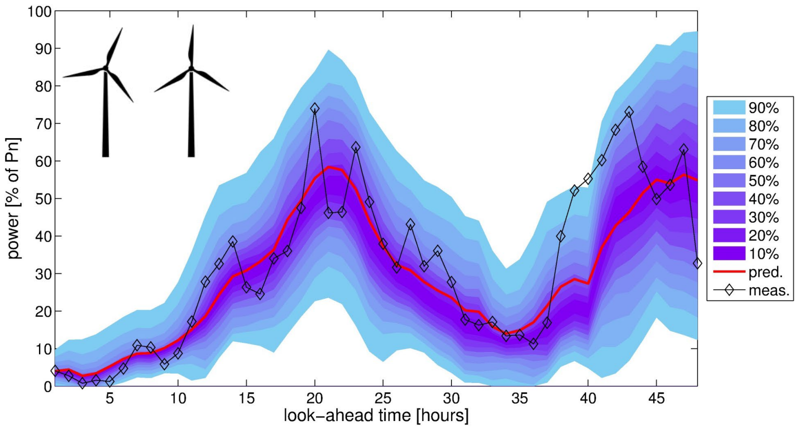
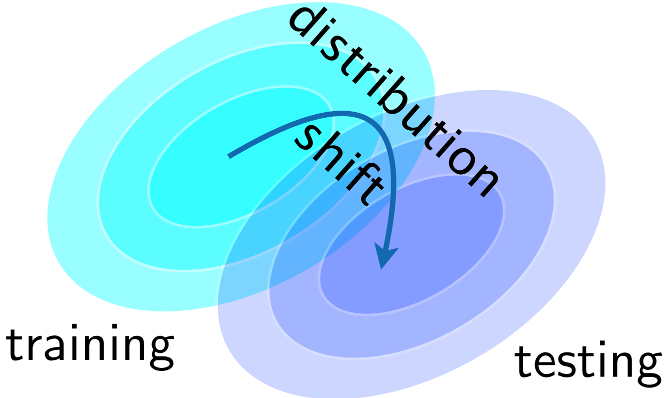
Automatic Control Laboratory

Capture, Propagate, & Control Distributional Uncertainty

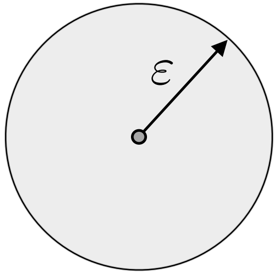
Florian Dörfler & Liviu Aolaritei
ETH Zürich

The Two Canonical Uncertainty Descriptions





Robust

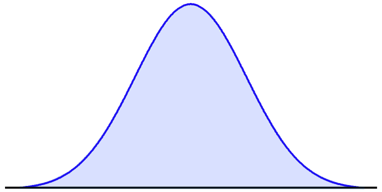


Norm-bounded uncertainty

✔ easy to propagate
tractable

✘ not very expressive
conservative

Stochastic



Fixed distribution

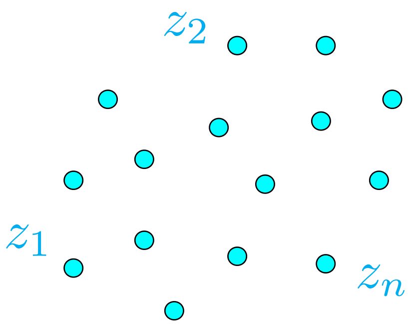
✔ more expressive
less conservative

✘ hard to propagate
less tractable

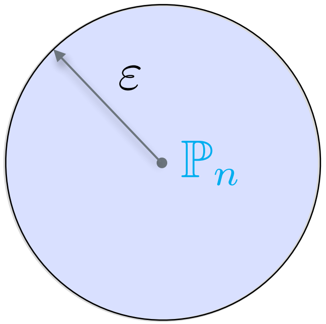
✘ additionally: both require prior knowledge & suffer from distribution shift … can we do better ?

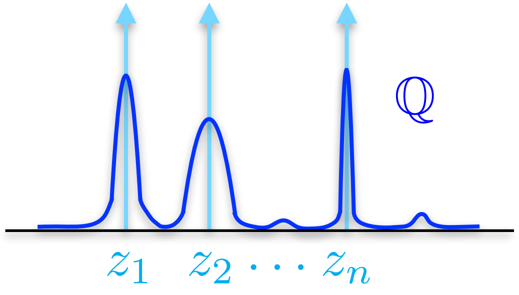
ETH zürich

Distributional Uncertainty via Optimal Transport



$\rightarrow \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{z_i} \quad \rightarrow \quad \mathbb{B}_\varepsilon^c(\mathbb{P}_n) :=$

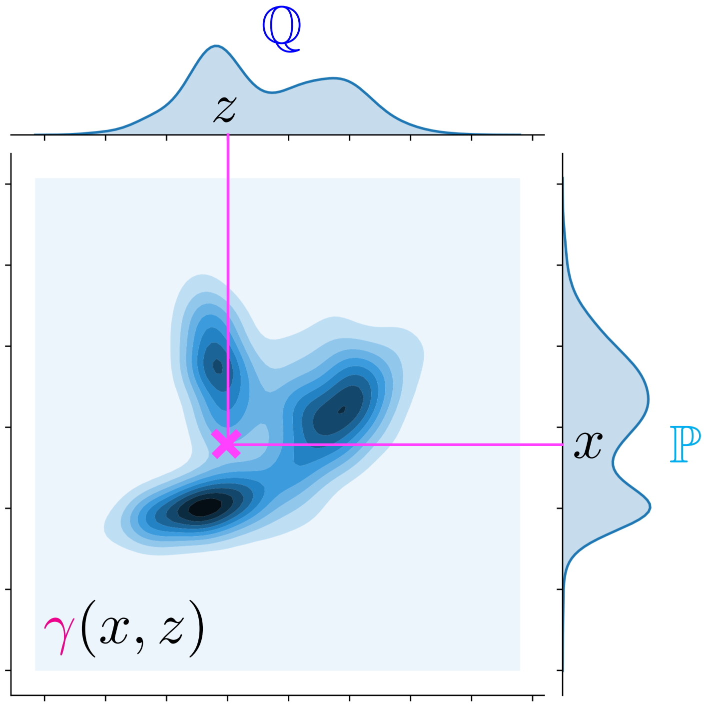




OT distance $W_c(\mathbb{Q}, \mathbb{P}) = \inf_{\gamma \in \Gamma(\mathbb{Q}, \mathbb{P})} \mathbb{E}_{(x,z)\sim\gamma}\left[c(x - z)\right]$

$\Gamma(\mathbb{Q}, \mathbb{P})$: set of joint distributions with marginals $\mathbb{Q}$ and $\mathbb{P}$

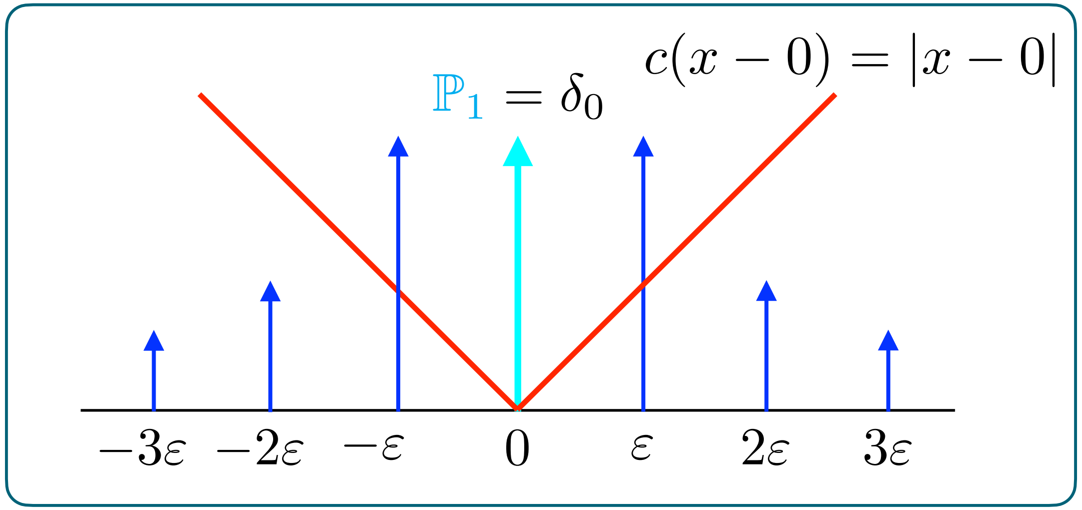
$c(x - z)$: transportation cost

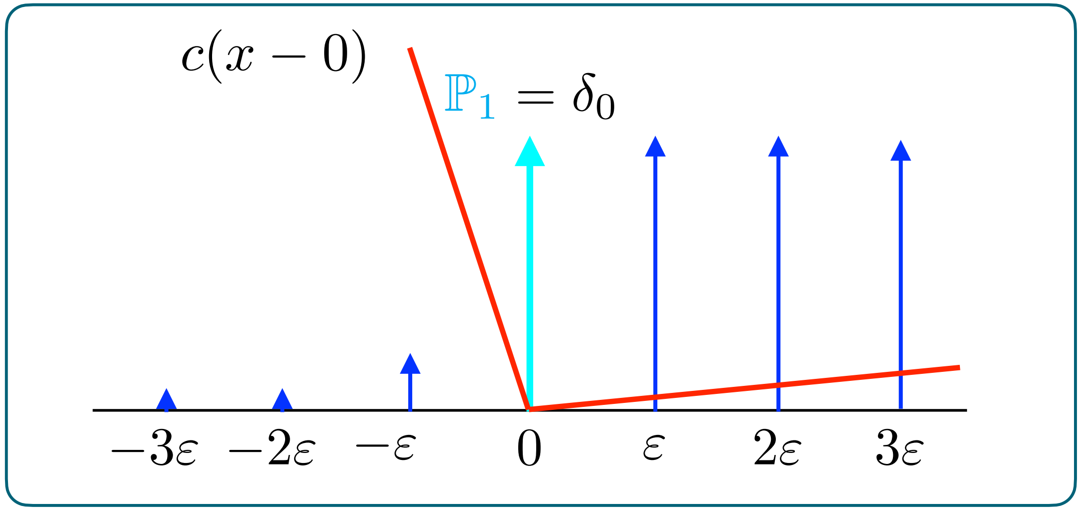


OT distance $W_c(\mathbb{Q}, \mathbb{P})$ can capture continuous (physics) & discrete (data) distributions $\rightarrow$ data-driven parameterization ✔

Geometric Properties of OT Ambiguity Sets

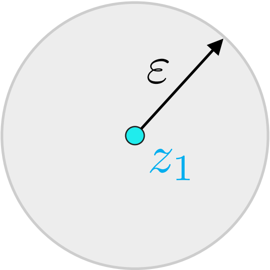
Example. $\mathbb{P}_1 = \delta_0$ → $W_c(\mathbb{Q}, \delta_0) = \mathbb{E}_{x\sim\mathbb{Q}}\left[c(x-0)\right]$ → $\mathbb{B}^c_\varepsilon(\delta_0) = \{\mathbb{Q} : \mathbb{E}_{x\sim\mathbb{Q}}\left[c(x-0)\right] \leq \varepsilon\}$





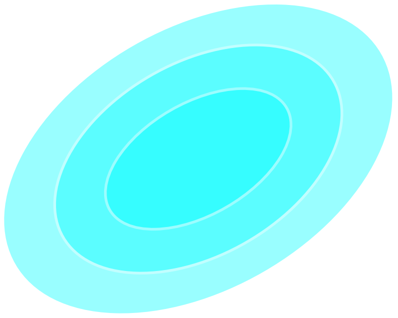
expressive: robustness in **space** & **likelihood**

Generalizes deterministic robustness
- bounded support → reduces conservativeness
- unbounded support → captures **black swans**

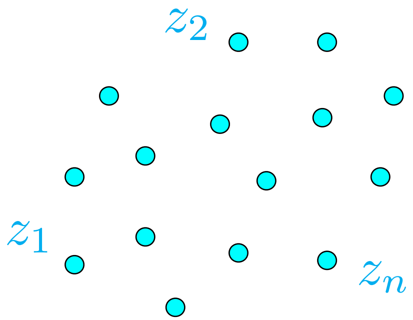


Capturing of OT Ambiguity Sets from Data

true distribution $\mathbb{P}_\star$



→



→

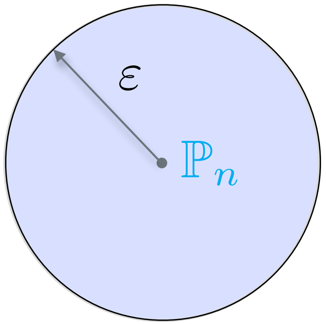
sample distribution $\mathbb{P}_n$

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{z_i}$$

→

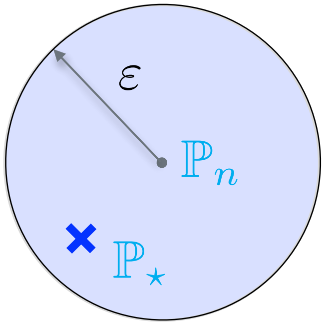
ambiguity set $\mathbb{B}^c_\varepsilon(\mathbb{P}_n)$

$$\mathbb{B}^c_\varepsilon(\mathbb{P}_n) :=$$



Statistical guarantees:

1. $\varepsilon = O\left(n^{-1/max\{2,d\}}\right)$ → $\mathbb{P}_\star \in \mathbb{B}^c_\varepsilon(\mathbb{P}_n)$



true distribution inside the OT ambiguity set ✔

2. $\varepsilon = O\left(n^{-1/2}\right)$ → $\mathbb{E}_{z\sim\mathbb{P}_\star}[\ell(z)] \le \sup_{\mathbb{Q}\in\mathbb{B}^c_\varepsilon(\mathbb{P}_n)} \mathbb{E}_{z\sim\mathbb{Q}}[\ell(z)] + O(n^{-1})$

true risk upper bounded by the robust risk ✔

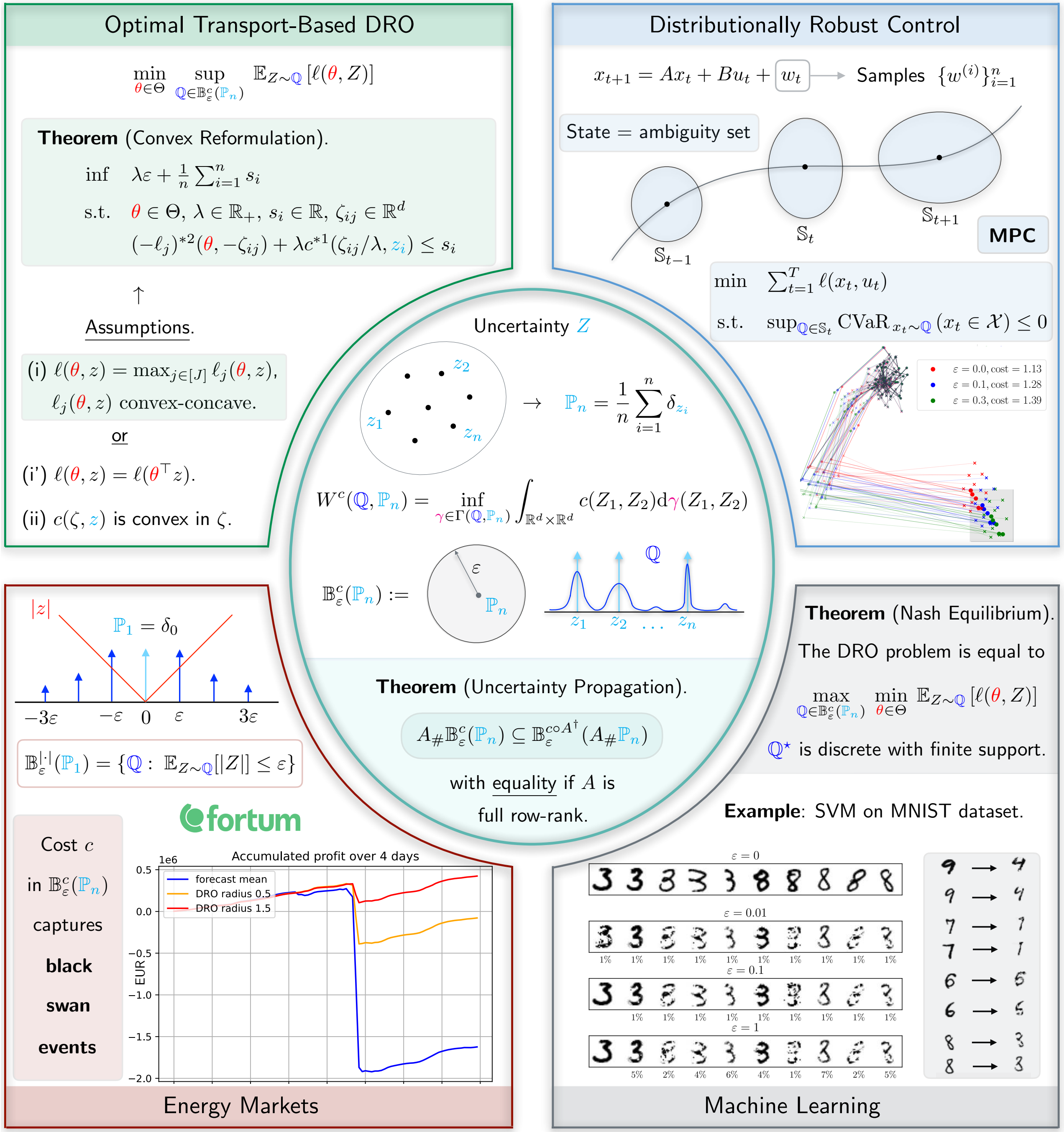
Outline

0. Capturing

1. Propagation

2. Computation

3. Applications:
   → today: control



Liviu Aolaritei

+ Nicolas Lanzetti, Marta Fochesato, Soroosh Shafiee, Antonio Terpin, Daniel Kuhn, John Lygeros, ...

ETH zürich

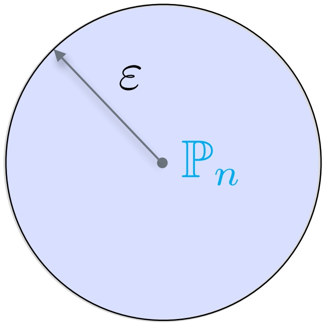
Propagation

Propagation of OT Ambiguity Sets

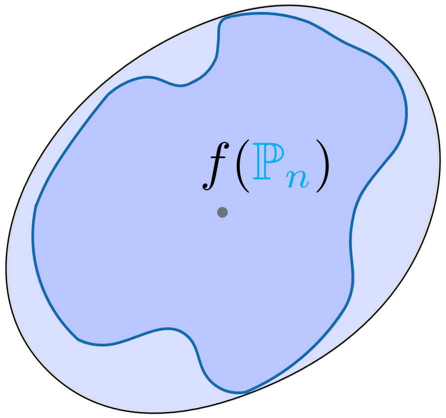
Pushforward operation: $Z \sim \mathbb{P} \;\rightarrow\; f(Z) \sim f_{\#}\mathbb{P}$

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{z_i} \;\rightarrow\; f_{\#}\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{f(z_i)}$$

$\mathbb{B}_{\varepsilon}^{c}(\mathbb{P}_n) :=$



$\rightarrow \quad f_{\#}\mathbb{B}_{\varepsilon}^{c}(\mathbb{P}_n) = ?$



Is this an OT ambiguity set?

If not, can we capture it tightly with another OT ambiguity set?

Example (deterministic).

$$\mathcal{B}_{\varepsilon}^{c}(z_0) := \{z \in \mathbb{R}^d : \; c(z - z_0) \leq \varepsilon\}$$

$f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ bijective

$$f(\mathcal{B}_{\varepsilon}^{c}(z_0)) = \{f(z) \in \mathbb{R}^d : \; c(z - z_0) \leq \varepsilon\}]$$

or

$$f(\mathcal{B}_{\varepsilon}^{c}(z_0)) = \{\tilde{z} \in \mathbb{R}^d : \; c(f^{-1}(\tilde{z}) - \underbrace{f^{-1}(f(z_0))}_{z_0}) \leq \varepsilon\} = \mathcal{B}_{\varepsilon}^{c \circ f^{-1}}(f(z_0))$$

Pushforward of OT Ambiguity Sets

Theorem. ($\mathbb{P}$ arbitrary)

1. If f is bijective, then

$$f_\# \mathbb{B}^c_\varepsilon(\mathbb{P}) = \mathbb{B}^{c \circ f^{-1}}_\varepsilon(f_\# \mathbb{P})$$

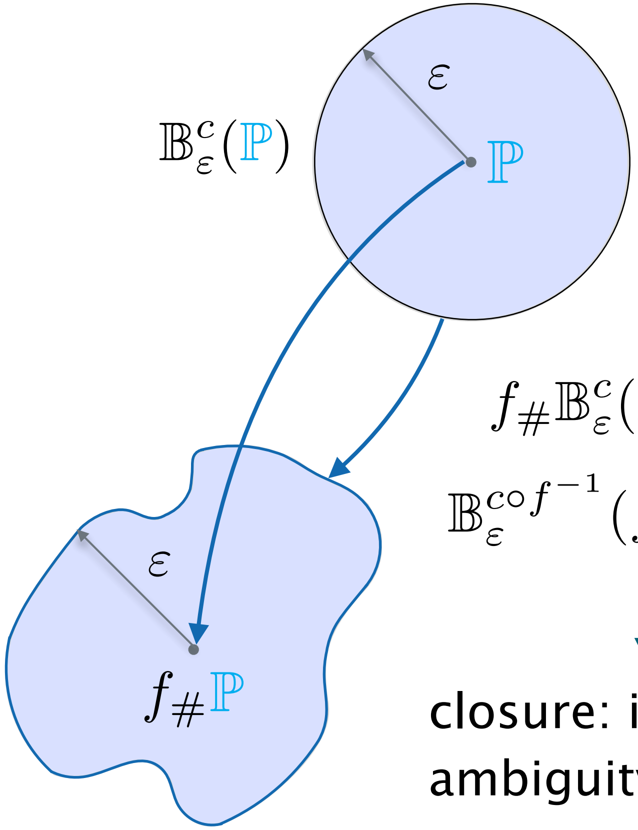
2. If f is injective/surjective, then

$$f_\# \mathbb{B}^c_\varepsilon(\mathbb{P}) \subseteq \mathbb{B}^{c \circ f^{-1}}_\varepsilon(f_\# \mathbb{P})$$

3. If $f(z) = Az$ with a general matrix A, then

$$A_\# \mathbb{B}^c_\varepsilon(\mathbb{P}) \subseteq \mathbb{B}^{c \circ A^\dagger}_\varepsilon(A_\# \mathbb{P})$$

with equality $\equiv$ if A is full row-rank



$f_\# \mathbb{B}^c_\varepsilon(\mathbb{P}) =$

$\mathbb{B}^{c \circ f^{-1}}_\varepsilon(f_\# \mathbb{P})$

↓

closure: is an OT ambiguity set ! ✔

(in a different geometry)

↓

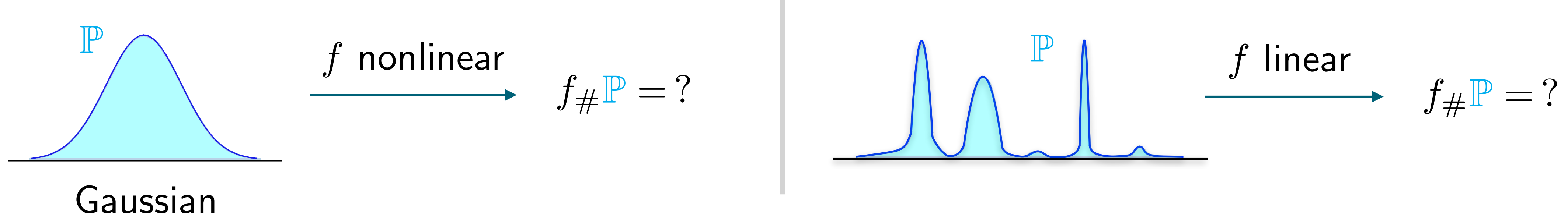
✔ easy to propagate

(as in the deterministic case)

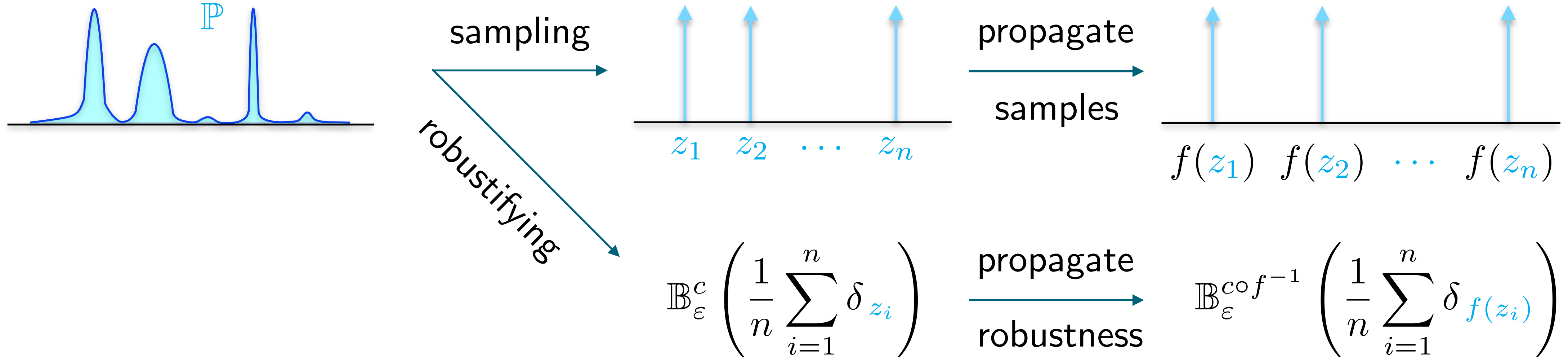
This Propagation Result Enables ...

Q: connection to … particle filter ?

1. **Robust Uncertainty Quantification**



Monte Carlo + Distributional Robustness



This Propagation Result Enables ...

2. **Distributionally Robust System Theory**

$x_0 \rightarrow$, $w_k \rightarrow$ $\boxed{x_{k+1} = Ax_k + w_k}$ $\rightarrow x_{k+1}$ $\rightarrow$ $x_{k+1} = Ax_k + w_k = A^{k+1}x_0 + \underbrace{\begin{bmatrix} A^k & \dots & A & I \end{bmatrix}}_{\mathcal{D}_k \text{ (full row-rank)}} \begin{bmatrix} w_0 \\ \dots \\ w_k \end{bmatrix}$

Capture uncertainty:

n noise trajectories

$$\widehat{\mathbf{w}}^{(i)} := \left\{ \left[\widehat{w}_0^{(i)}, \dots, \widehat{w}_k^{(i)} \right] \right\}_{i=1}^{n}$$

$$\rightarrow \frac{1}{n}\sum_{i=1}^{n} \delta_{\widehat{\mathbf{w}}^{(i)}} \rightarrow \mathbb{B}_\varepsilon^c\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{\widehat{\mathbf{w}}^{(i)}}\right)$$

Propagation: The distributional uncertainty of the state x_{k+1} is **exactly** captured by

$$\mathbb{B}_\varepsilon^{c\circ\mathcal{D}_k^\dagger}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{A^{k+1}x_0 + \mathcal{D}_k\widehat{\mathbf{w}}^{(i)}}\right)$$

Geometry: Decomposition of Dynamics: Nominal z_k + Error e_k

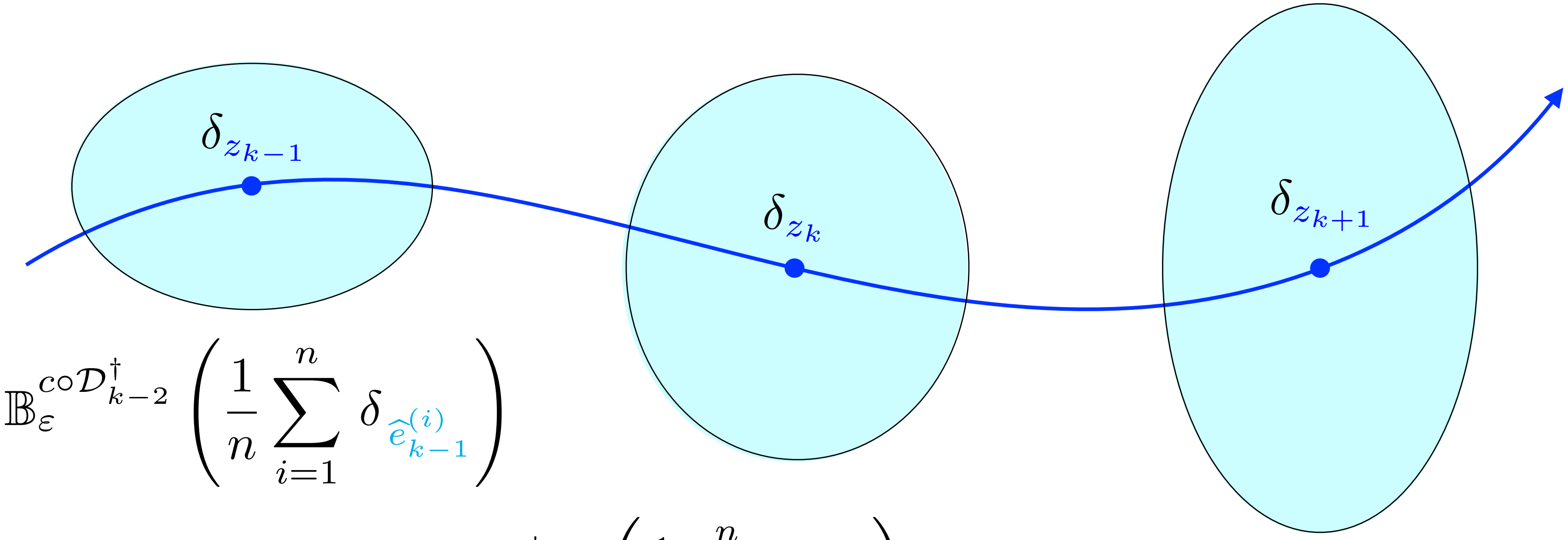
nominal: $z_{k+1} = A z_k$

error: $e_{k+1} = \mathcal{D}_k \begin{bmatrix} w_0 \\ \cdots \\ w_{k-1} \\ w_k \end{bmatrix}$

→ OT ambiguity set of $x_{k+1} = z_{k+1} + e_{k+1}$: $\mathbb{B}_\varepsilon^{c \circ \mathcal{D}_k^\dagger}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{\underbrace{A^{k+1}x_0}_{z_{k+1}} + \underbrace{\mathcal{D}_k \widehat{\mathbf{w}}^{(i)}}_{\widehat{e}_{k+1}^{(i)}}}\right)$

↓

samples of e_{k+1}



$\mathbb{B}_\varepsilon^{c \circ \mathcal{D}_{k-2}^\dagger}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{\widehat{e}_{k-1}^{(i)}}\right)$

$\mathbb{B}_\varepsilon^{c \circ \mathcal{D}_{k-1}^\dagger}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{\widehat{e}_{k}^{(i)}}\right)$

$\mathbb{B}_\varepsilon^{c \circ \mathcal{D}_{k}^\dagger}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{\widehat{e}_{k+1}^{(i)}}\right)$

Q: distributional stability or contraction of ambiguity sets ?

This Propagation Result Enables ...

$$\mathcal{C}_k := \left[(A+BK)^k B \quad \dots \quad (A+BK)B \quad B\right]$$

$$\mathcal{D}_k := \left[(A+BK)^k \quad \dots \quad (A+BK) \quad I\right]$$ full rank

3. **Distributionally Robust Control**

$$x_{k+1} = A x_k + B u_k + w_k = (A+BK)^{k+1} x_0 + \mathcal{C}_k \underbrace{\begin{bmatrix} v_0 \\ \dots \\ v_{k-1} \\ v_k \end{bmatrix}}_{\mathbf{v}_{[0:k]}} + \mathcal{D}_k \begin{bmatrix} w_0 \\ \dots \\ w_{k-1} \\ w_k \end{bmatrix}$$

$$u_k \to K x_k + v_k$$

Capture noise uncertainty

n noise trajectories

$$\widehat{\mathbf{w}}^{(i)}_{[0:k]} := \left\{ \left[\widehat{w}^{(i)}_0, \dots, \widehat{w}^{(i)}_k \right] \right\}_{i=1}^{n}$$

$$\to \frac{1}{n} \sum_{i=1}^{n} \delta_{\widehat{\mathbf{w}}^{(i)}_{[0:k]}} \to \mathbb{B}^c_\varepsilon \left(\frac{1}{n} \sum_{i=1}^{n} \delta_{\widehat{\mathbf{w}}^{(i)}_{[0:k]}} \right)$$

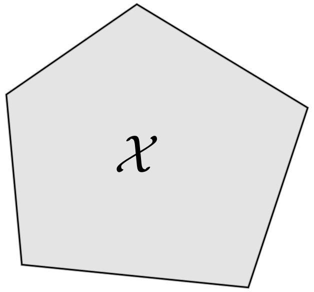
Proposition. The distributional uncertainty of the state x_{k+1} is **exactly** captured by

$$\mathbb{S}_{k+1} := \mathbb{B}^{c \circ \mathcal{D}_k^\dagger}_\varepsilon \left(\frac{1}{n} \sum_{i=1}^{n} \delta_{(A+BK)^{k+1} x_0 + \mathcal{C}_k \mathbf{v}_{[0:k]} + \mathcal{D}_k \widehat{\mathbf{w}}^{(i)}_{[0:k]}} \right)$$

Q: optimal or robust control ?

Computation

Example: Distributionally Robust CVaR Constraints



$$\mathcal{X} := \left\{ x \in \mathbb{R}^d : \max_j a_j^\top x + b_j \leq 0 \right\}$$

→ **Constraint**: $\sup_{\mathbb{Q} \in \mathbb{S}_k} \text{CVaR}_\delta^{\mathbb{Q}} \left(\max_j a_j^\top x_k + b_j \right) \leq 0$

↓

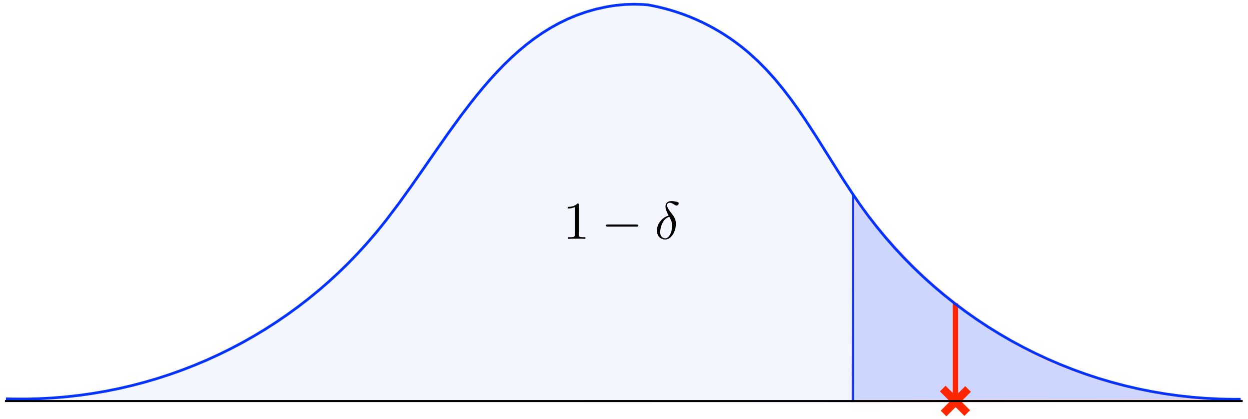
$\Pr\left(x_k \in \mathcal{X}\right) \geq 1 - \delta$ & benign tail risk

OT ambiguity set $\mathbb{S}_k$ captures the distributional uncertainty of x_k

CVaR$_\delta^{\mathbb{Q}}$ $\left(\max_j a_j^\top x_k + b_j\right)$ is of the form:

$$\inf_{\theta \in \Theta} \mathbb{E}_{z \sim \mathbb{Q}} \left[\ell(\theta, z)\right]$$

with $\ell(\theta, z) =$ max of affine functions



(expectation of the δ-tail) CVaR$_\delta$

Distributionally Robust Optimization (DRO)

Strong dual formulation

$$\inf_{\lambda \geq 0} \; \lambda \varepsilon + \frac{1}{n} \sum_{i=1}^{n} \sup_{\zeta \in \mathbb{R}^d} \; \ell(\theta, \zeta) - \lambda c(\zeta - z_i)$$

$$\inf_{\theta \in \Theta} \; \sup_{\mathbb{Q} \in \mathbb{B}_\varepsilon^c(\mathbb{P}_n)} \; \mathbb{E}_{z \sim \mathbb{Q}} [\ell(\theta, z)]$$

infinite-dimensional

$$\ell(\theta, z) = \max_j \; \underbrace{\ell_j(\theta, z)}$$

convex-concave

Theorem. Let c be convex. Then the DRO problem has the same infimum as

$$\begin{aligned} \inf \quad & \lambda \varepsilon + \tfrac{1}{n} \textstyle\sum_{i=1}^{n} s_i \\ \text{s.t.} \quad & \theta \in \Theta, \; \lambda \in \mathbb{R}_+, \; s_i \in \mathbb{R}, \; \zeta_{ij} \in \mathbb{R}^d, & \forall i, j \\ & (-\ell_j)^{*2}(\theta, -\zeta_{ij}) + \zeta_{ij}^\top z_i + \lambda c^*(\zeta_{ij}/\lambda) \leq s_i, & \forall i, j \end{aligned}$$

✔ efficiently computable s.t. reasonable assumptions (very active area)

Control

Decomposition of Dynamics

LTI Stochastic System: $x_{k+1} = A x_k + B u_k + w_k = (A+BK)^{k+1} x_0 + \mathcal{C}_k \begin{bmatrix} v_0 \\ \cdots \\ v_{k-1} \\ v_k \end{bmatrix} + \mathcal{D}_k \begin{bmatrix} w_0 \\ \cdots \\ w_{k-1} \\ w_k \end{bmatrix}$

$u_k \rightarrow K x_k + v_k$

$\mathcal{D}_k \rightarrow$ full row-rank

Tube formulation: $x_k = z_k + e_k$

$z_k =$ nominal state, $e_k =$ error state

Nominal Dynamics: $z_{k+1} = (A+BK) z_k + B v_k$

$z_0 = x_0$

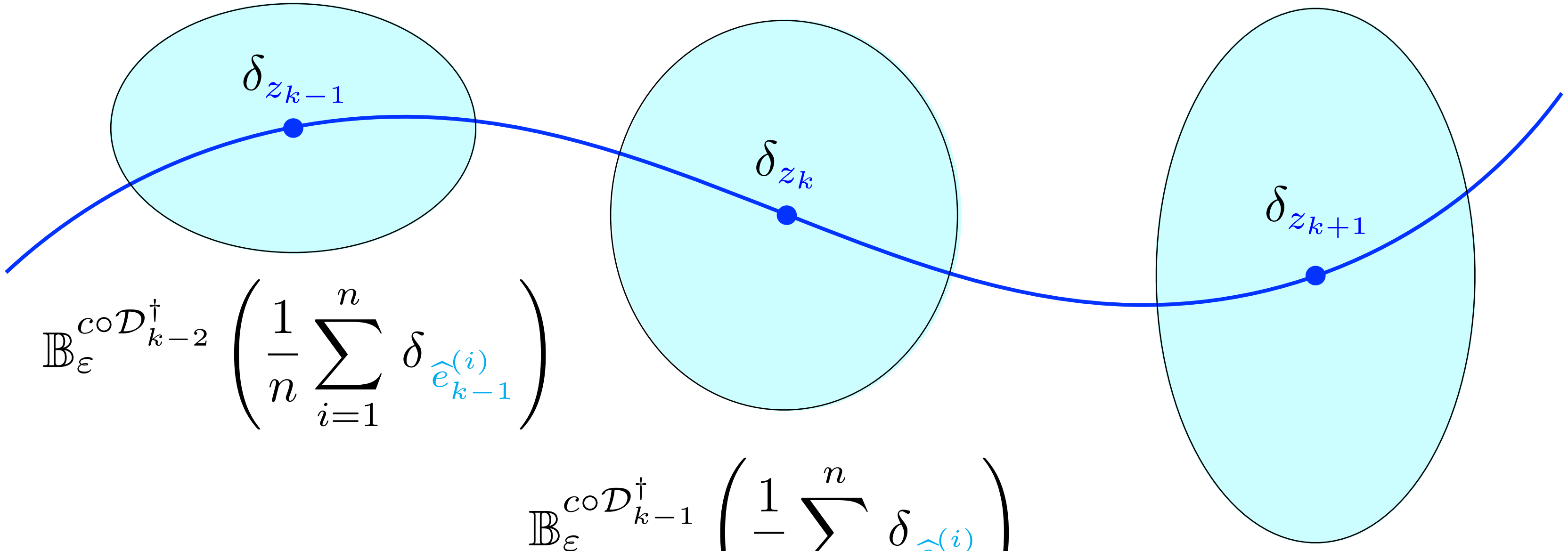
Error Dynamics: $e_{k+1} = \mathcal{D}_k \begin{bmatrix} w_0 \\ \cdots \\ w_{k-1} \\ w_k \end{bmatrix}$

$e_0 = 0$

Evolution of Decomposed of Dynamics: Nominal z_k + Error e_k

OT ambiguity set of x_{k+1} is $\mathbb{S}_{k+1} := \mathbb{B}_{\varepsilon}^{c\circ\mathcal{D}_k^{\dagger}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\underbrace{(A+BK)^{k+1}x_0+\mathcal{C}_k\mathbf{v}_{[0:k]}}_{z_{k+1}}+\underbrace{\mathcal{D}_k\widehat{\mathbf{w}}_{[0:k]}^{(i)}}_{\widehat{e}_{k+1}^{(i)}}}\right)$

$\widehat{e}_{k+1}^{(i)}$ → samples of e_{k+1}



$$\mathbb{B}_{\varepsilon}^{c\circ\mathcal{D}_{k-2}^{\dagger}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{e}_{k-1}^{(i)}}\right)$$

$$\mathbb{B}_{\varepsilon}^{c\circ\mathcal{D}_{k-1}^{\dagger}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{e}_{k}^{(i)}}\right)$$

$$\mathbb{B}_{\varepsilon}^{c\circ\mathcal{D}_{k}^{\dagger}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{e}_{k+1}^{(i)}}\right)$$

Nominal dynamics

$+$

Error OT ambiguity set

Wasserstein Tube MPC

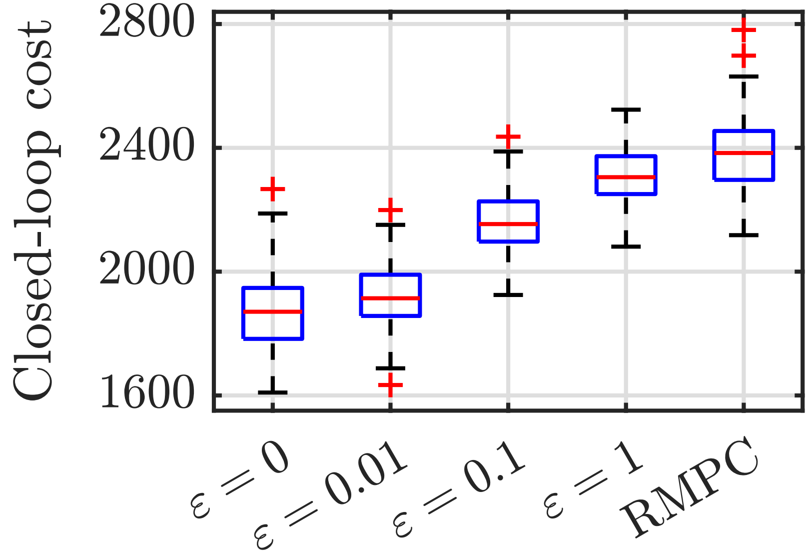
WT-MPC:

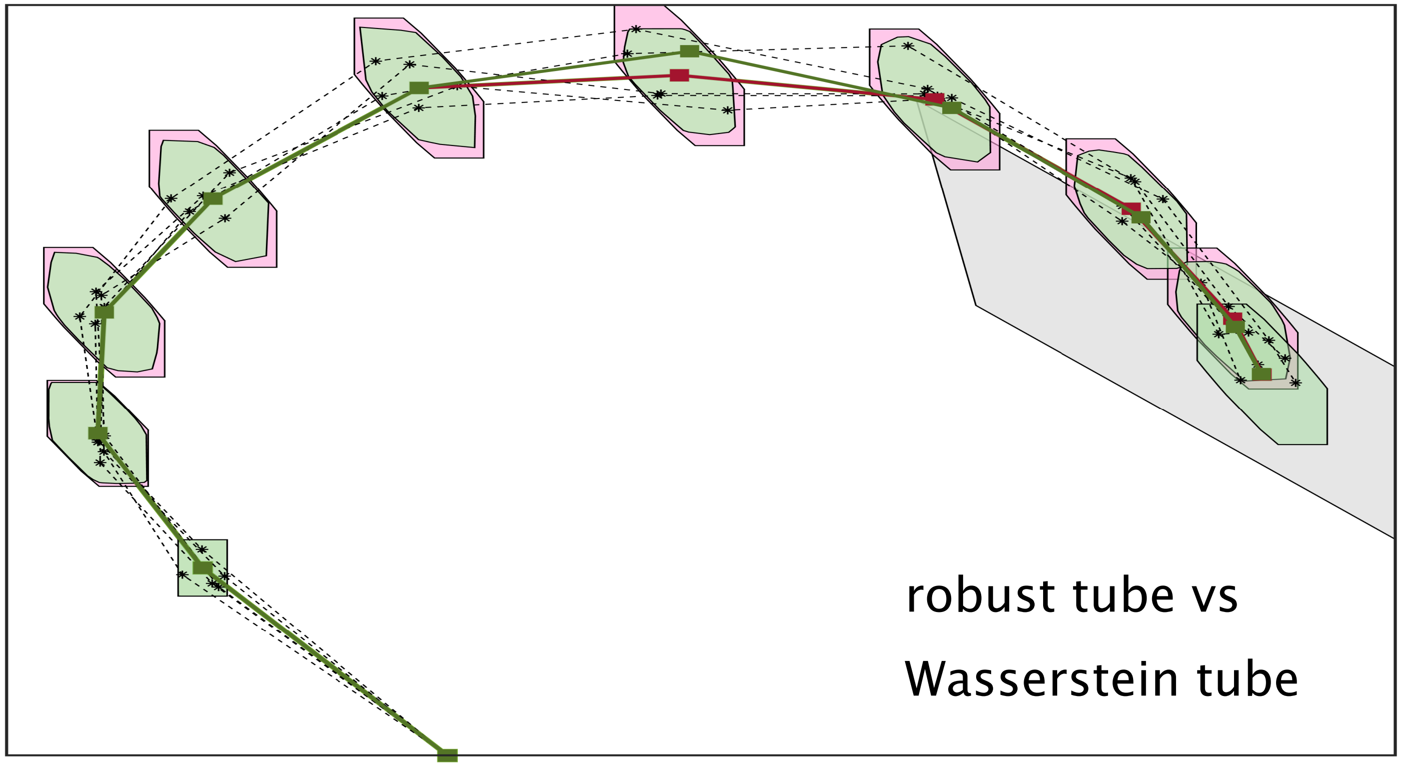
$$
\begin{aligned}
\min \quad & \sum_{t=0}^{N-1} \left(\|z_{k|t}\|_Q^2 + \|v_{k|t}\|_R^2 \right) \\
\text{s.t.} \quad & v_{k|t}, z_{k|t} \\
& z_{k+1|t} = (A + BK) z_{k|t} + B v_{k|t} \\
& K z_{k|t} + v_{k|t} \in \mathcal{U} \ominus K \mathcal{E}_k \\
& \boxed{\sup_{\mathbb{Q} \in \mathbb{S}_k} \mathrm{CVaR}_\delta^{\mathbb{Q}} \left(\max_j a_j^\top x_{k|t} + b_j \right) \leq 0} \\
& z_{N|t} \in \mathcal{Z}_f \\
& z_{0|t} = x_t
\end{aligned}
$$

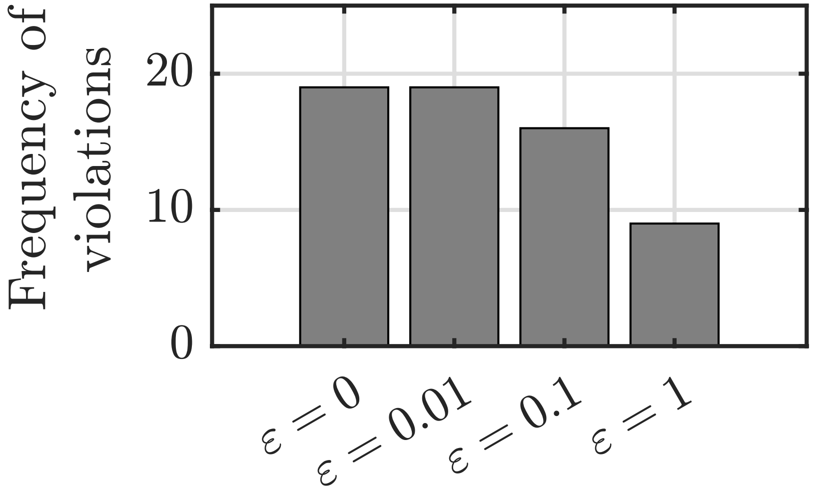
→ Set of convex deterministic constraints on $z_{k|t}$

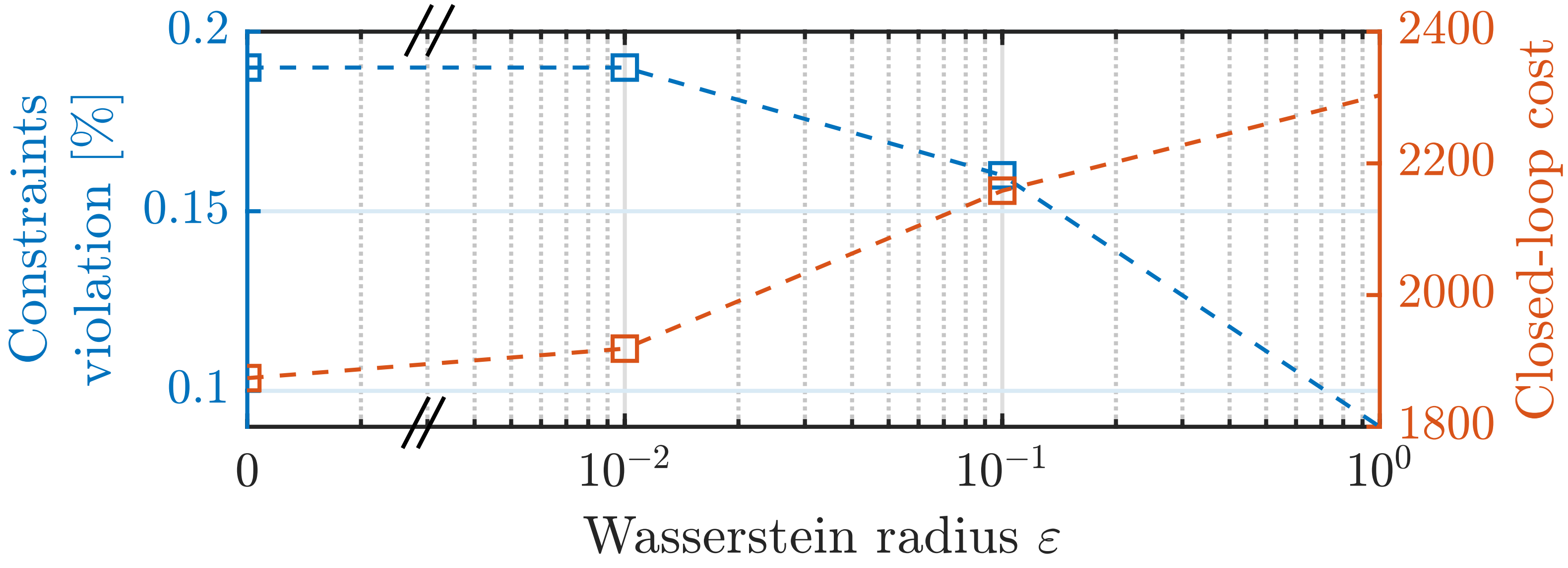
Theorem (informal).
- the optimal control problem is **efficiently solvable**,
- the receding-horizon implementation is **recursively feasible** under appropriate assumptions on $\mathcal{Z}_f$ & constraint tightening,
- … and **closed-loop stability** is still open ☺

Wasserstein Tube MPC





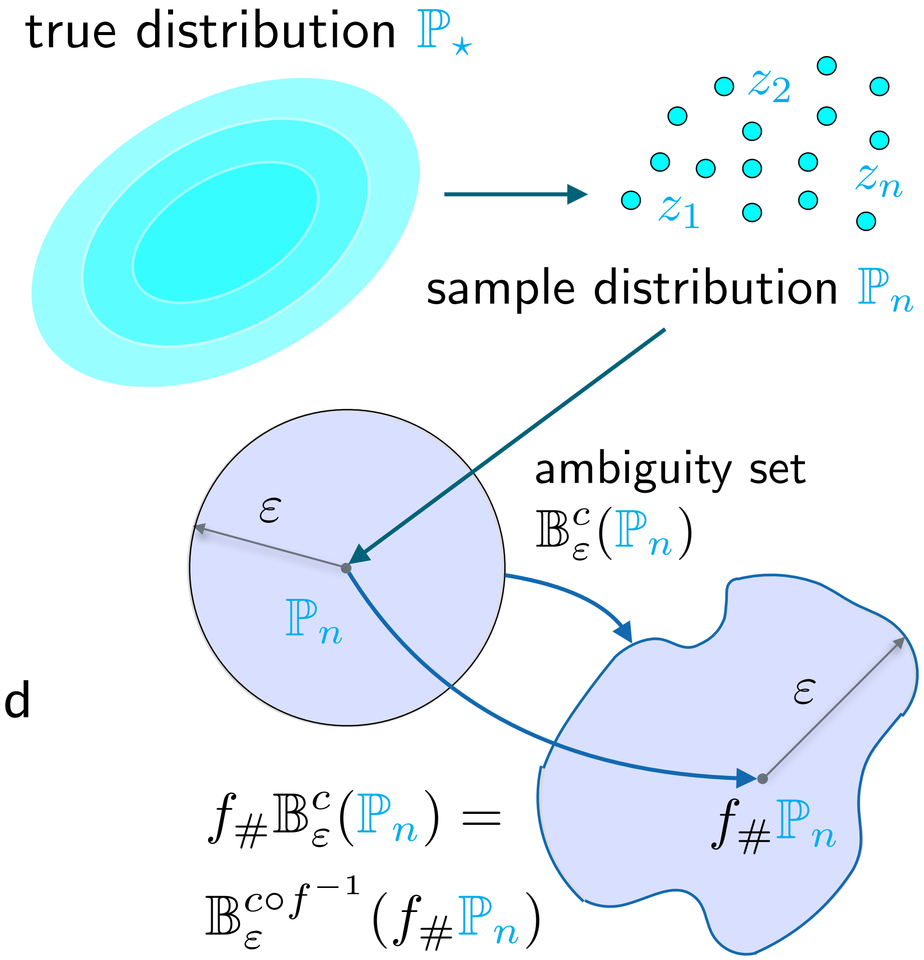




ETH zürich

Summary

- pitfalls of robust & stochastic uncertainty descriptions ✗
- OT ambiguity capturing of distributional uncertainty
- propagation results in yet another OT ambiguity set
- today's application to control: Wasserstein tube MPC



Conclusions: better uncertainty model ✓

- parametrizable from data
- strong statistical properties
- efficiently computable
- expressive: space & likelihood
- closure under propagation
- robust to distribution shift

Aolaritei, Lanzetti, Chen, Dörfler, Uncertainty Propagation via Optimal Transport Ambiguity Sets, 2022

Aolaritei, Lanzetti, Dörfler, Capture, Propagate, and Control Distributional Uncertainty, 2023

Aolaritei, Fochesato, Lygeros, Dörfler, Wasserstein Tube MPC with Exact Uncertainty Propagation, 2023

Shafieezadeh-Abadeh, Aolaritei, Dörfler, Kuhn, New Perspectives on Regularization and Computation in Optimal Transport-Based Distributionally Robust Optimization, 2023