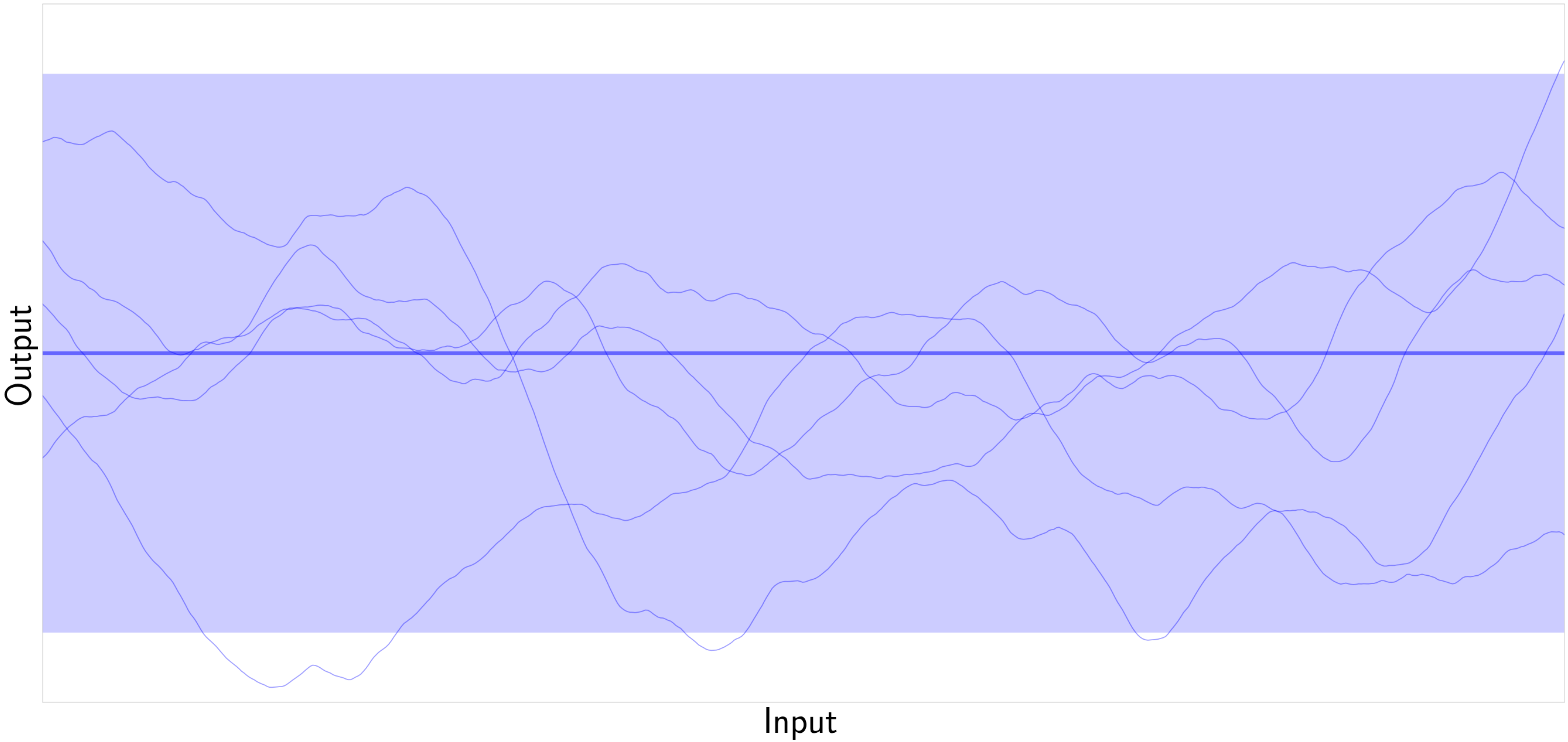


$$g(x) = \sum_{n \geq 0} \alpha_n \, k(x, x_n)$$

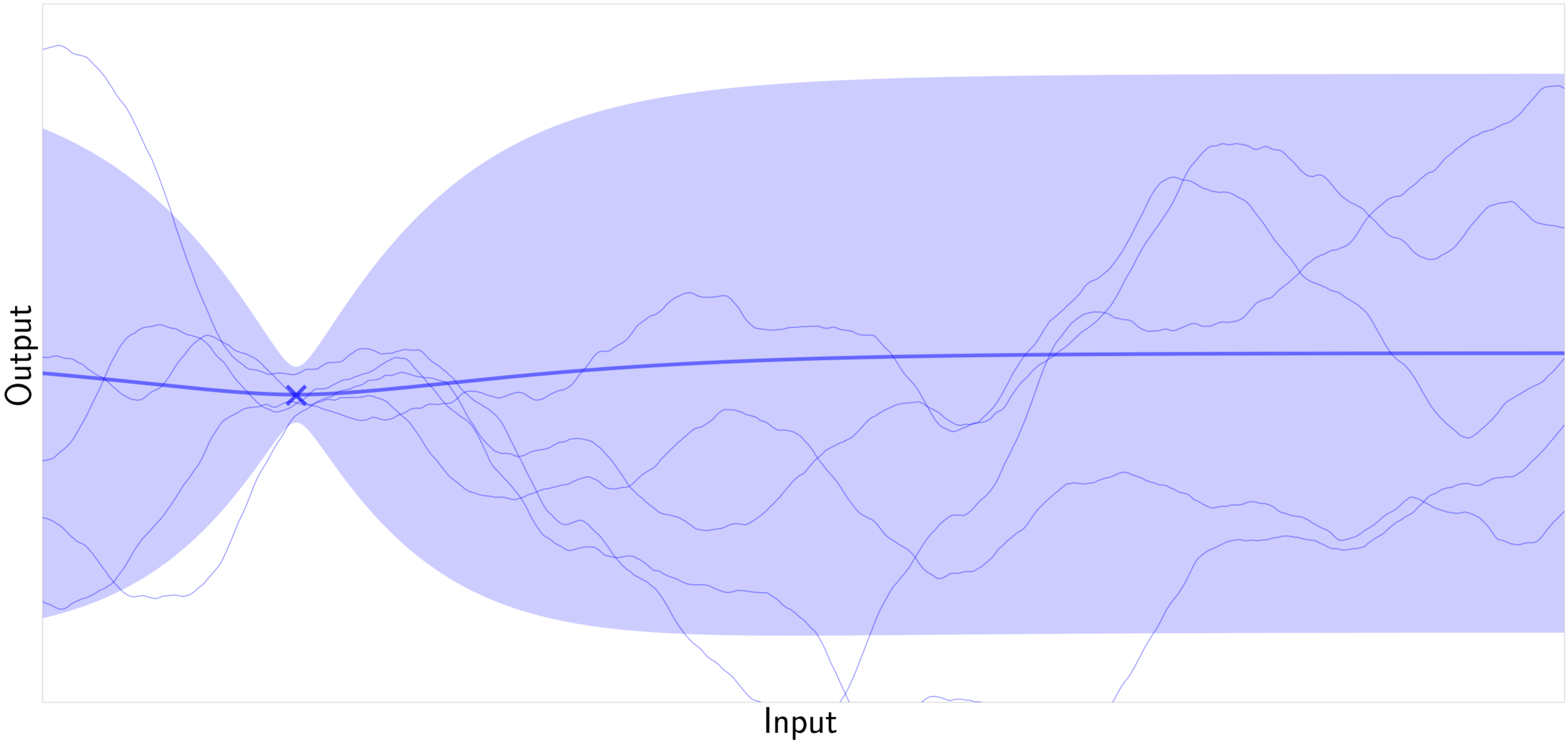
$$\sum_{n>0} \sum_{m>0} \alpha_n \, \alpha_m \, k(x_n, x_m) \leq B$$

ϵ_t sub-Gaussian

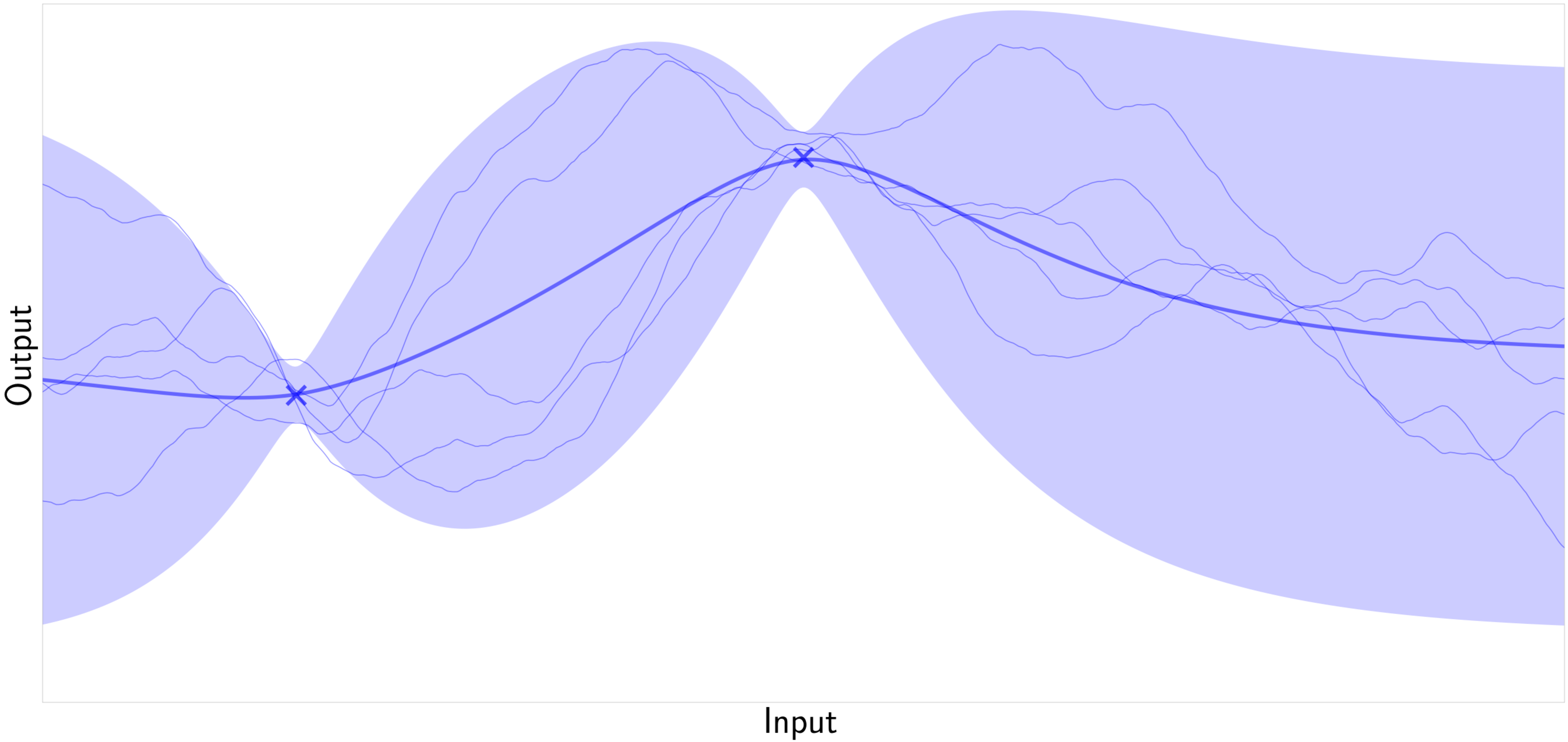
Gaussian process



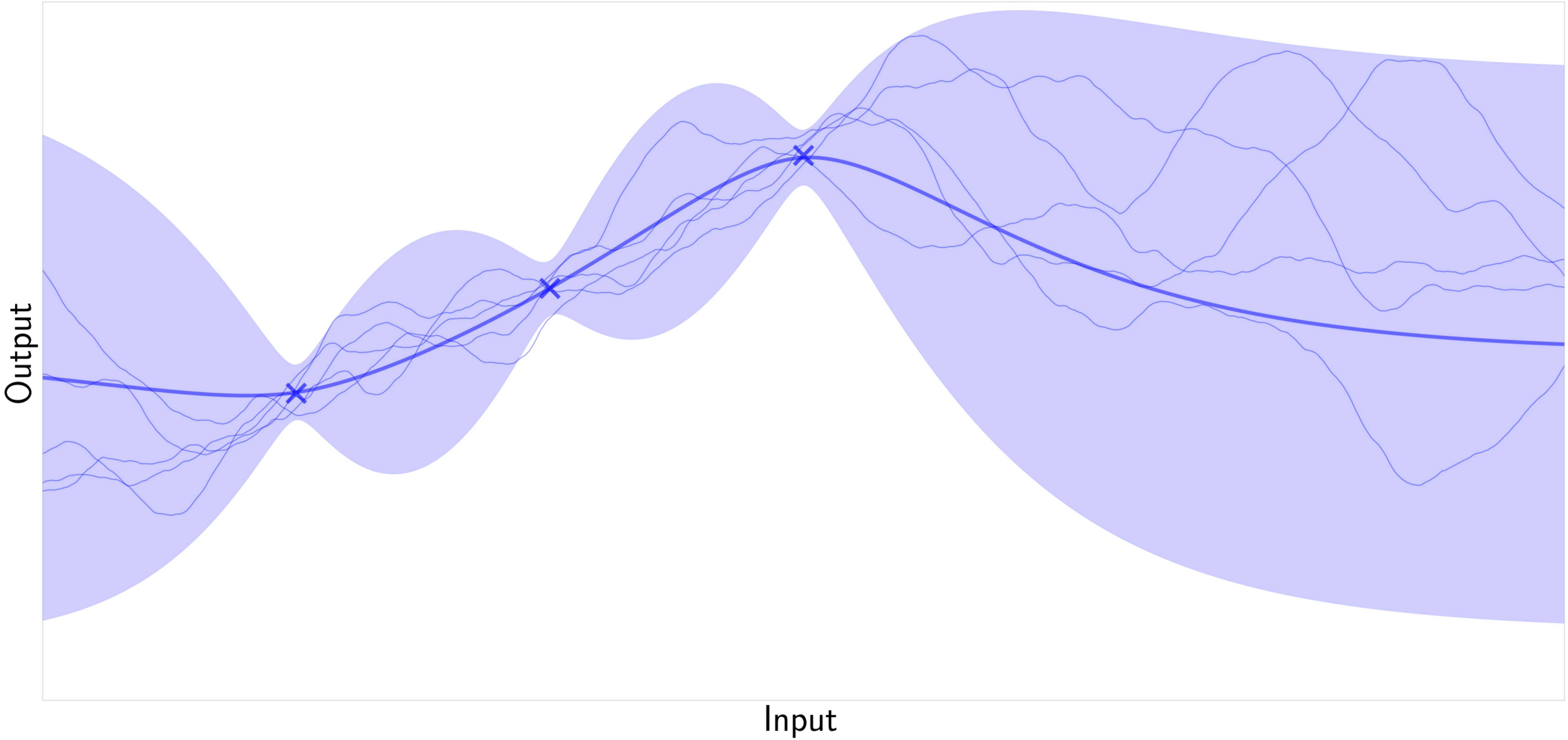
Gaussian process



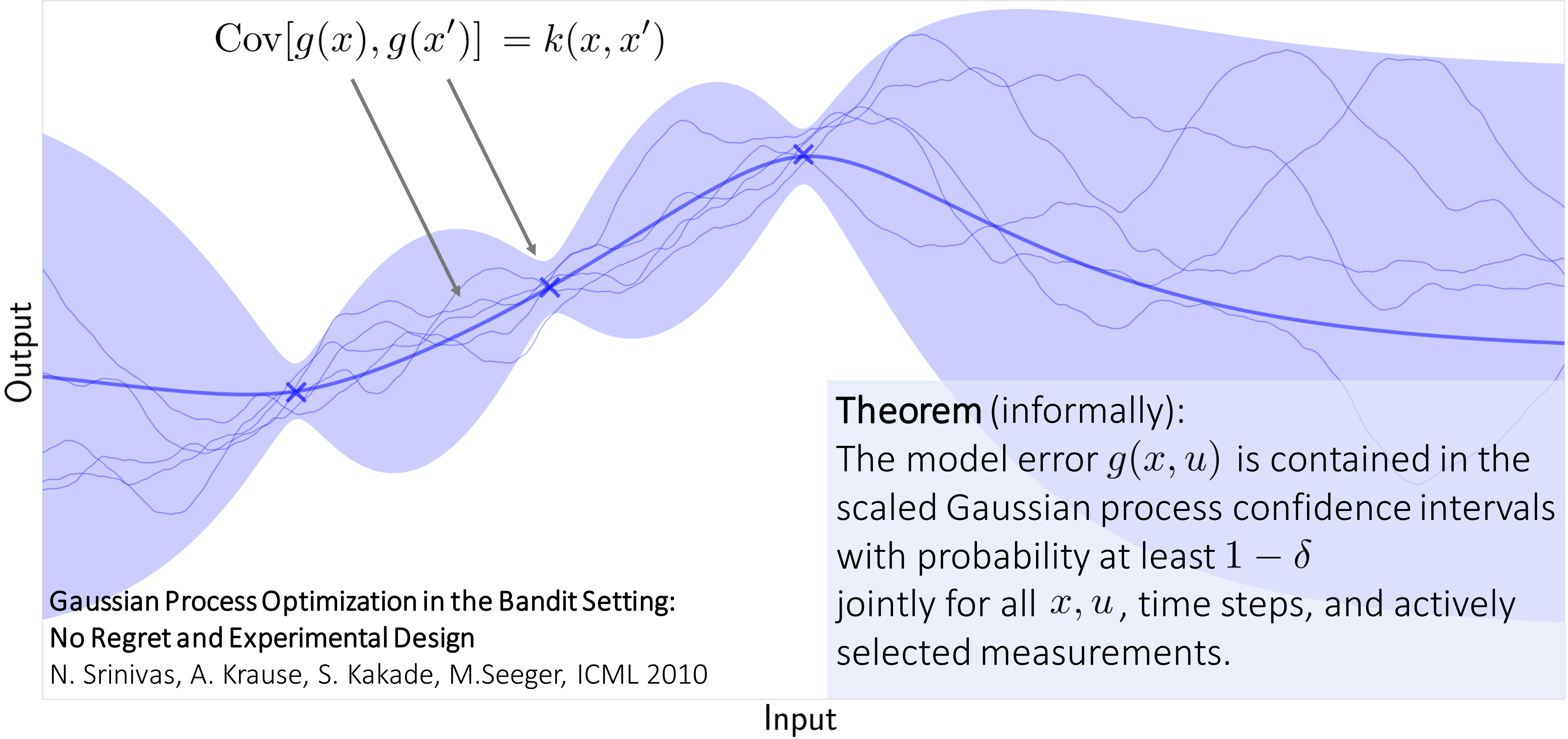
Gaussian process



Gaussian process



Gaussian process



Theorem (informally):
The model error $g(x, u)$ is contained in the scaled Gaussian process confidence intervals with probability at least $1 - \delta$ jointly for all x, u, time steps, and actively selected measurements.

Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design
N. Srinivas, A. Krause, S. Kakade, M.Seeger, ICML 2010

A Bayesian dynamics model

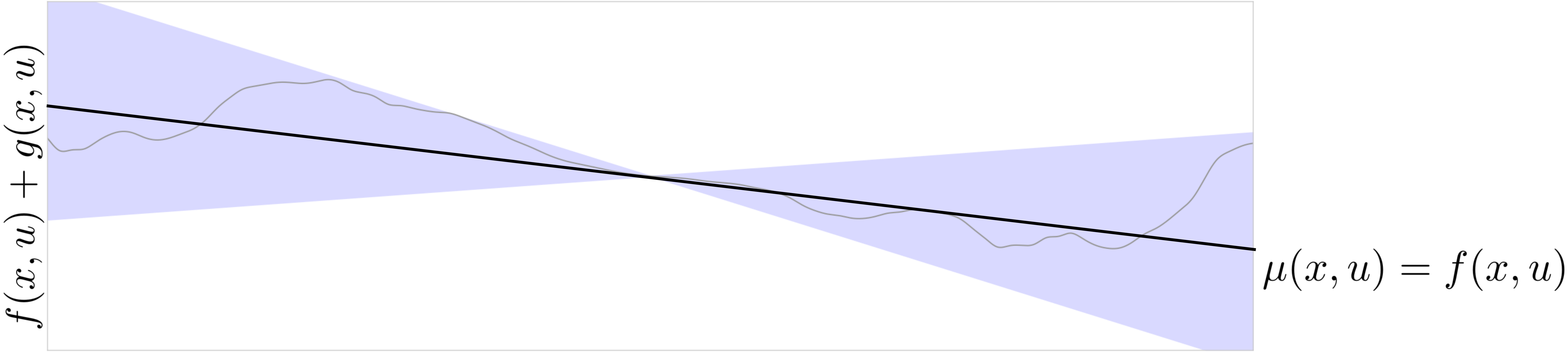
Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a \ priori \ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$

A Bayesian dynamics model

Dynamics

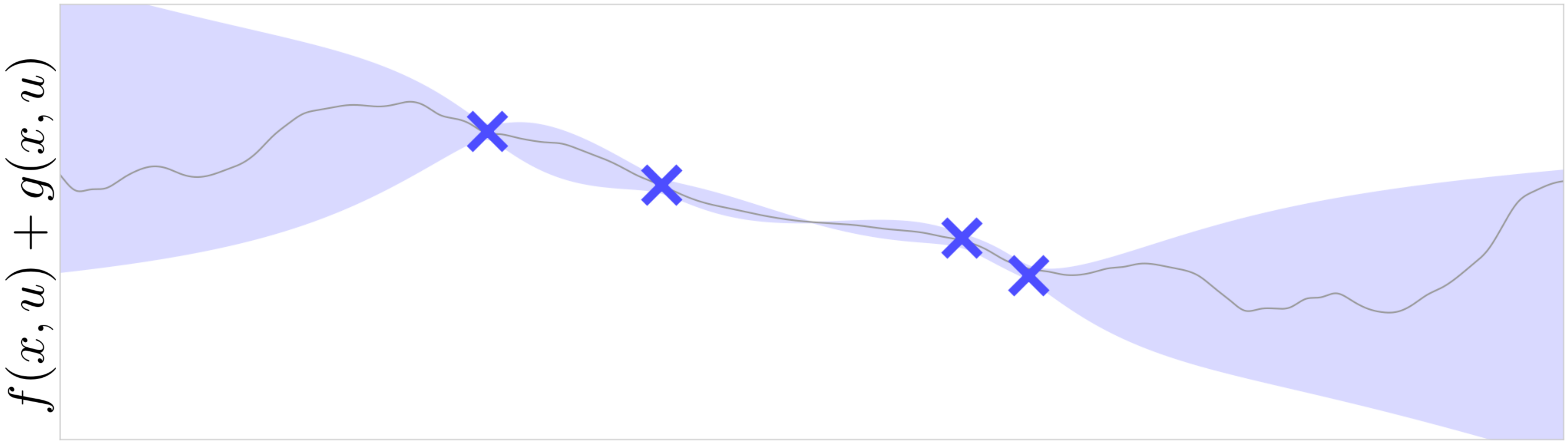
$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori\ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$



A Bayesian dynamics model

Dynamics

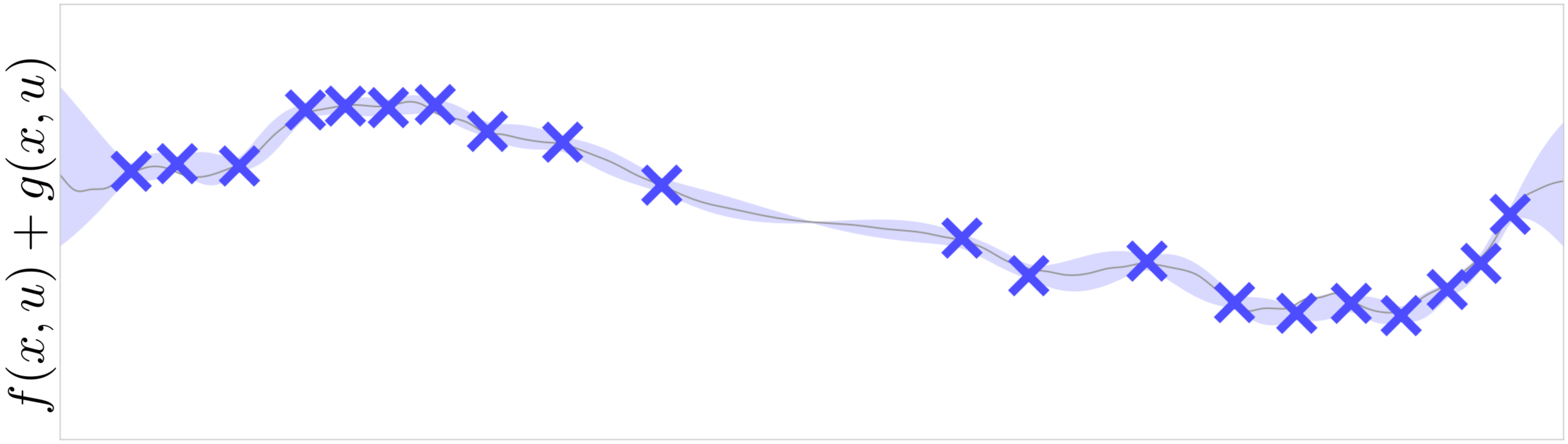
$$x_{t+1} = \underbrace{f(x_t, u_t)}_{\textit{a priori} \text{ model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$



A Bayesian dynamics model

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori \text{ model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}} + \epsilon_t$$



Overview

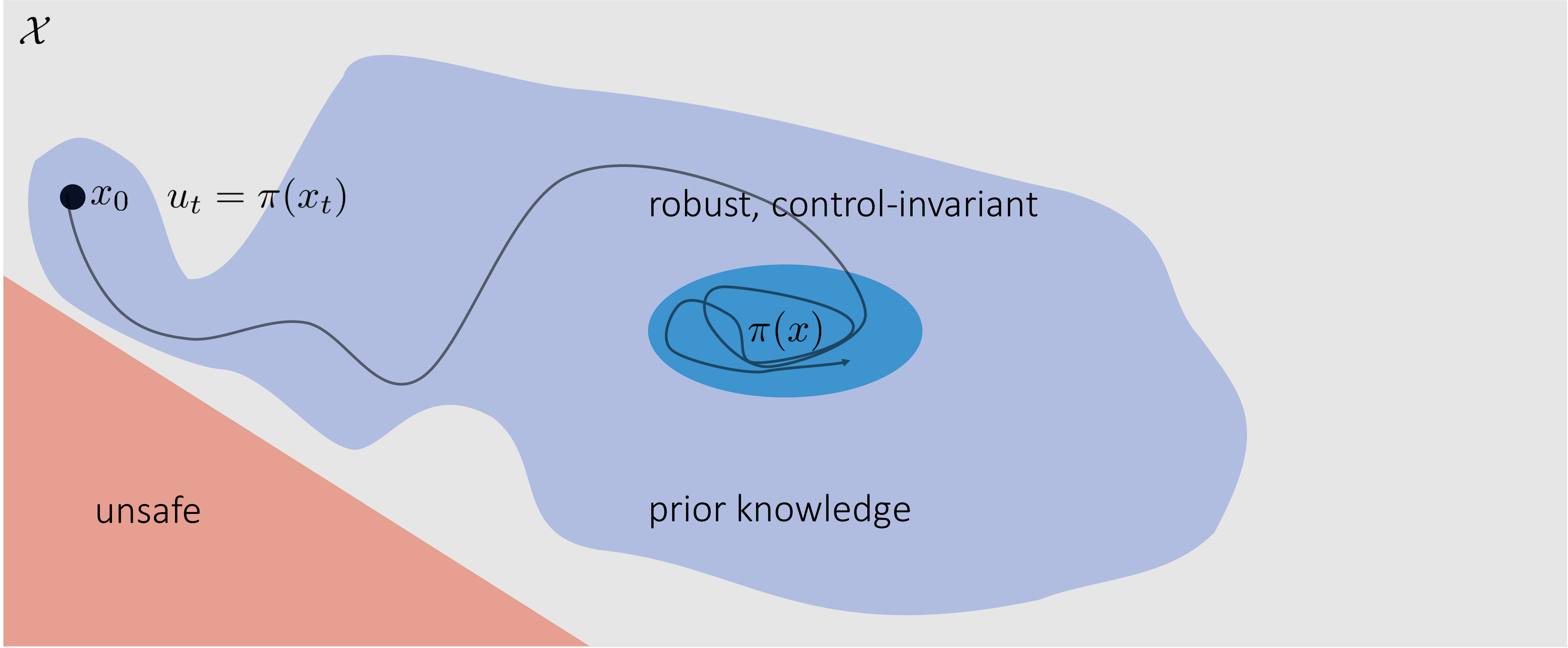
Understand model errors and learning dynamics

Define safety, analyze a model for safety

Algorithm to safely acquire data and optimize task

Safe Model-based Reinforcement Learning

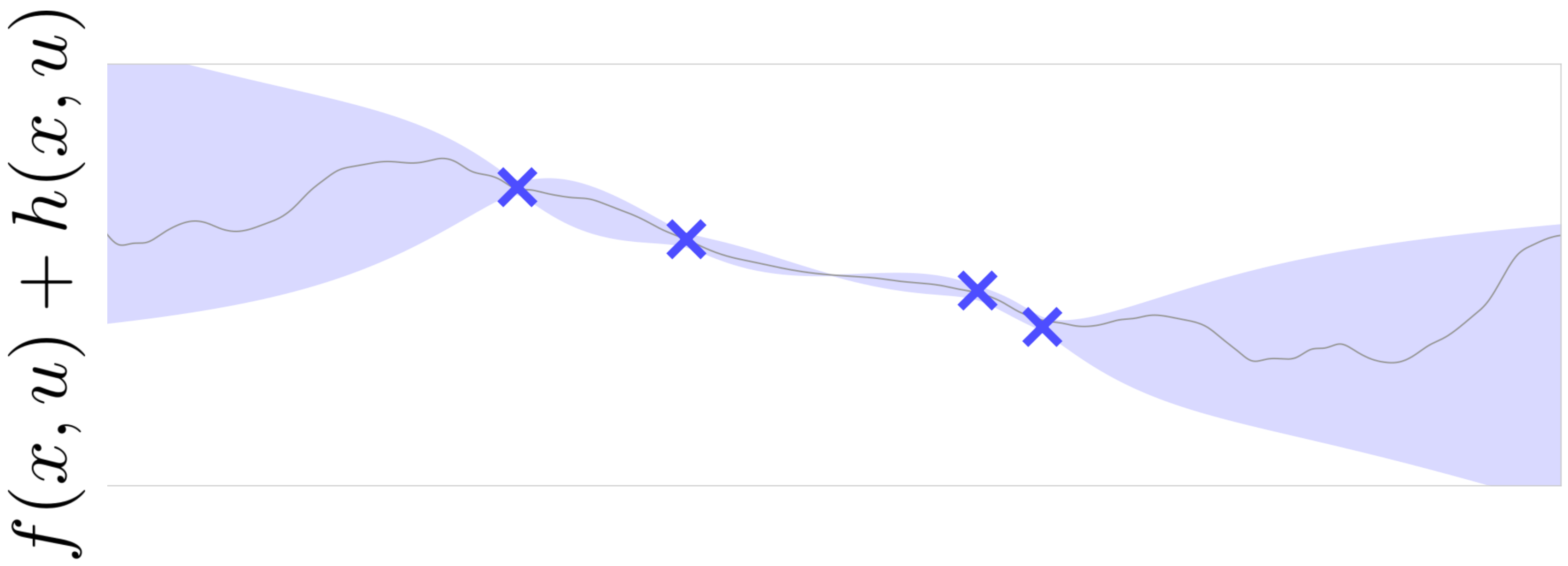
Safety definition



Safety for learned models

Dynamics

$$x_{t+1} = \underbrace{f(x_t, u_t)}_{a\ priori\ \text{model}} + \underbrace{g(x_t, u_t)}_{\text{unknown model}}$$



+

Policy

$$u_t = \pi(x_t)$$

Stability?
Region of attraction?

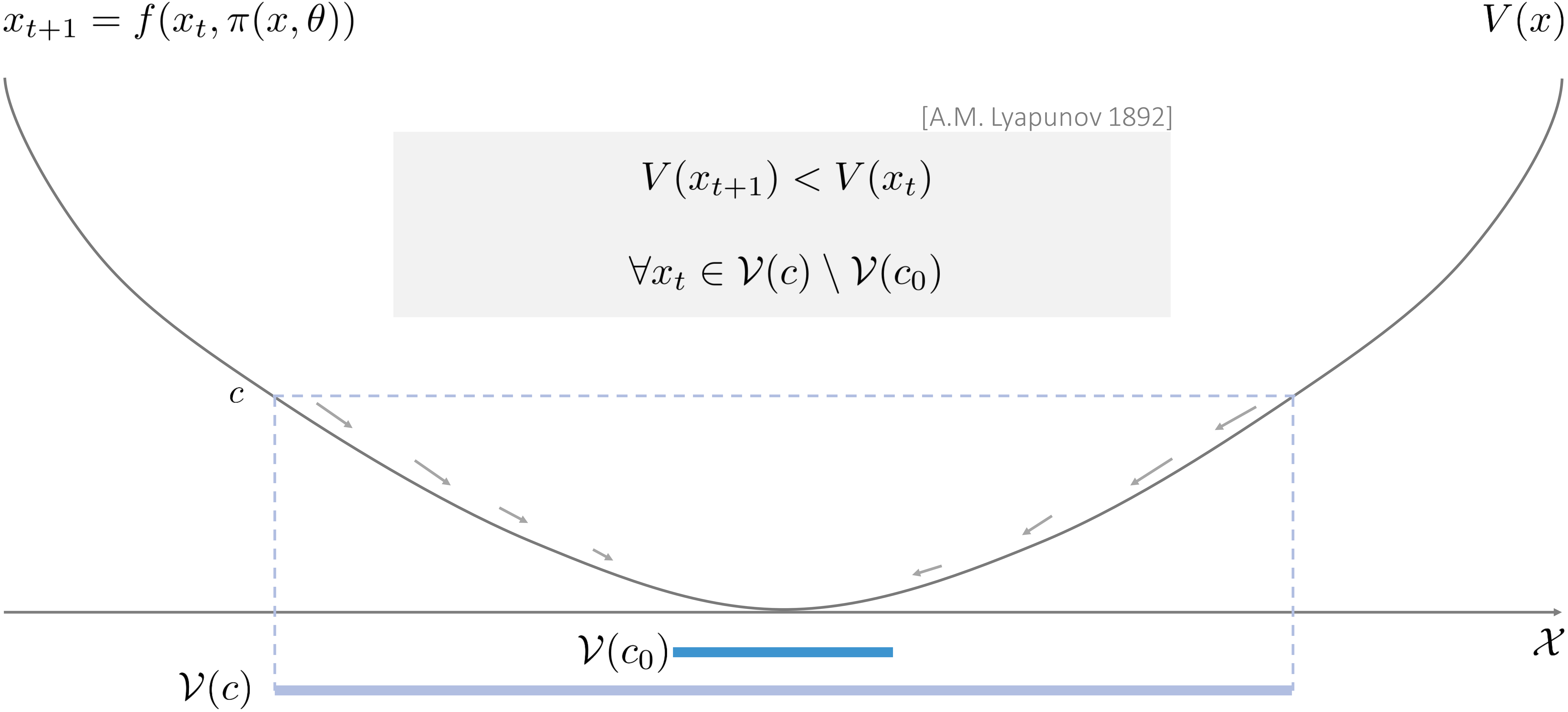
Lyapunov functions

$$x_{t+1} = f(x_t, \pi(x, \theta))$$

[A.M. Lyapunov 1892]

$$V(x_{t+1}) < V(x_t)$$

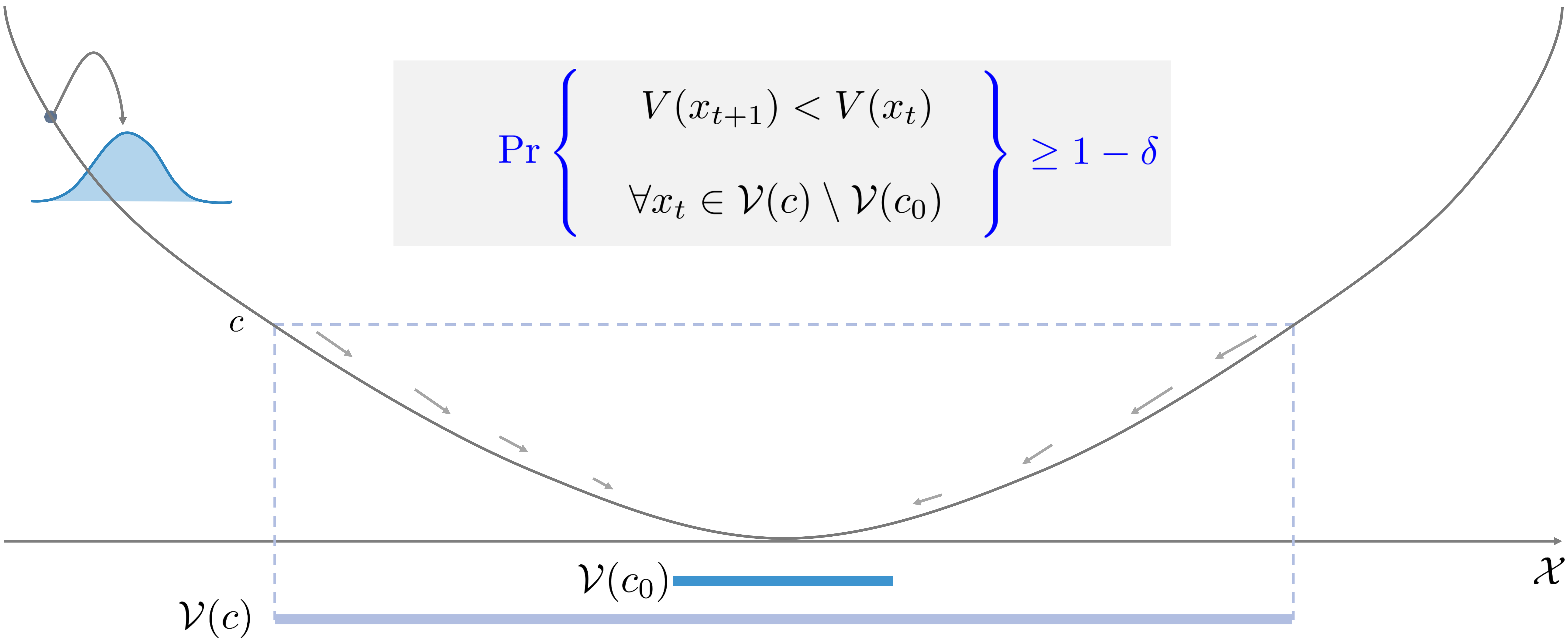
$$\forall x_t \in \mathcal{V}(c) \setminus \mathcal{V}(c_0)$$



Lyapunov functions

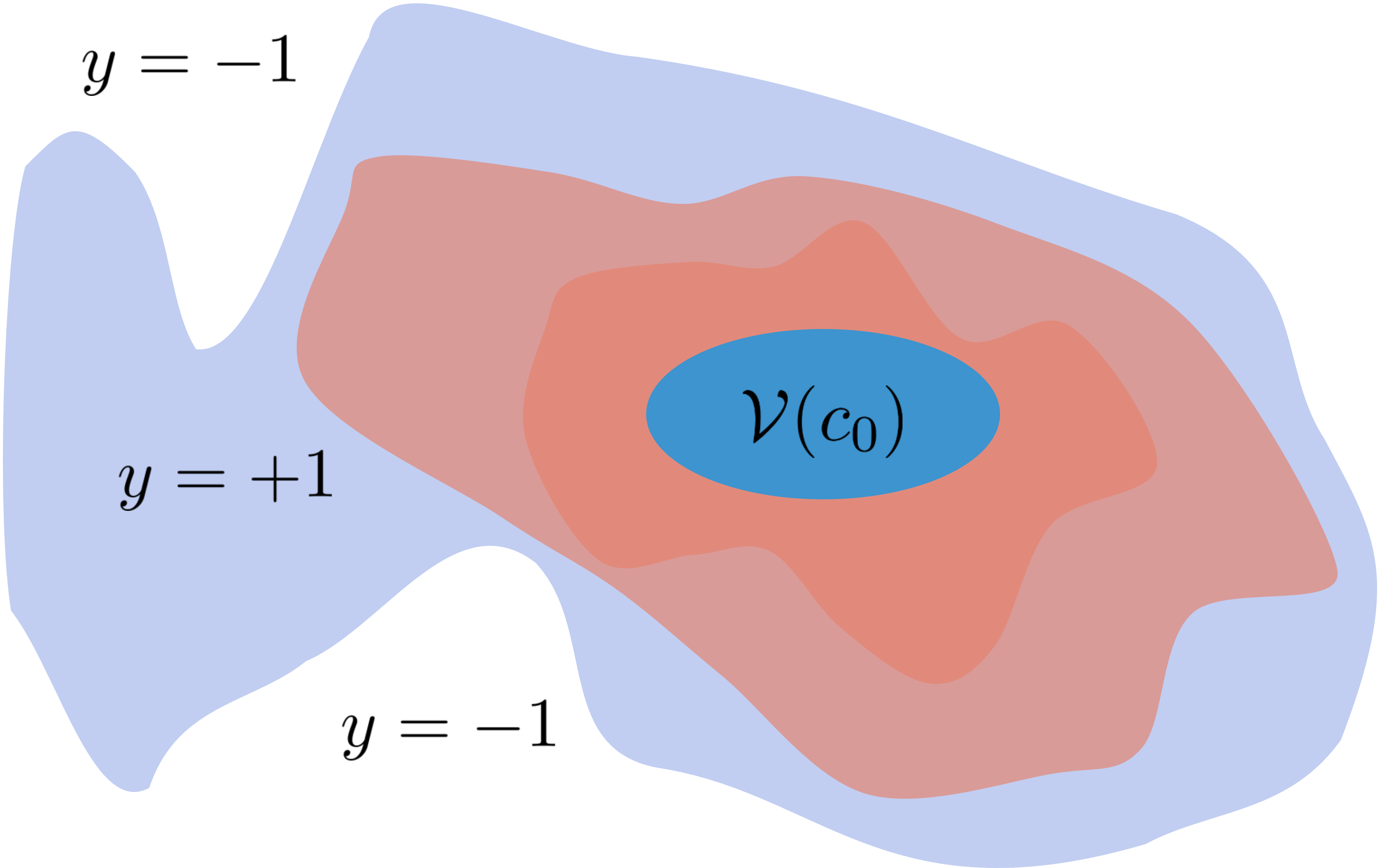
$$x_{t+1} = f(x_t, \pi(x, \theta)) \; + \; g(x_t, \pi(x, \theta))$$

$$\Pr \left\{ \begin{array}{c} V(x_{t+1}) < V(x_t) \\ \forall x_t \in \mathcal{V}(c) \setminus \mathcal{V}(c_0) \end{array} \right\} \geq 1 - \delta$$



Learning Lyapunov functions

Finding the right Lyapunov function is difficult!



$$V(x) = \phi_\theta(x)^{\mathrm{T}} \phi_\theta(x)$$

Weights - positive-definite
Nonlinearities - trivial nullspace

Classification problem

The Lyapunov Neural Network: Adaptive Stability Certification for Safe Learning of Dynamic Systems
S.M. Richards, F. Berkenkamp, A. Krause, CoRL 2018

Overview

Understand model errors and learning dynamics

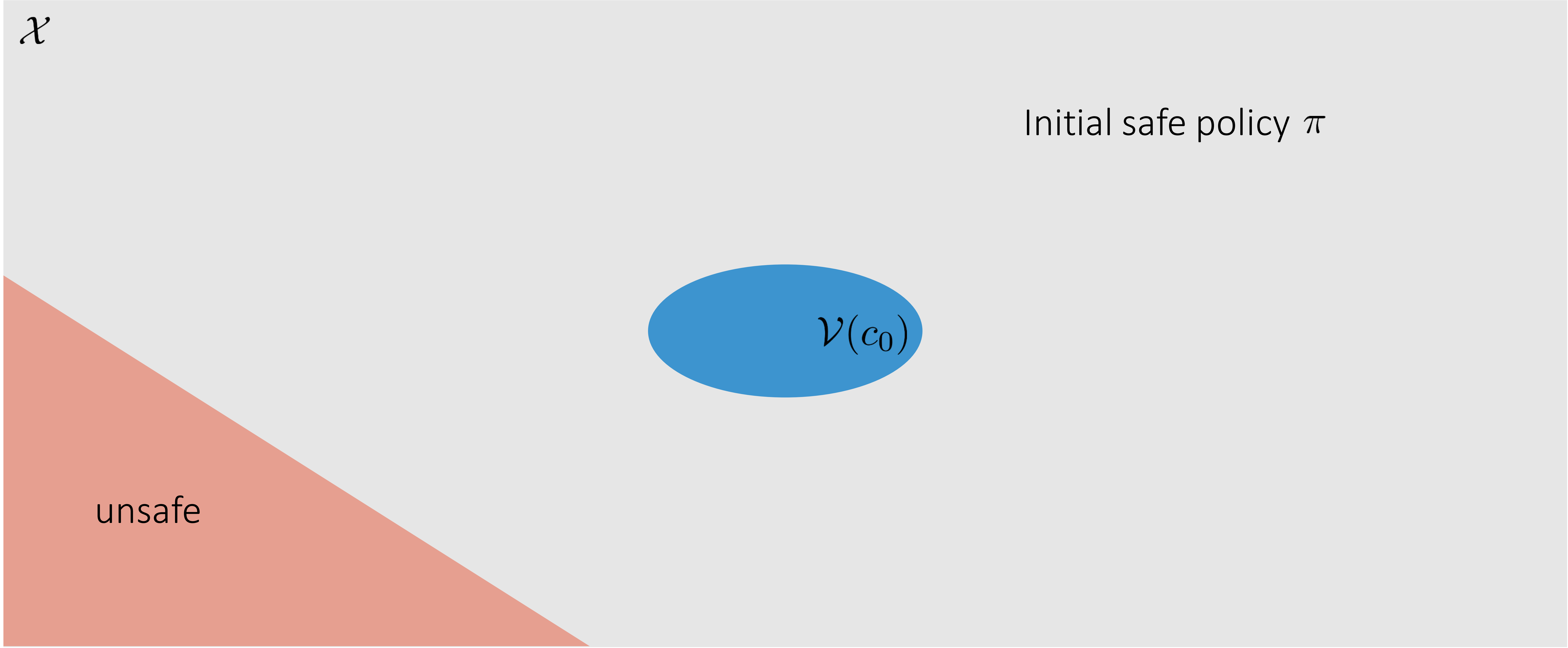
Define safety, analyze a model for safety

Algorithm to safely acquire data and optimize task

Safe Model-based Reinforcement Learning

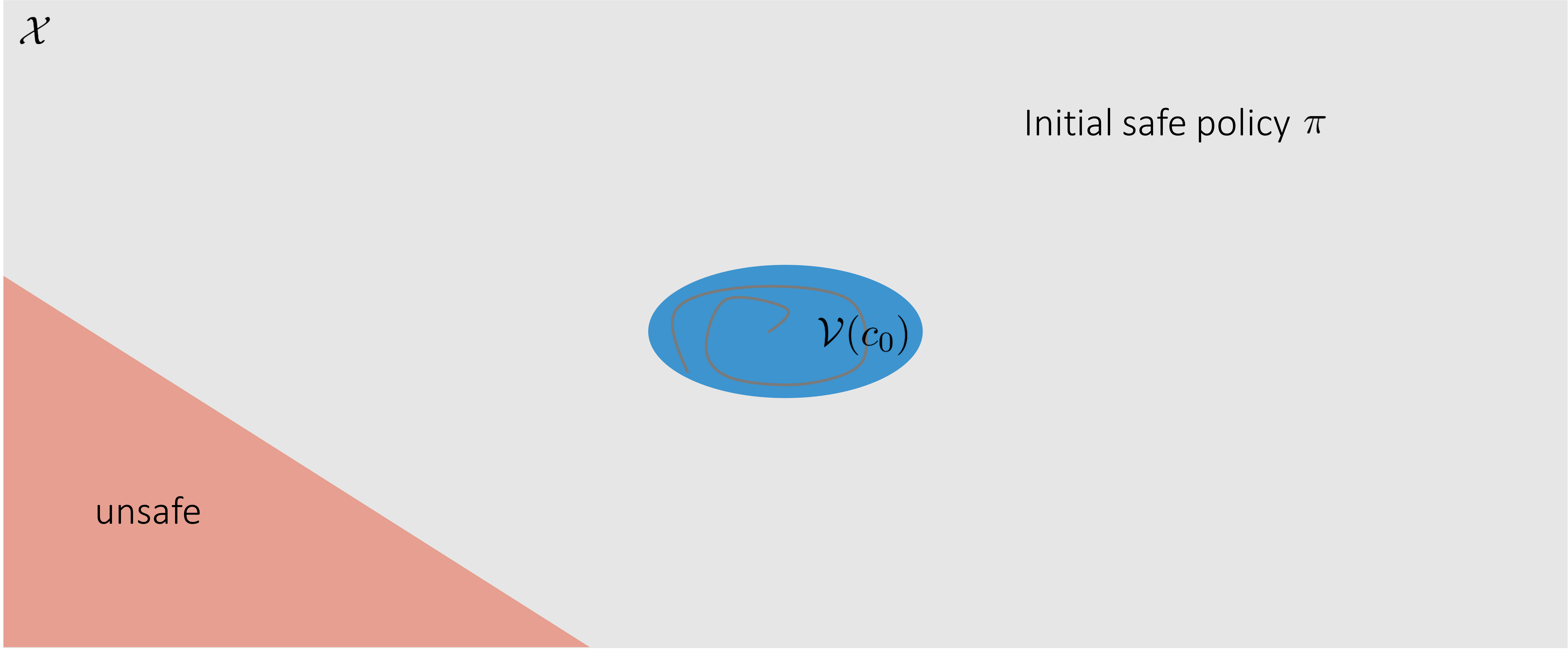
Safety definition

Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017



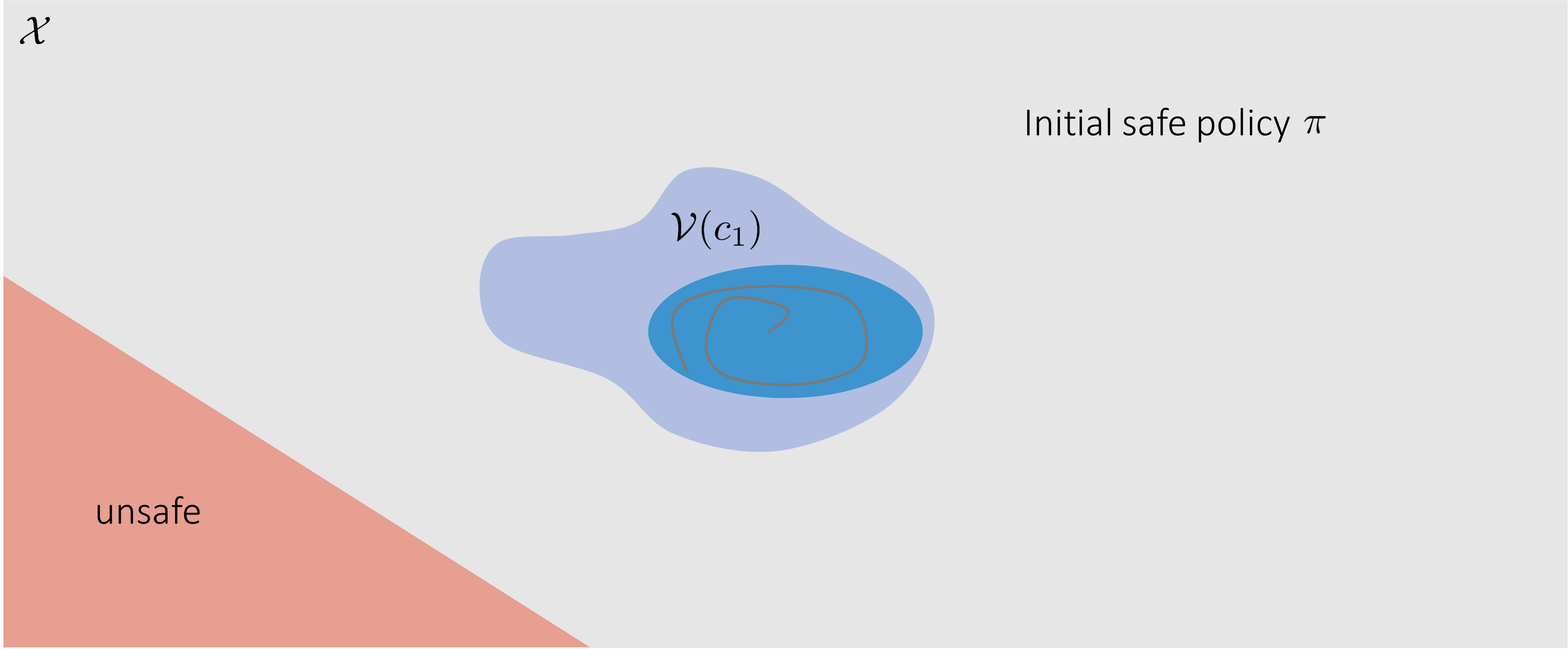
Safety definition

Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017



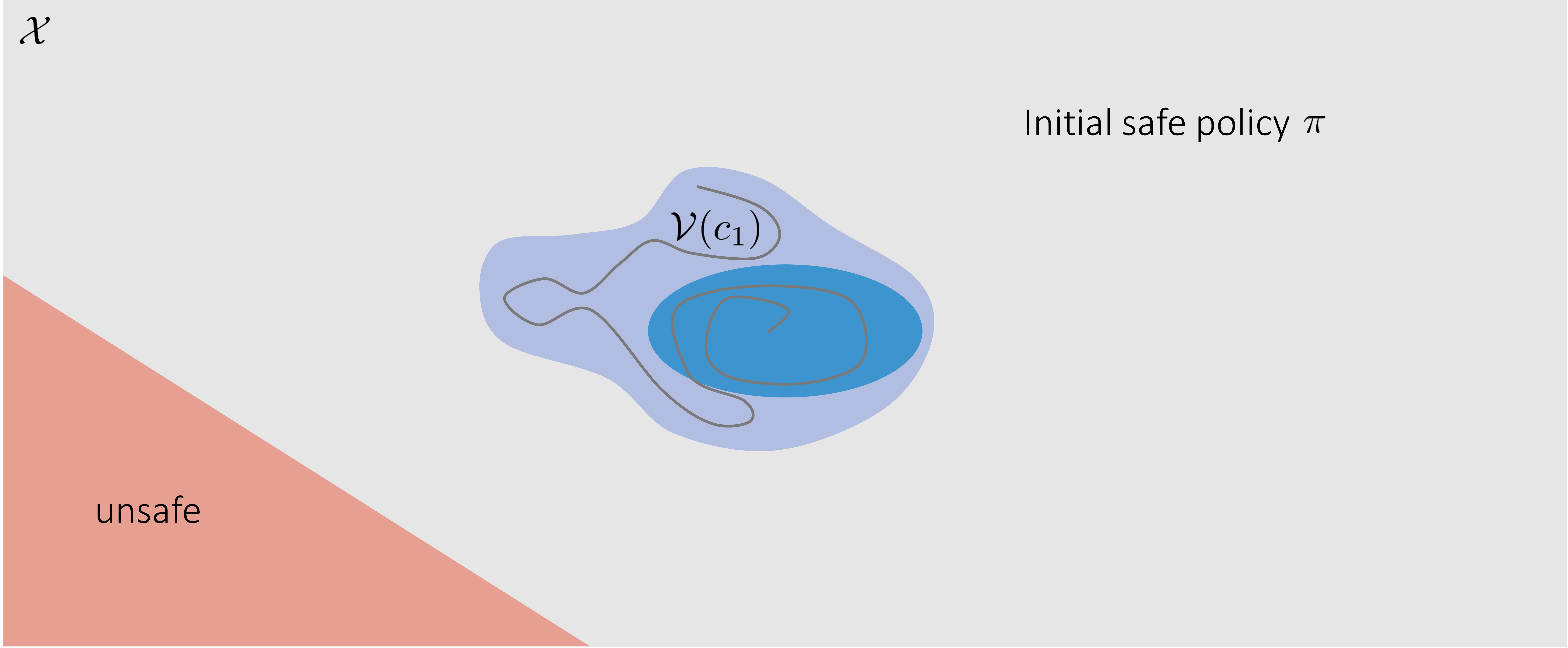
Safety definition

Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017



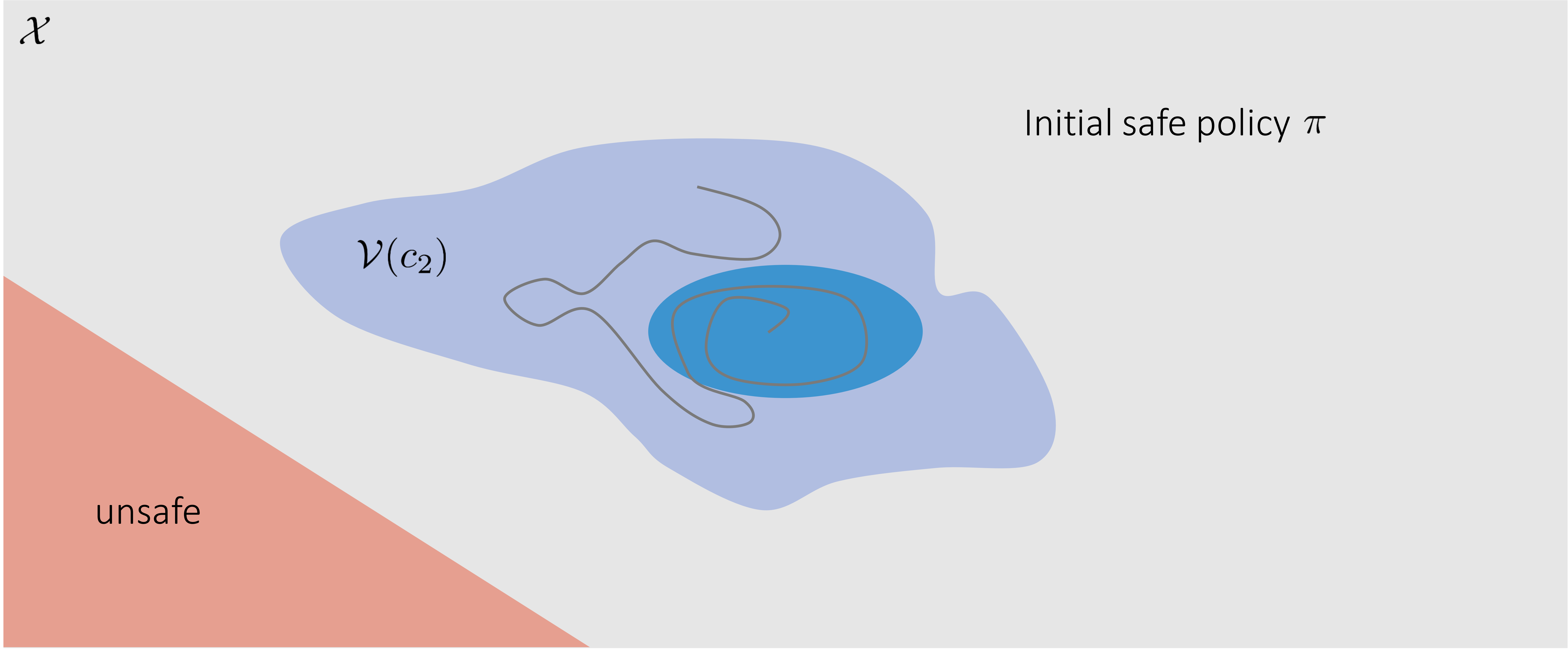
Safety definition

Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017



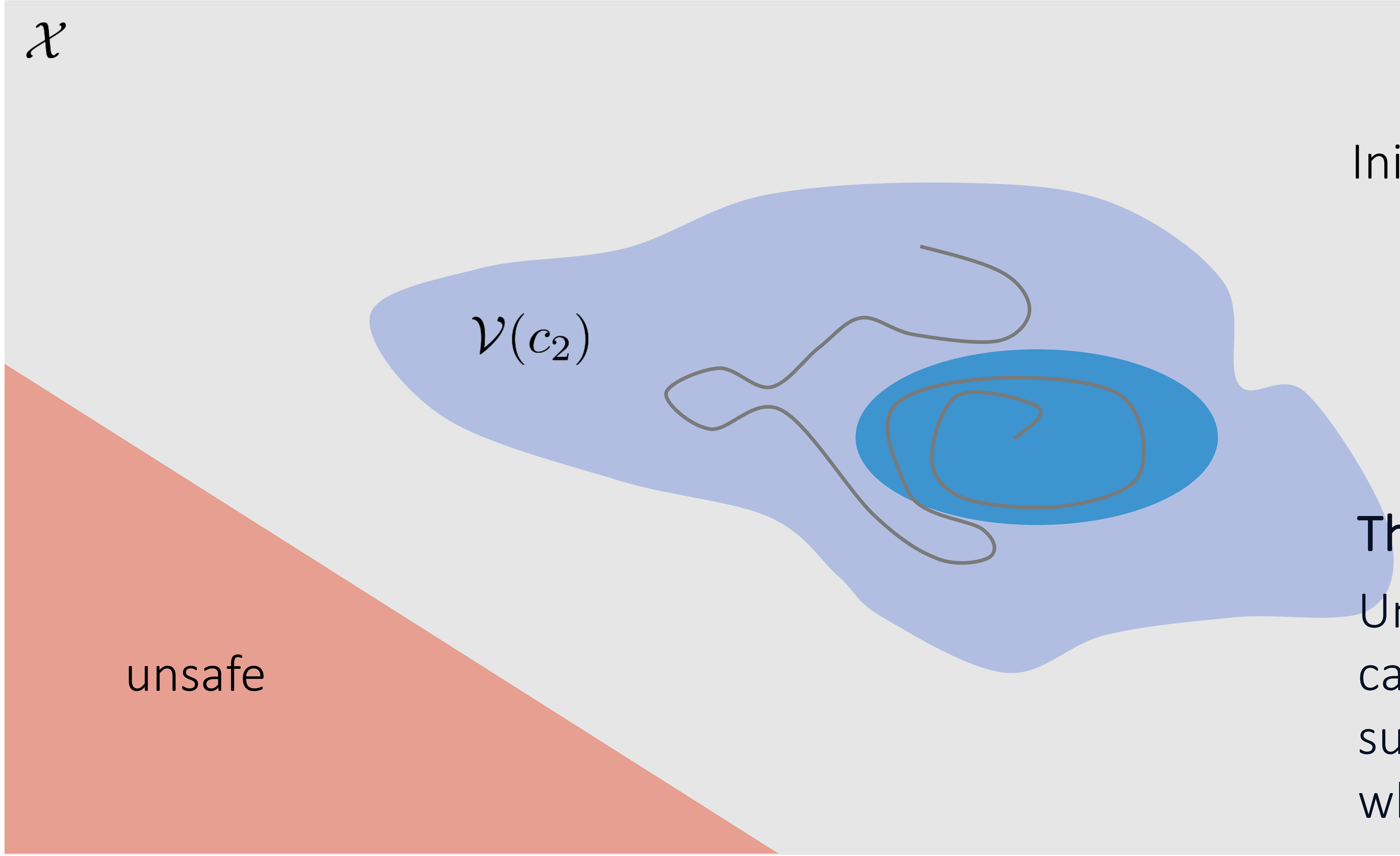
Safety definition

Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017



Safety definition

Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017

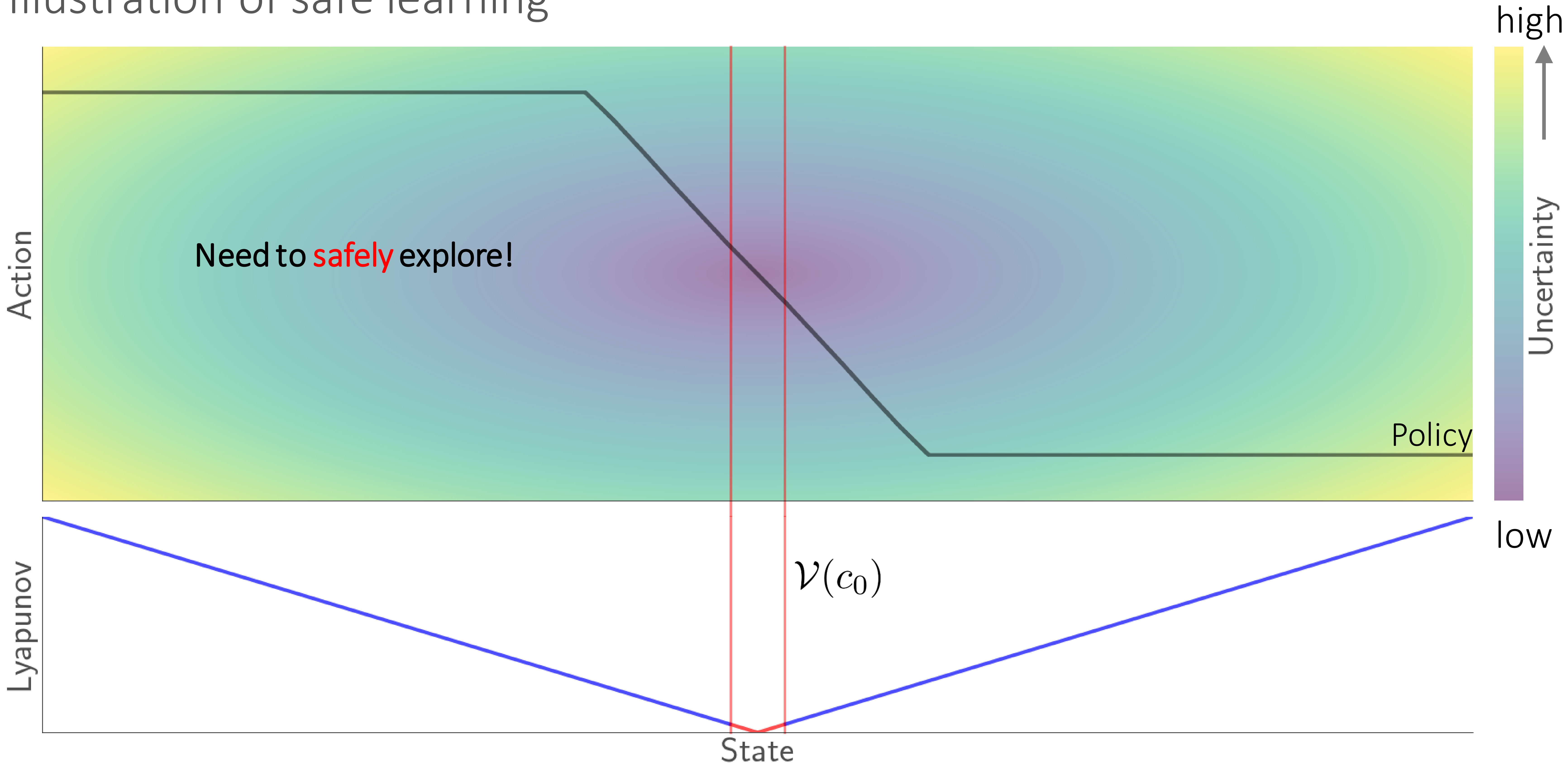


Initial safe policy π

Theorem (informally):

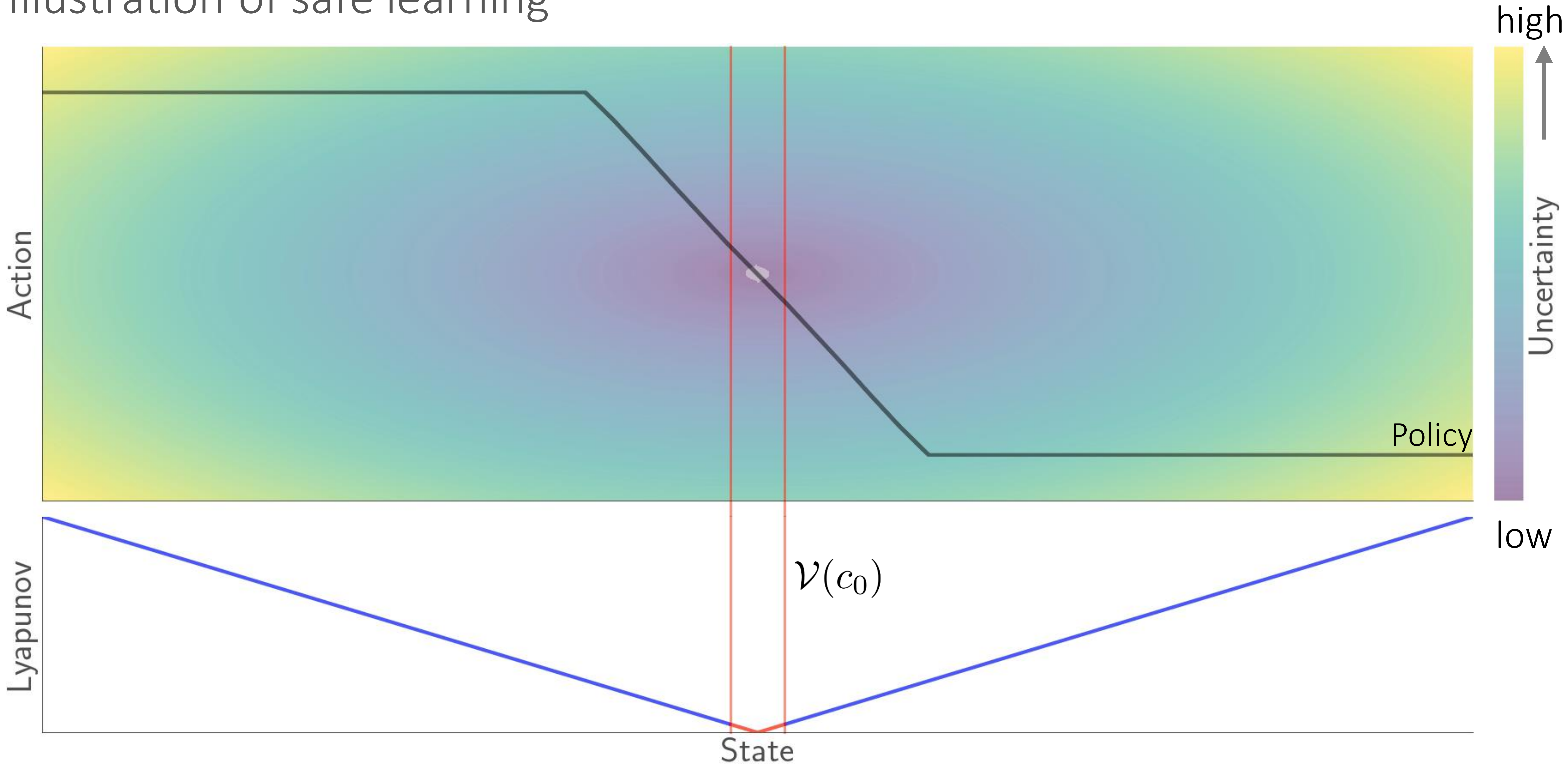
Under suitable conditions can identify (near-)maximal subset of X on which π is stable, while never leaving the safe set

Illustration of safe learning



Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017

Illustration of safe learning



Safe Model-based Reinforcement Learning with Stability Guarantees
F. Berkenkamp, M. Turchetta, A.P. Schoellig, A. Krause, NIPS, 2017

Model predictive control

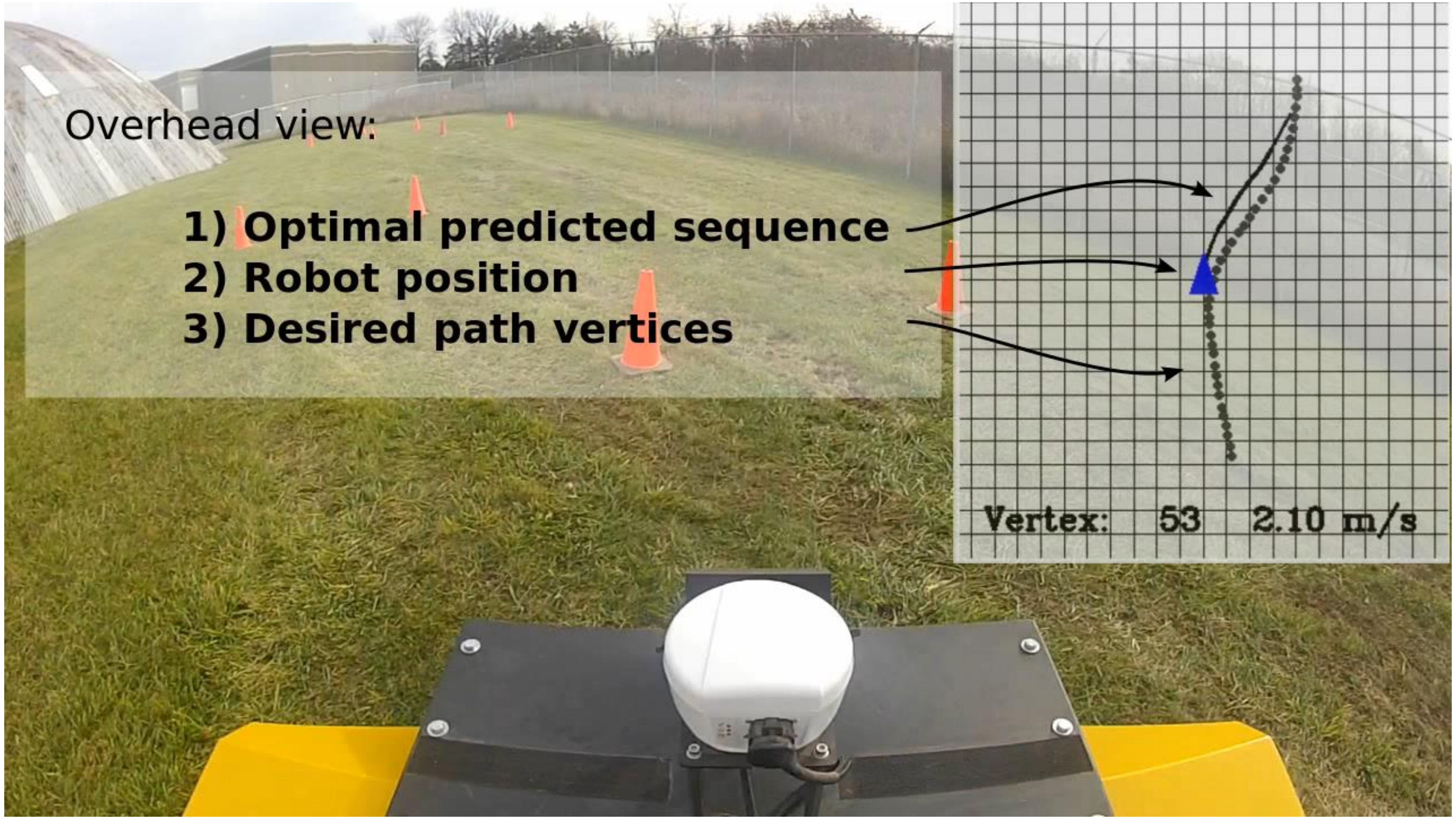
$$
\begin{aligned}
\underset{\{u_0, u_1, \dots, u_{N-1}\}}{\text{minimize}} \quad & \sum_{k=0}^{N-1} J(x_k, u_k) + J_N(x_N) && \text{mission objective} \\
\text{subject to} \quad & x_0 = \overline{x}_0 && \text{system state} \\
& x_{k+1} = f(x_k, u_k) && \text{system dynamics} \\
& x_k \in \mathcal{X}_k && \text{state constraints} \\
& u_k \in \mathcal{U}_k && \text{input constraints}
\end{aligned}
$$

Makes decisions based on predictions about the future

Includes input / state constraints

Model predictive control on a robot

https://youtu.be/3xRNmNv5Efk



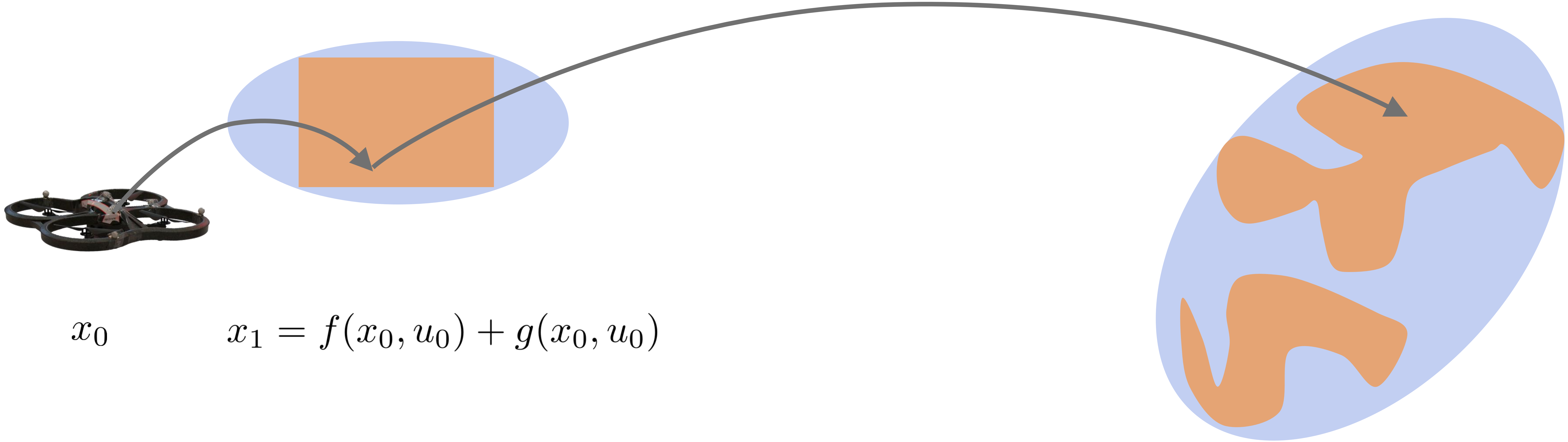
Robust constrained learning-based NMPC enabling reliable mobile robot path tracking
C.J. Ostafew, A.P. Schoellig, T.D. Barfoot, IJRR, 2016

Model predictive control

$$
\begin{aligned}
\underset{\{u_0, u_1, \dots, u_{N-1}\}}{\text{minimize}} \quad & \sum_{k=0}^{N-1} J(x_k, u_k) + J_N(x_N) && \text{mission objective} \\
\text{subject to} \quad & x_0 = \overline{x}_0 && \text{system state} \\
& x_{k+1} = f(x_k, u_k) + g(x_k, u_k) && \text{system dynamics} \\
& x_k \in \mathcal{X}_k && \text{state constraints} \\
& u_k \in \mathcal{U}_k && \text{input constraints}
\end{aligned}
$$

Problem: True dynamics $f(x, u) + g(x, u)$ are unknown!

Prediction under uncertainty



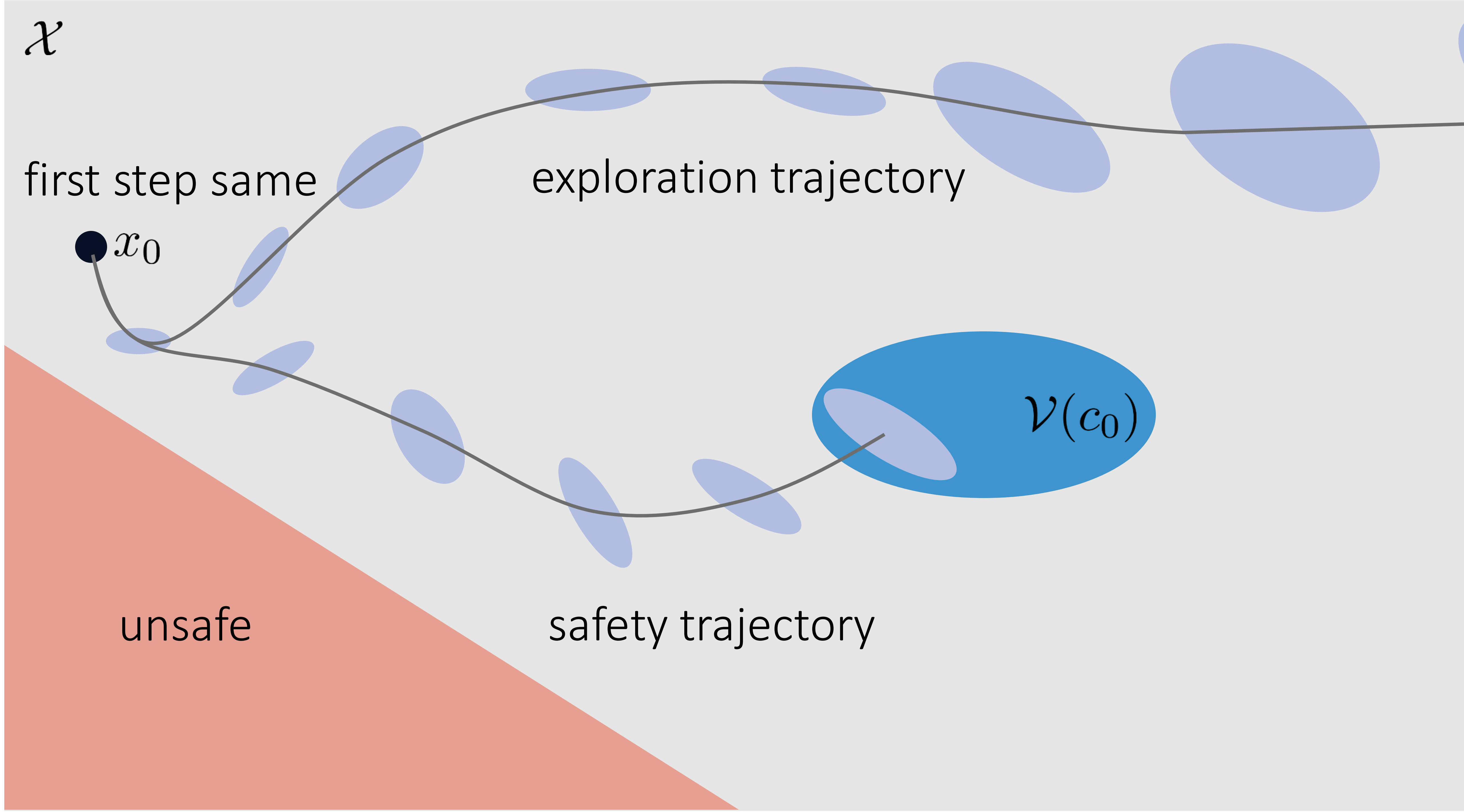
$$x_0 \qquad x_1 = f(x_0, u_0) + g(x_0, u_0)$$

$$x_2 = f(x_1, u_1) + g(x_1, u_1)$$

Outer approximation contains true dynamics for all time steps with probability at least $1 - \delta$

Learning-based Model Predictive Control for Safe Exploration
T. Koller, F. Berkenkamp, M. Turchetta, A. Krause, CDC, 2018

Safe model-based learning framework



Theorem (informally):

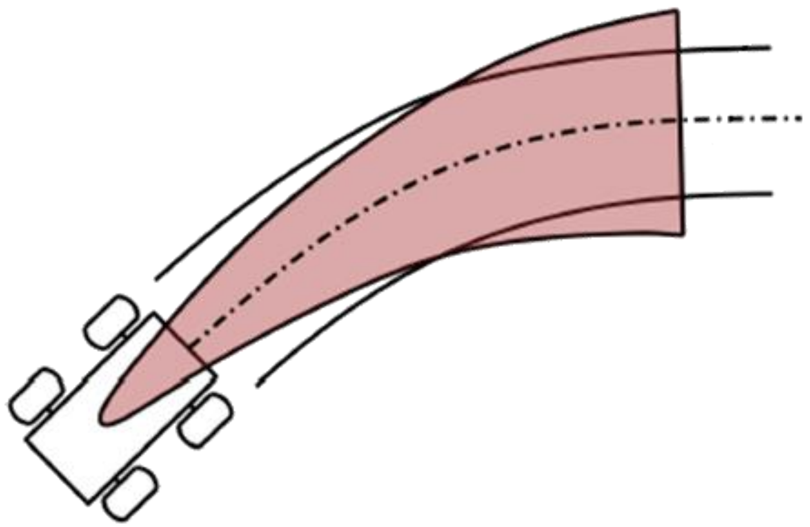
Under suitable conditions can always guarantee that we are able to return to the safe set

Exploration via expected performance

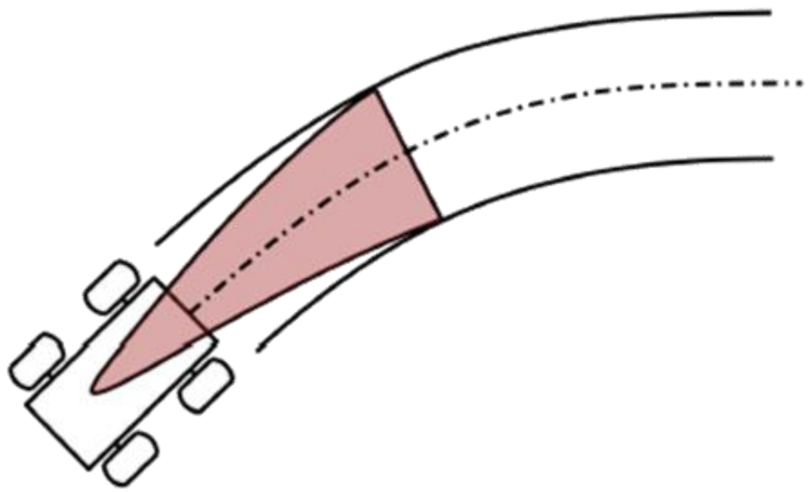
We design our cost functions to be helpful for optimization

Exploration objective: $$\underset{\{u_0, u_1, \ldots, u_{N-1}\}}{\text{minimize}} \; \mathrm{E}\left[\sum_{k=0}^{N-1} J(x_k, u_k) + J_N(x_N)\right]$$

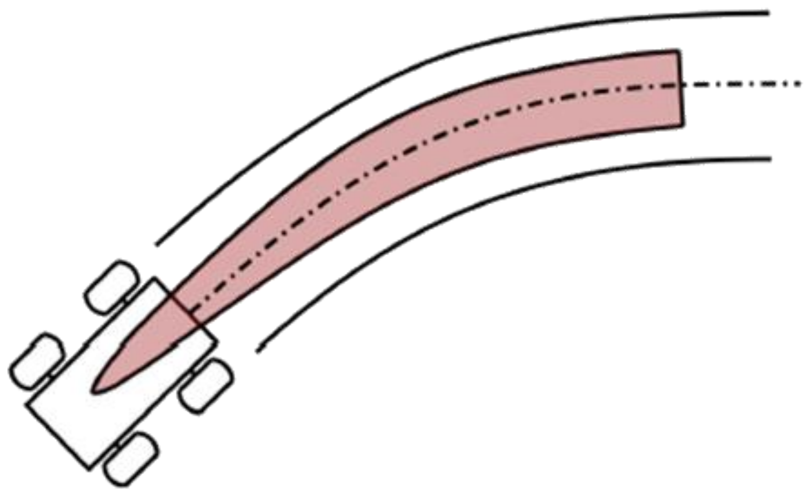
subject to safety constraints



Driving too fast



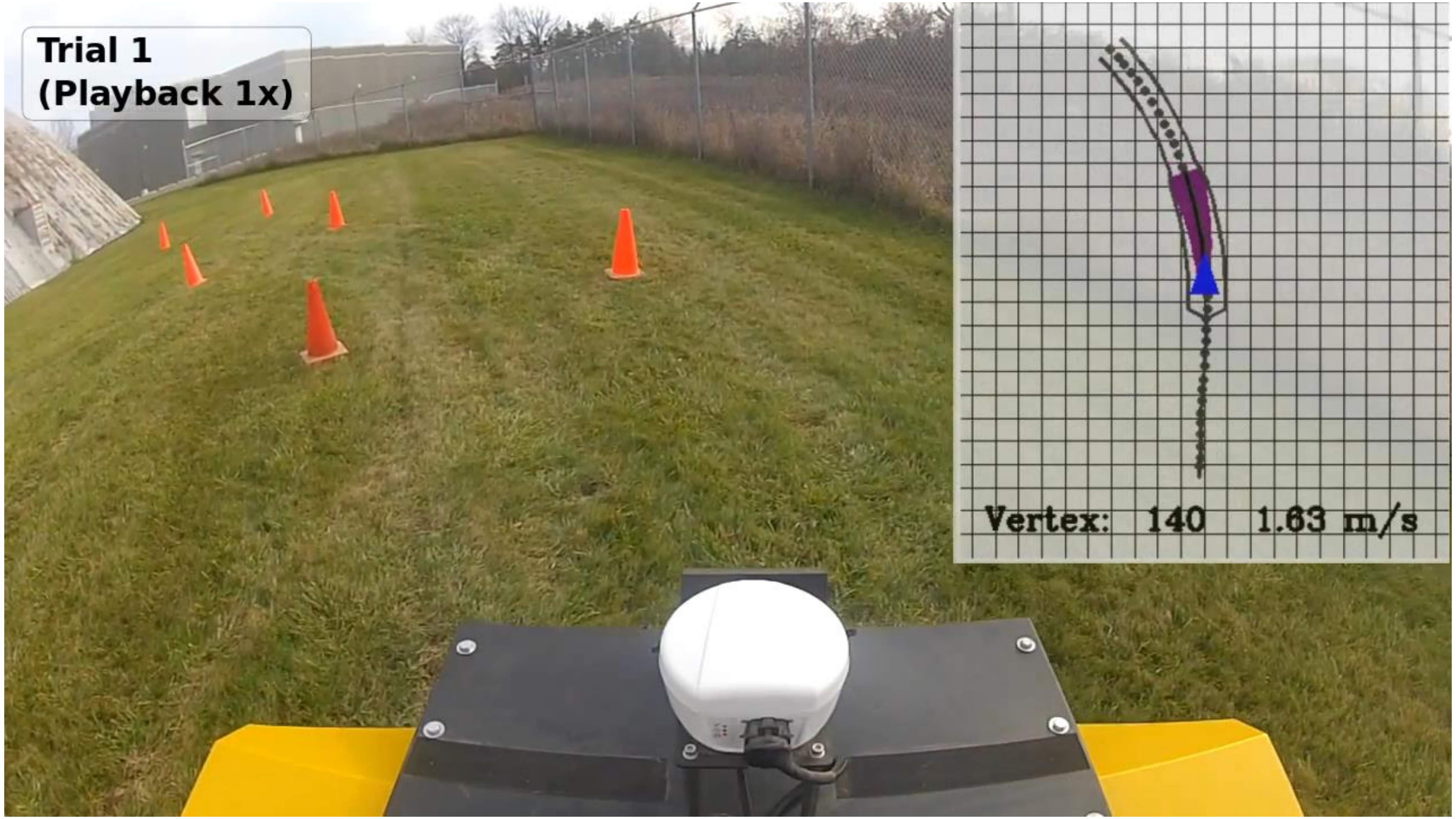
Slow down for safety



Faster driving after learning

Example

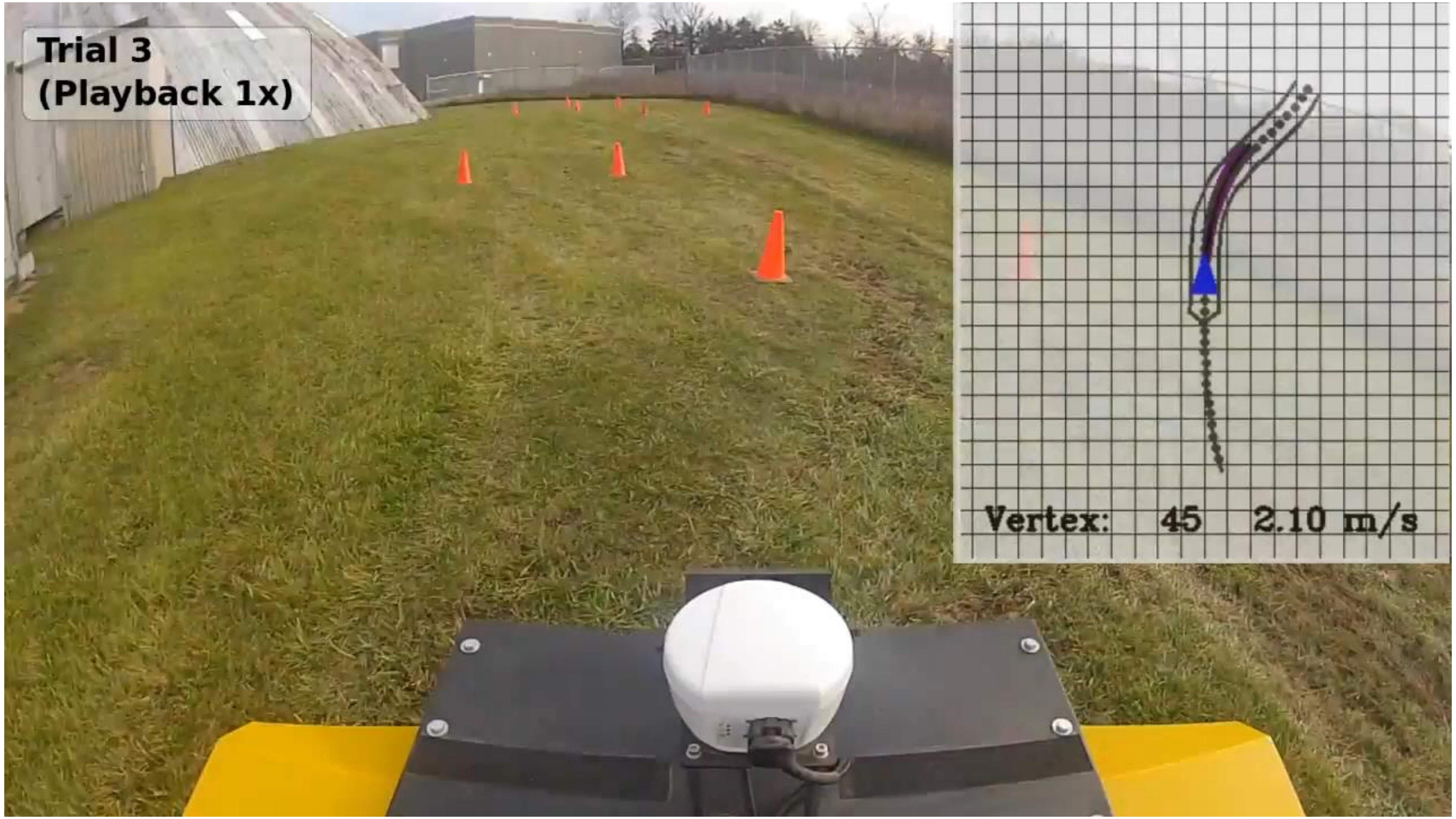
https://youtu.be/3xRNmNv5Efk



Robust constrained learning-based NMPC enabling reliable mobile robot path tracking
C.J. Ostafew, A.P. Schoellig, T.D. Barfoot, IJRR, 2016

Example

https://youtu.be/3xRNmNv5Efk



Robust constrained learning-based NMPC enabling reliable mobile robot path tracking

C.J. Ostafew, A.P. Schoellig, T.D. Barfoot, IJRR, 2016

Summary

Understand model and learning dynamics

RKHS / Gaussian processes

reliable confidence intervals

Define safety, analyze a model for safety

Lyapunov stability

stability of learned models

Algorithm to safely acquire data and optimize task

Model predictive control

Uncertainty propagation, safe active learning

Safe Model-based Reinforcement Learning

https://berkenkamp.me www.las.inf.ethz.ch www.dynsyslab.org