# Learning Data-Driven Uncertainty sets with Mean Robust Optimization

Bartolomeo Stellato

PRINCETON UNIVERSITY

ICSP 2025

# Joint work with

**Irina Wang**
Princeton ORFE

**Marta Fochesato**
ETH Zürich

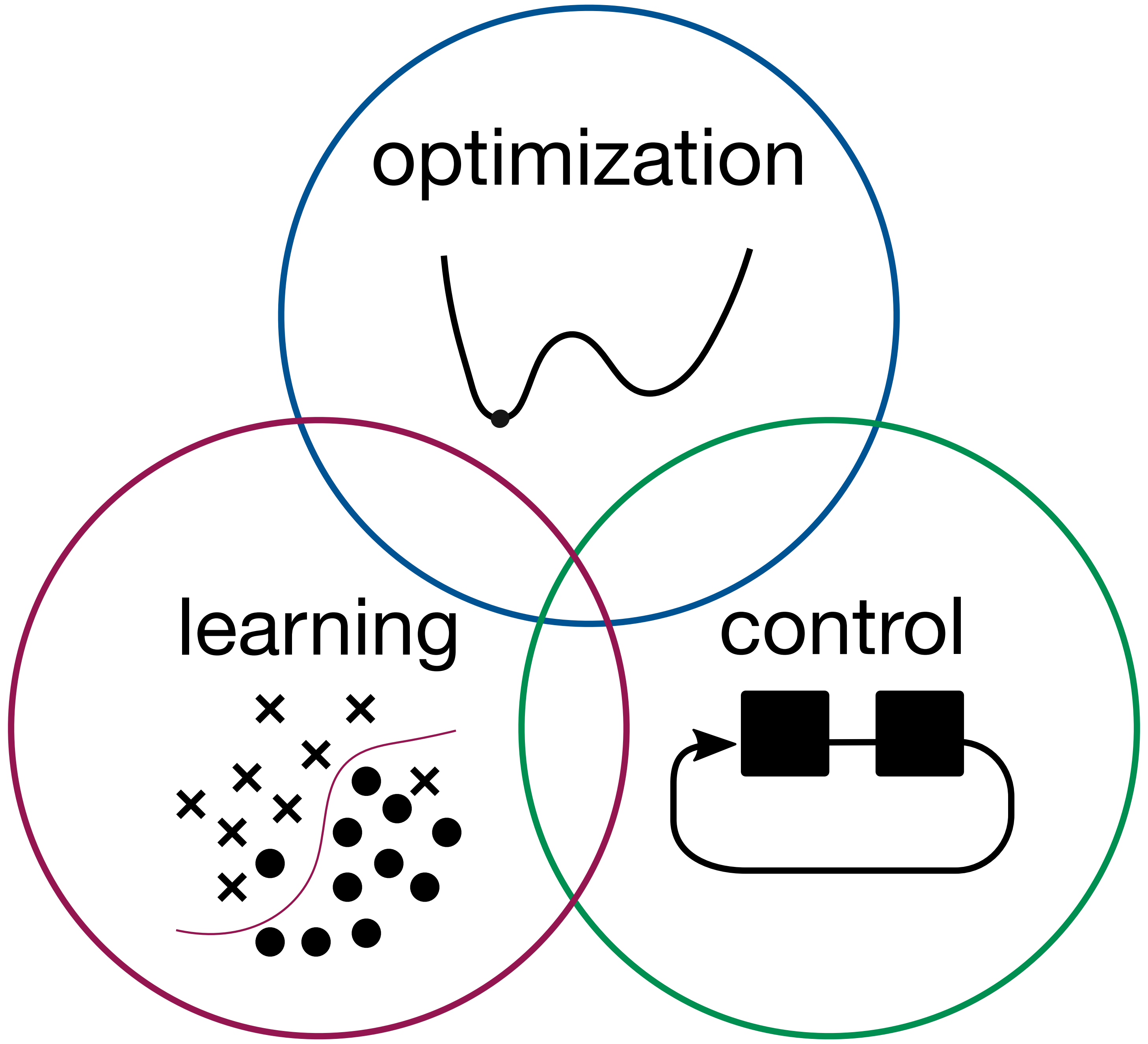
**Bart Van Parys**
CWI - the Netherlands

Marta Fochesato acknowledges the generous support of G-Research

# Most applications require fast and safe decisions in presence of uncertainty

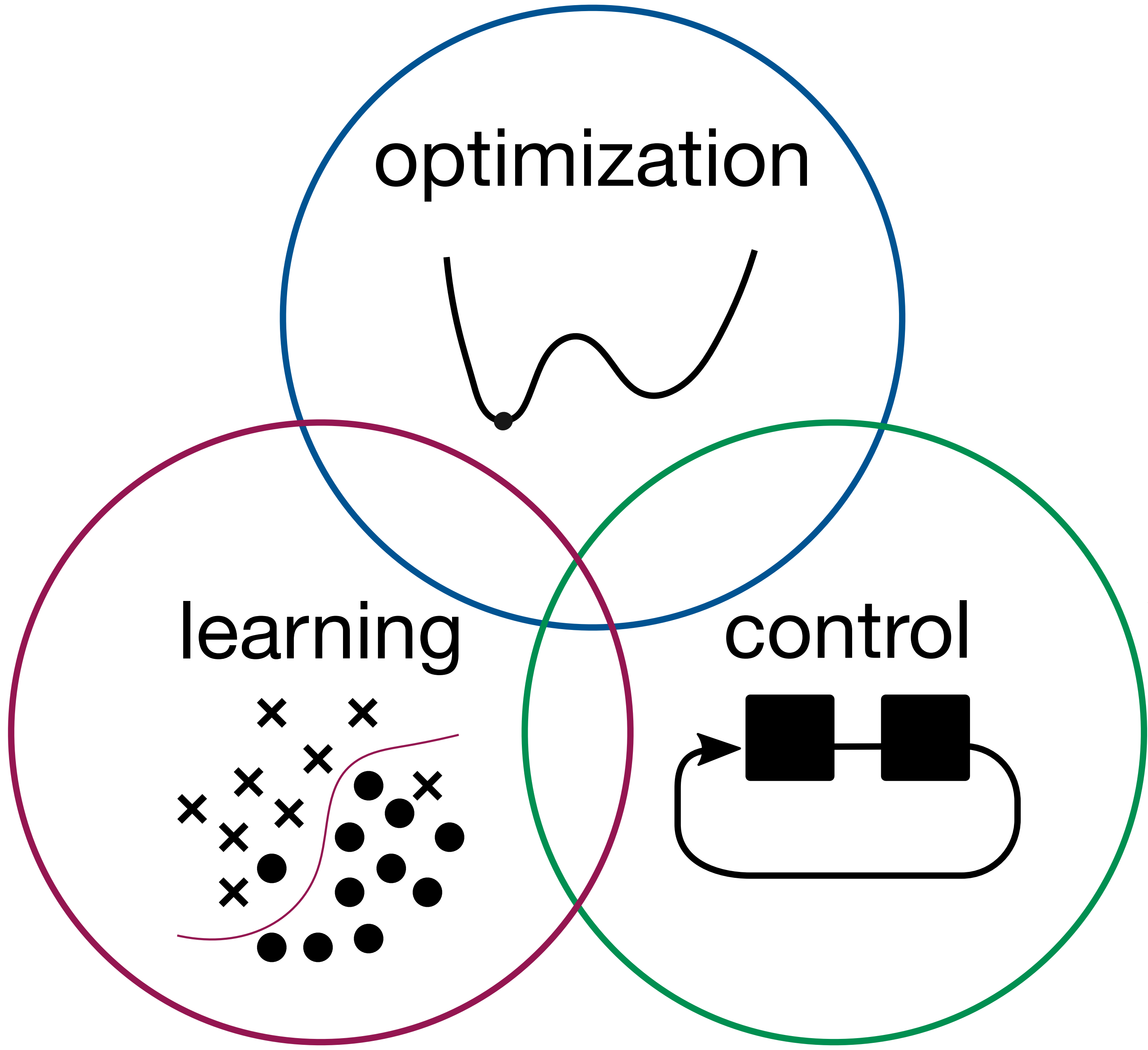
# My research in a 🥜shell

**theory**



# My research in a 🥜shell

## theory


optimization

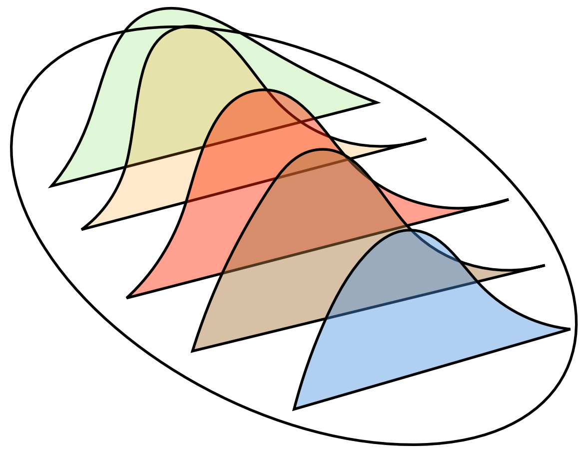
learning

control

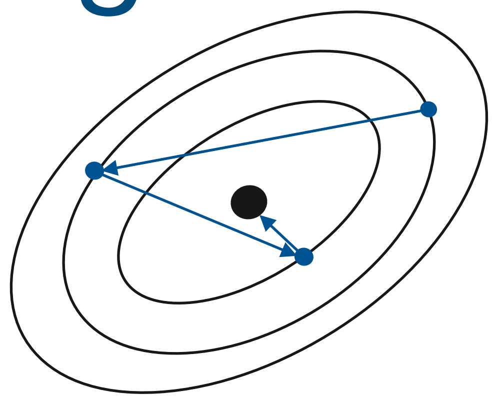
↔

## methodology

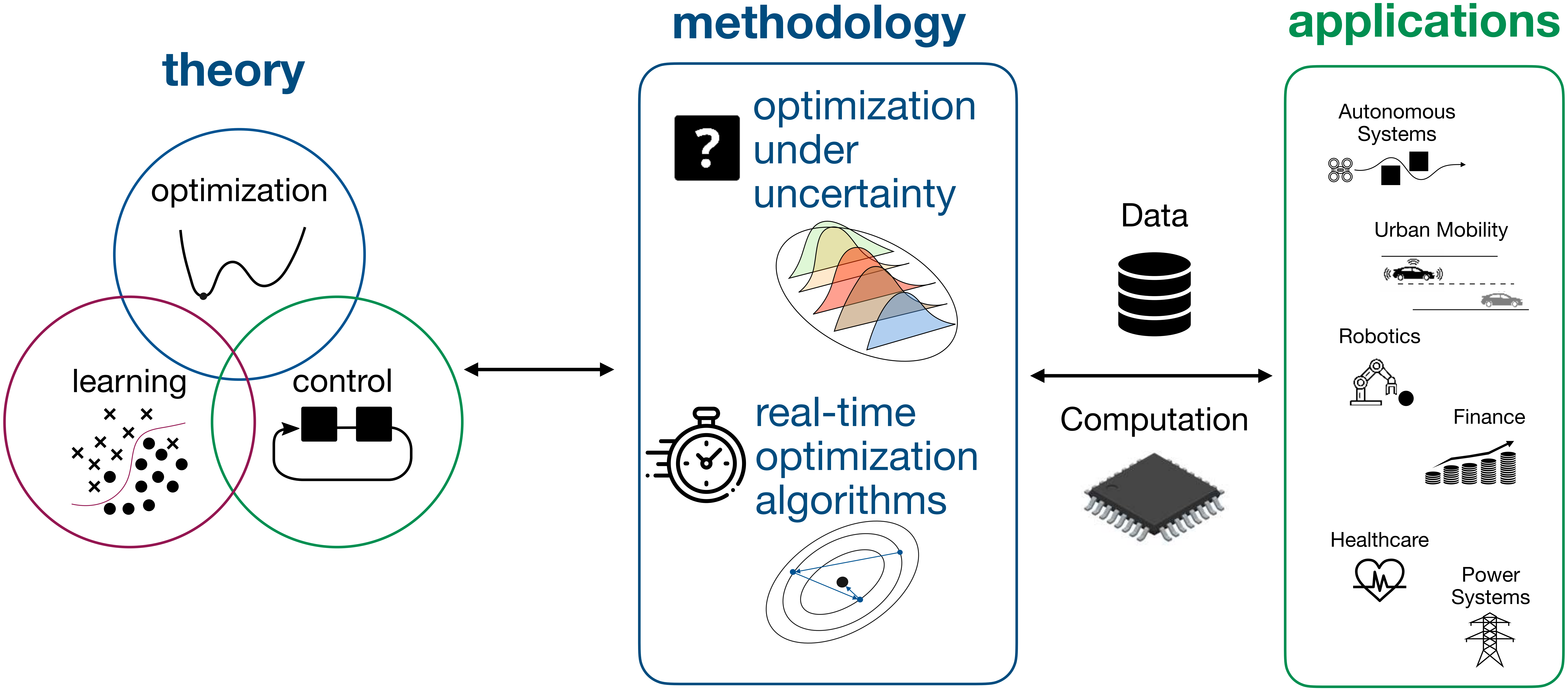
optimization under uncertainty



real-time optimization algorithms



# My research in a 🥜shell



## theory

optimization

learning

control

## methodology

optimization under uncertainty

real-time optimization algorithms

Data

Computation

## applications

Autonomous Systems

Urban Mobility

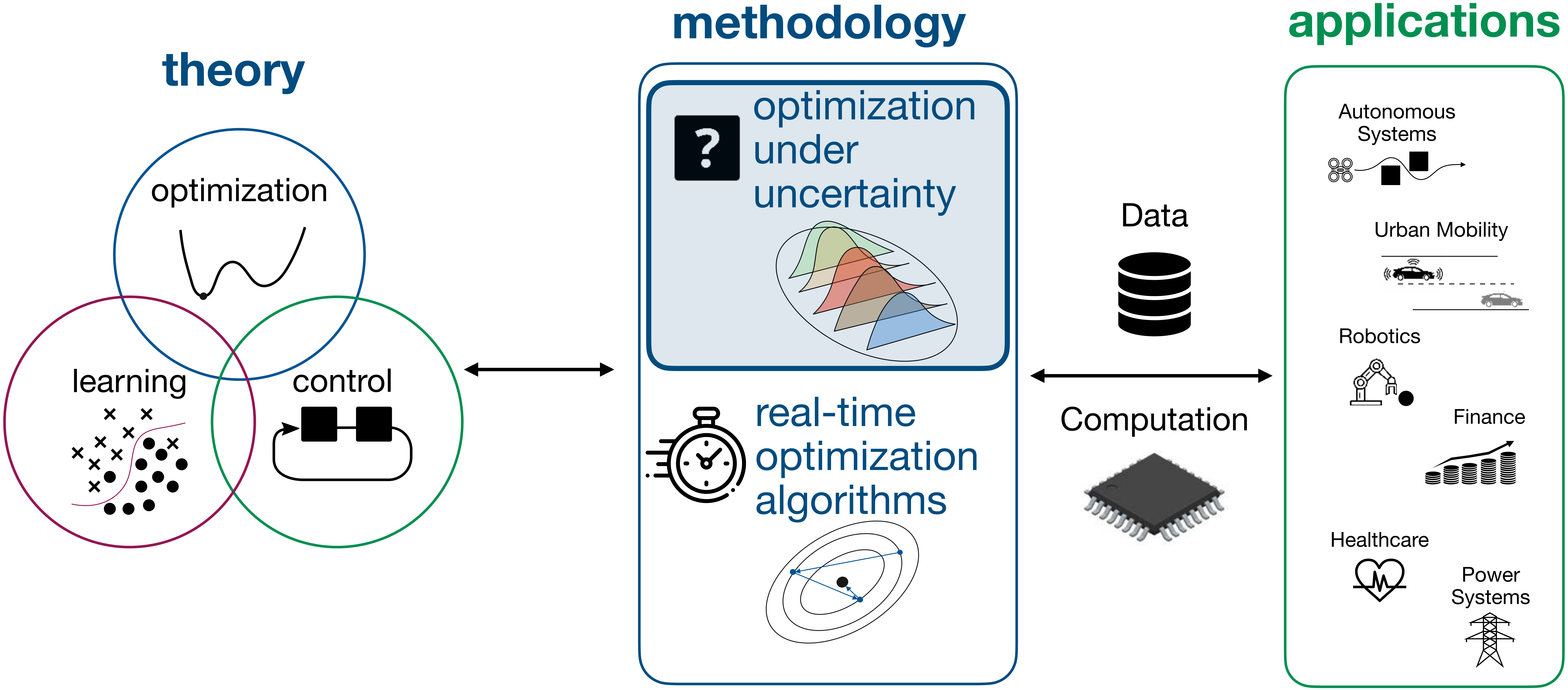
Robotics

Finance

Healthcare

Power Systems

# My research in a 🥜shell



**theory**

optimization

learning

control

**methodology**

optimization under uncertainty

real-time optimization algorithms

Data

Computation

**applications**

- Autonomous Systems
- Urban Mobility
- Robotics
- Finance
- Healthcare
- Power Systems

# Data can help us but we have to use it wisely!

# Data can help us but we have to use it wisely!

Moore's Law is slowing down

The New York Times

*Moore's Law Running Out of Room, Tech Looks for a Successor*

The Economist explains

The end of Moore's law

# Data can help us but we have to use it wisely!

## Moore's Law is slowing down

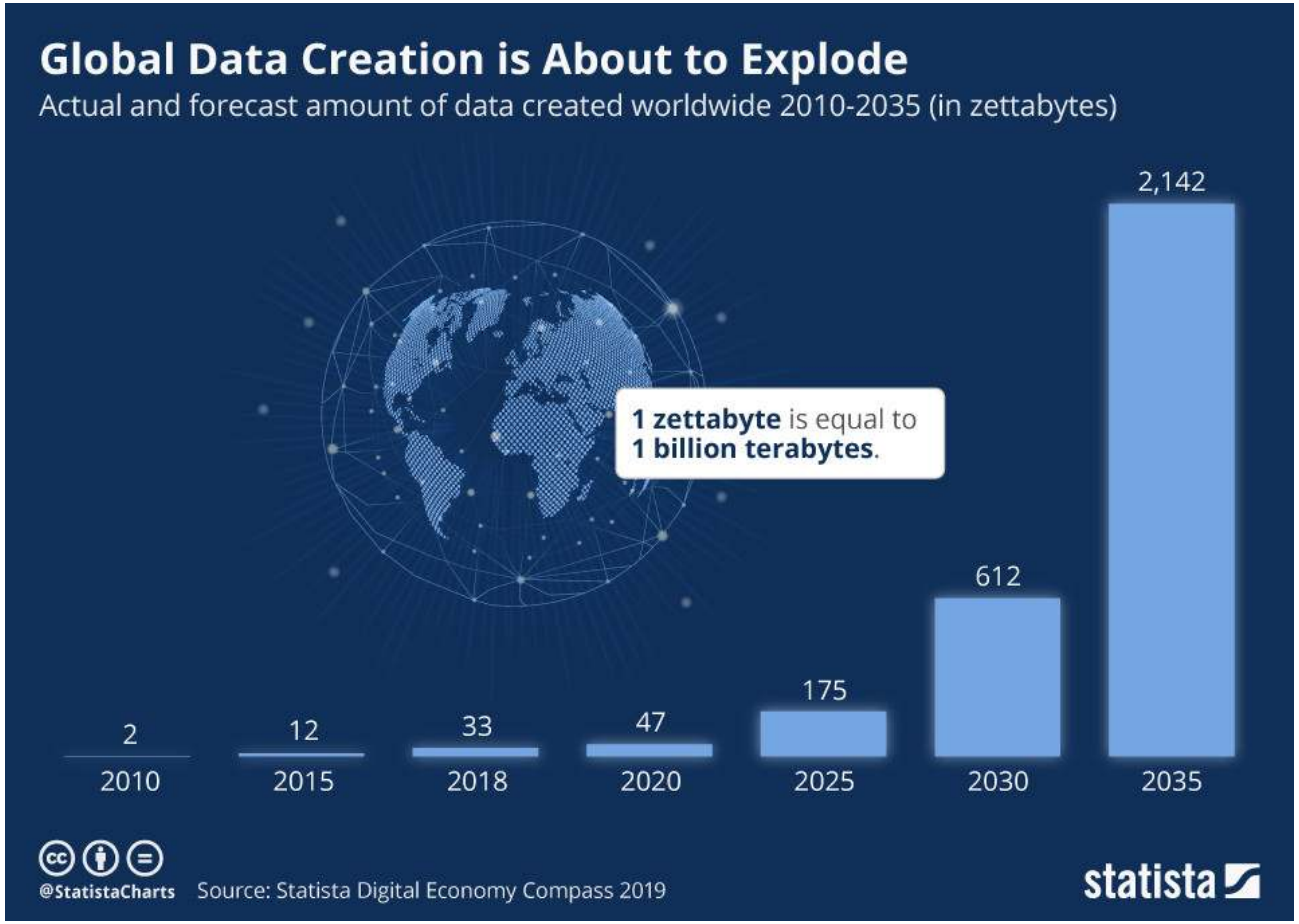
The New York Times

*Moore's Law Running Out of Room, Tech Looks for a Successor*

The Economist explains

The end of Moore's law

## Data availability is exploding



**Global Data Creation is About to Explode**

Actual and forecast amount of data created worldwide 2010-2035 (in zettabytes)

Source: Statista Digital Economy Compass 2019

# Problem setup with uncertain constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x,u) \leq 0 \end{array}$$

# Problem setup with uncertain constraints

optimization
variable

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x, u) \leq 0 \end{array}$$

# Problem setup with uncertain constraints

optimization
variable

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x, u) \leq 0 \end{array}$$

uncertain
parameter

# Problem setup with uncertain constraints

optimization
variable

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x, u) \leq 0 \end{array}$$

uncertain
parameter

How do we guarantee constraint satisfaction?

# Finite sample probabilistic guarantees in expectation

$$\mathbf{E}(g(x, u)) \leq 0$$

# Finite sample probabilistic guarantees in expectation

$$\mathbf{E}(g(x,u)) \leq 0$$

$u \sim P$

(but we never know $P$!)

# Finite sample probabilistic guarantees in expectation

$$\mathbf{E}(g(x,u)) \leq 0$$

$u \sim P$

(but we never know $P$!)

Data

$\mathcal{D} = \{d_i\}_{i=1}^{N}$

# Finite sample probabilistic guarantees in expectation

$$\mathbf{E}(g(x, u)) \leq 0$$

$u \sim P$

(but we never know $P$!)

Data

$\mathcal{D} = \{d_i\}_{i=1}^N$

Data-driven probabilistic guarantees

$$\mathbf{P}^N \left(\mathbf{E}(g(\hat{x}_N, u)) \leq 0\right) \geq 1 - \beta$$

# Finite sample probabilistic guarantees in expectation

$$\mathbf{E}(g(x,u)) \leq 0$$

$$u \sim P$$

(but we never know $P$!)

Data

$$\mathcal{D} = \{d_i\}_{i=1}^{N}$$

Data-driven probabilistic guarantees

product distribution →

$$\boxed{\mathbf{P}^N} \left(\mathbf{E}(g(\hat{x}_N, u)) \leq 0\right) \geq 1 - \beta$$

# Finite sample probabilistic guarantees in expectation

$$\mathbf{E}(g(x, u)) \leq 0$$

$u \sim P$

(but we never know $P$!)

Data

$\mathcal{D} = \{d_i\}_{i=1}^{N}$

Data-driven probabilistic guarantees

product distribution

$$\mathbf{P}^N \left(\mathbf{E}(g(\hat{x}_N, u)) \leq 0\right) \geq 1 - \beta$$

data-driven solution

# Finite sample probabilistic guarantees in expectation

$$\mathbf{E}(g(x, u)) \leq 0$$

$u \sim P$

(but we never know $P$!)

Data

$$\mathcal{D} = \{d_i\}_{i=1}^{N}$$

**Data-driven probabilistic guarantees**

$$\mathbf{P}^N \left(\mathbf{E}(g(\hat{x}_N, u)) \leq 0\right) \geq 1 - \beta$$

product distribution

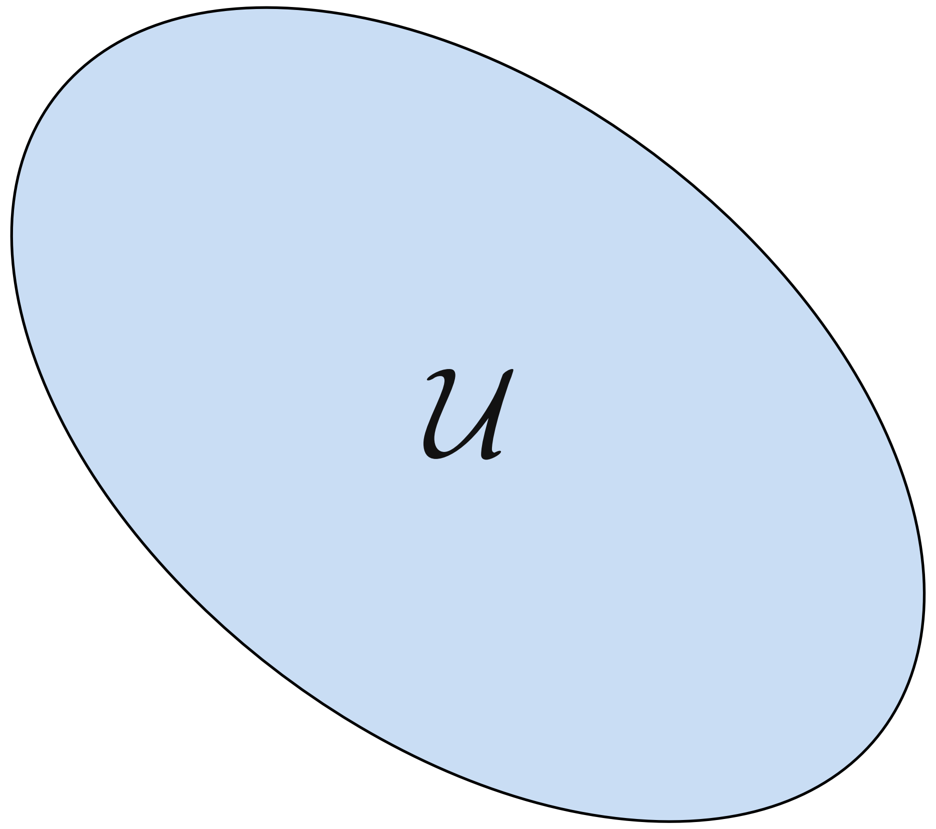
data-driven solution

probability of constraint satisfaction

# Robust optimization recipe

### Recipe

1. Pick uncertainty set $\mathcal{U}$
2. Ensure constraint satisfaction $\forall u \in \mathcal{U}$

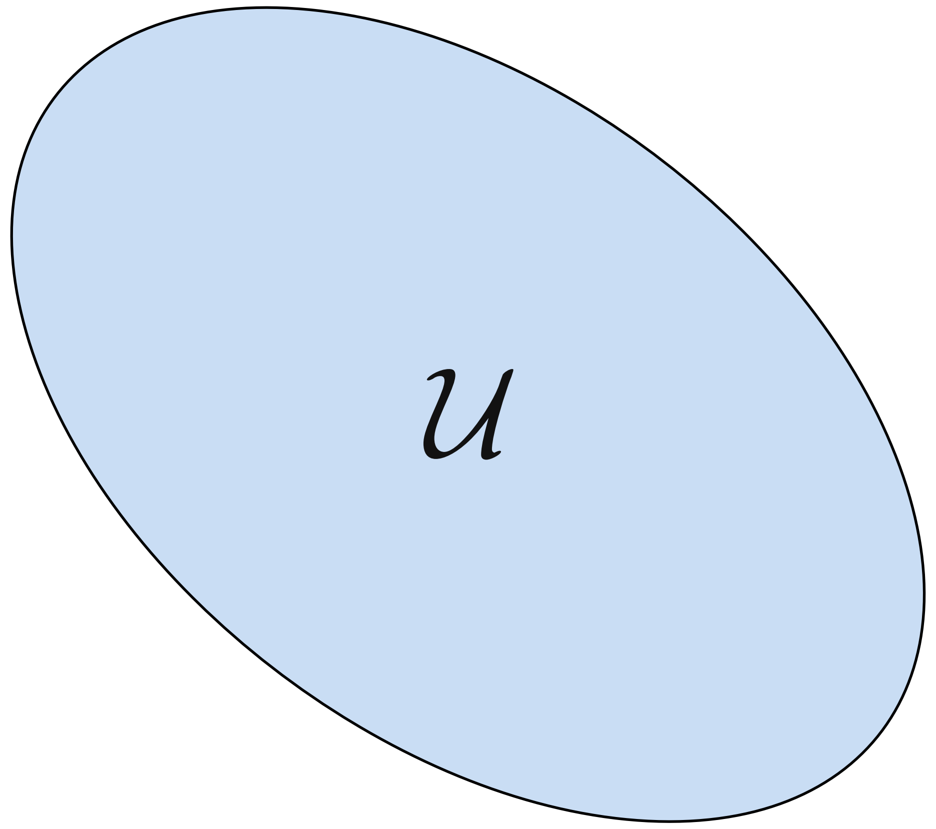


$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x, u) \leq 0, \quad \forall u \in \mathcal{U} \end{array}$$

# Robust optimization recipe

### Recipe

1. Pick uncertainty set $\mathcal{U}$
2. Ensure constraint satisfaction $\forall u \in \mathcal{U}$

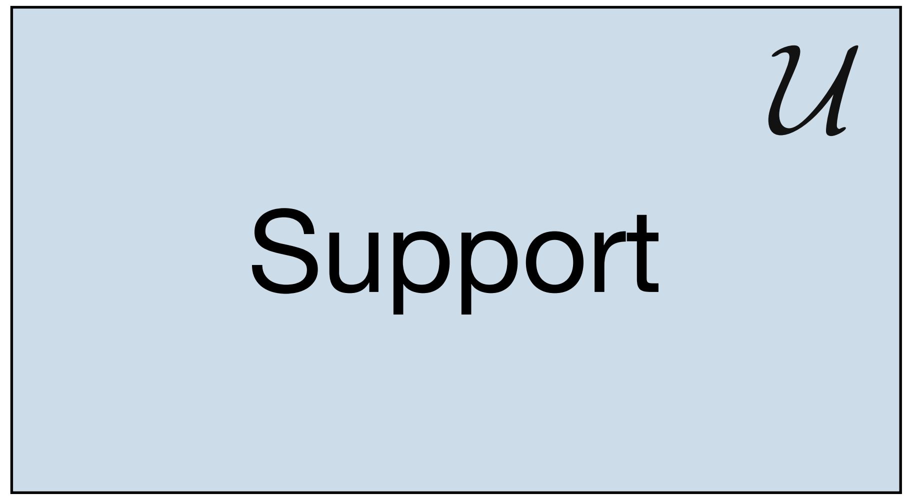


$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x,u) \leq 0, \quad \forall u \in \mathcal{U} \end{array}$$

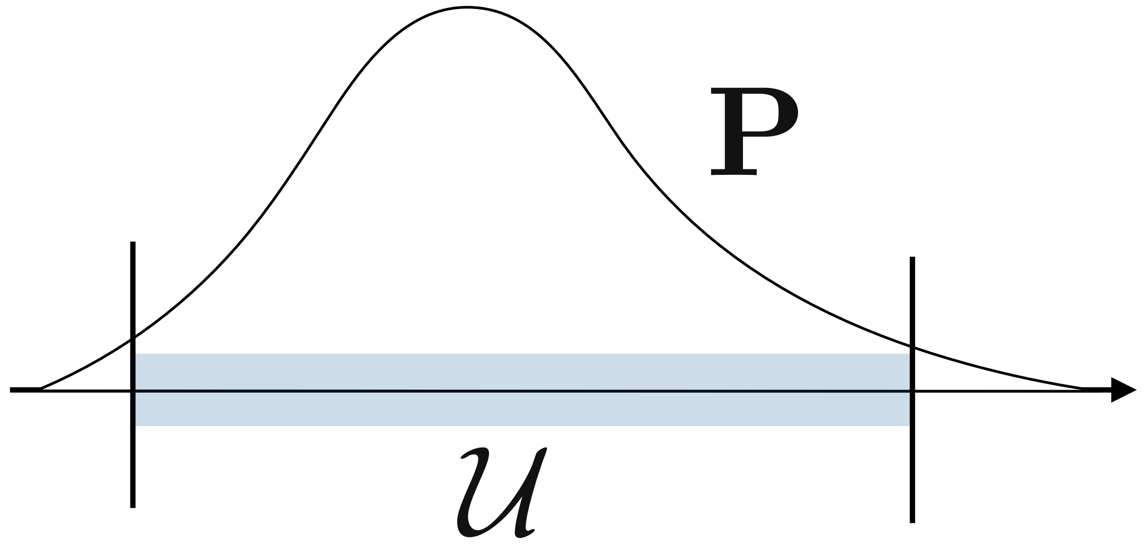
How do we pick the uncertainty set?

# Constructing the uncertainty set is difficult

*data-free approaches*



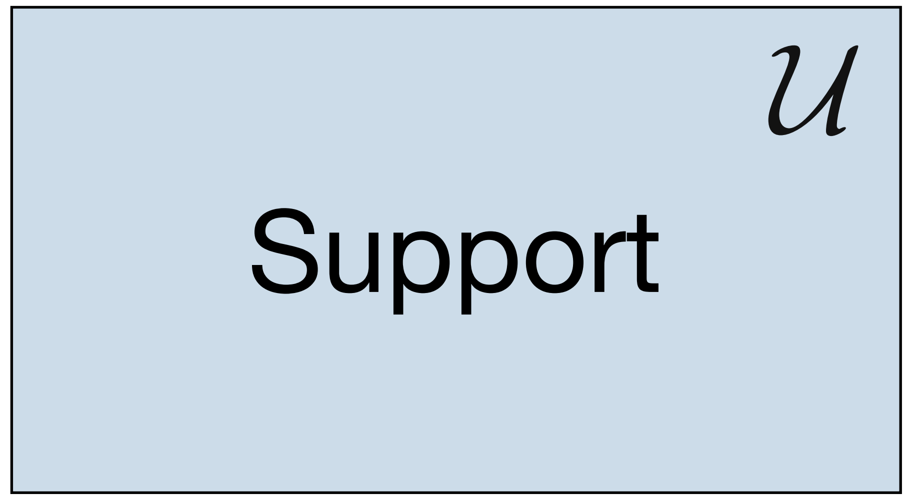
Worst-case approach



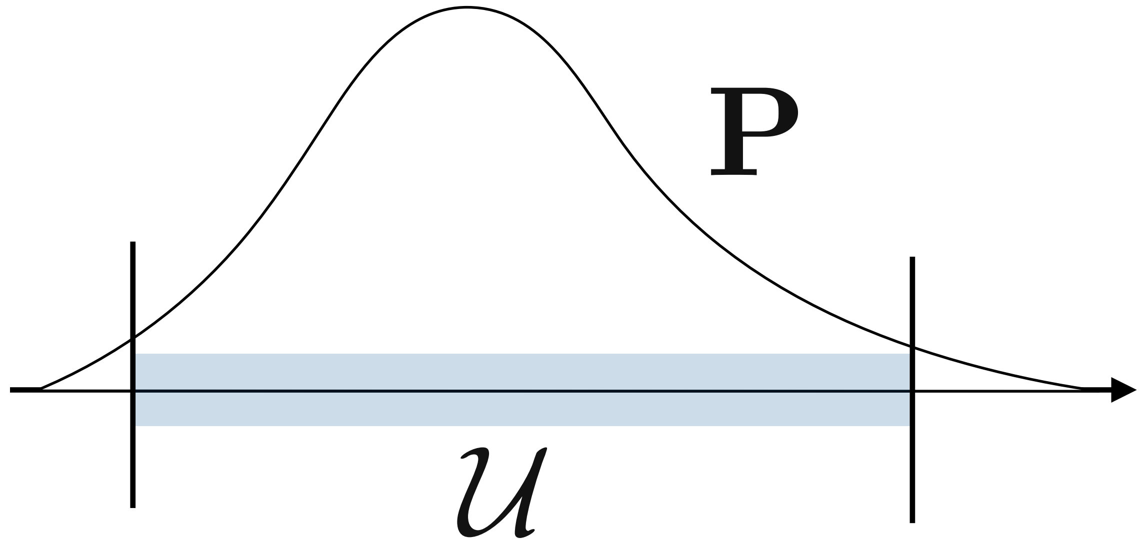
Probabilistic approach

# Constructing the uncertainty set is difficult

*data-free approaches*



Worst-case approach



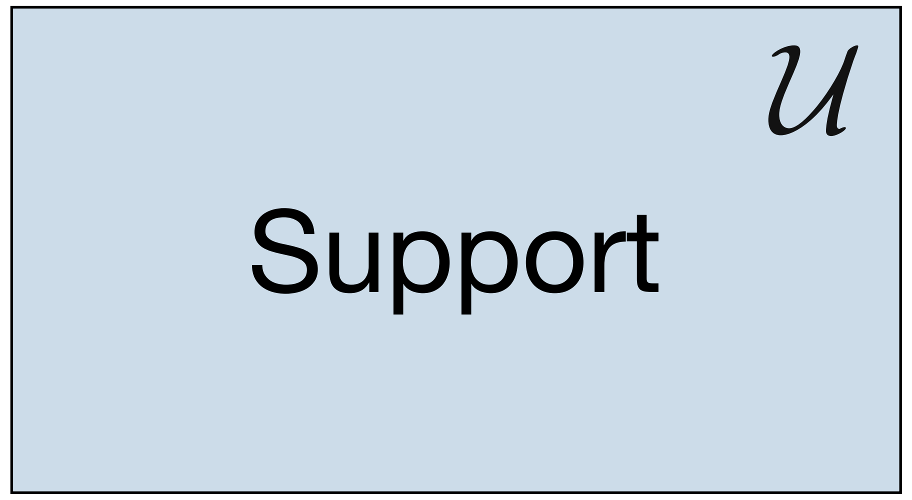
Probabilistic approach

**rely on *a-priori* pessimistic assumptions**

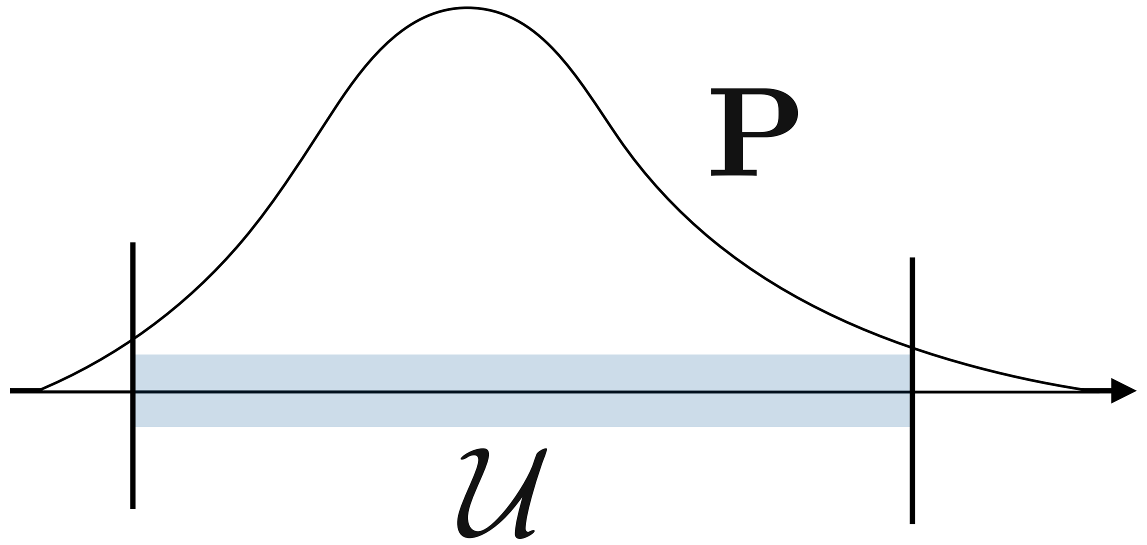
El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

# Constructing the uncertainty set is difficult

*data-free approaches*



Worst-case approach

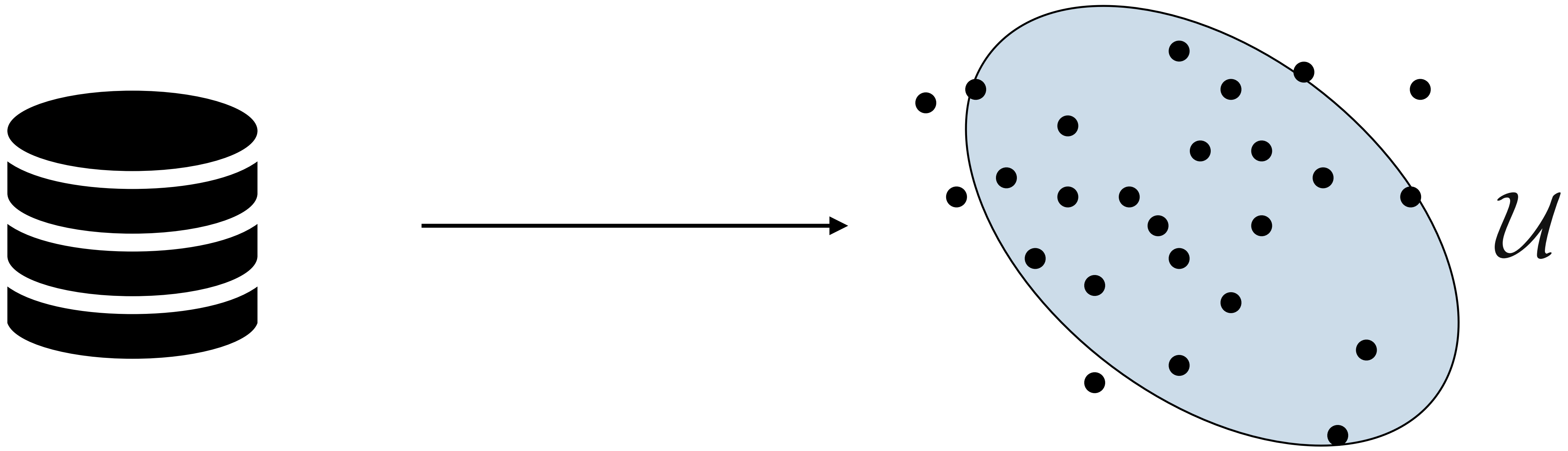


Probabilistic approach

rely on *a-priori* pessimistic assumptions

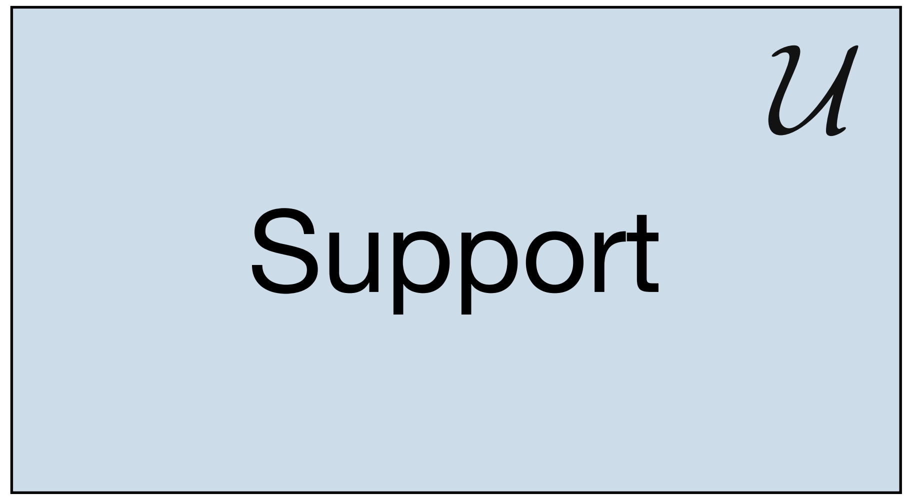
El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

*data-driven approaches*

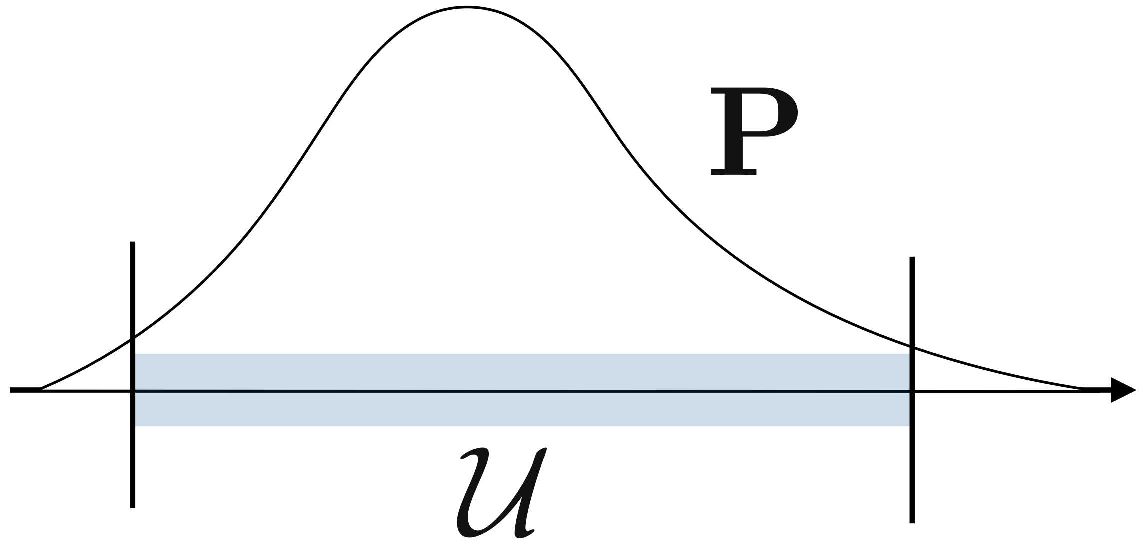


# Constructing the uncertainty set is difficult

***data-free approaches***



Worst-case approach

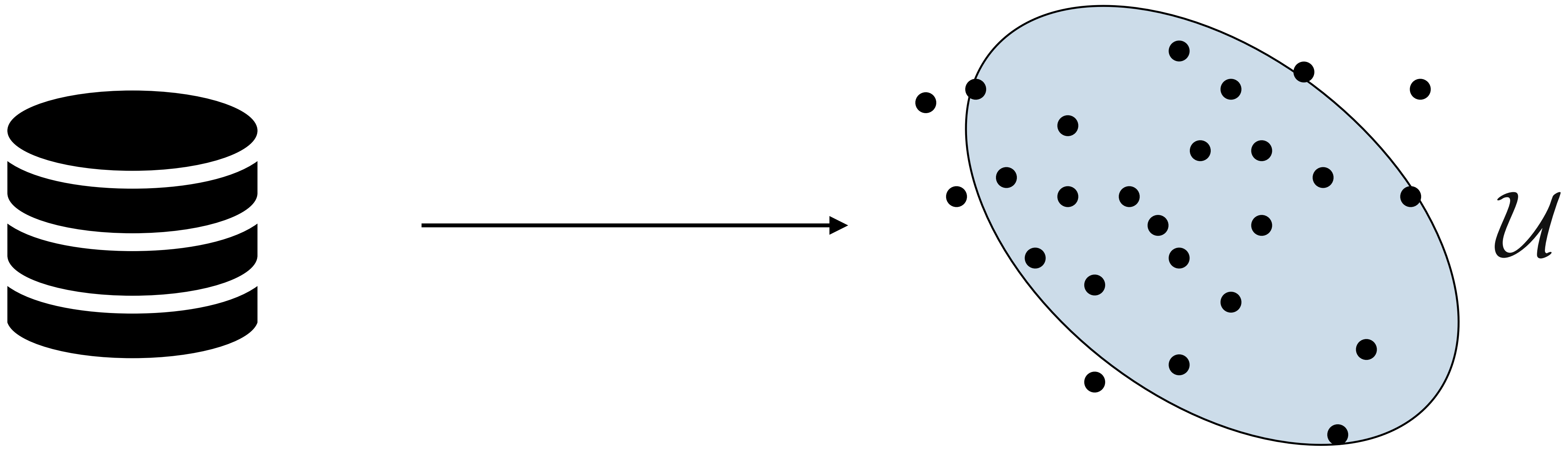


Probabilistic approach

**rely on *a-priori* pessimistic assumptions**

El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

***data-driven approaches***

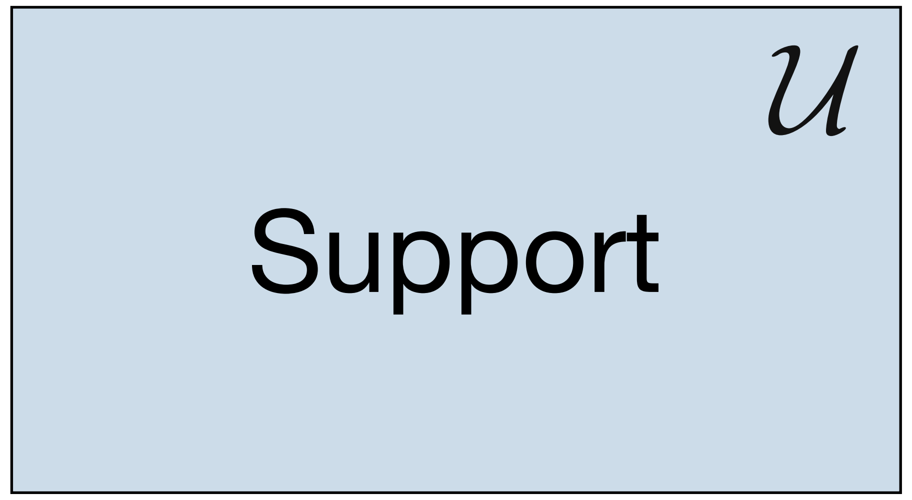


**hypothesis testing**

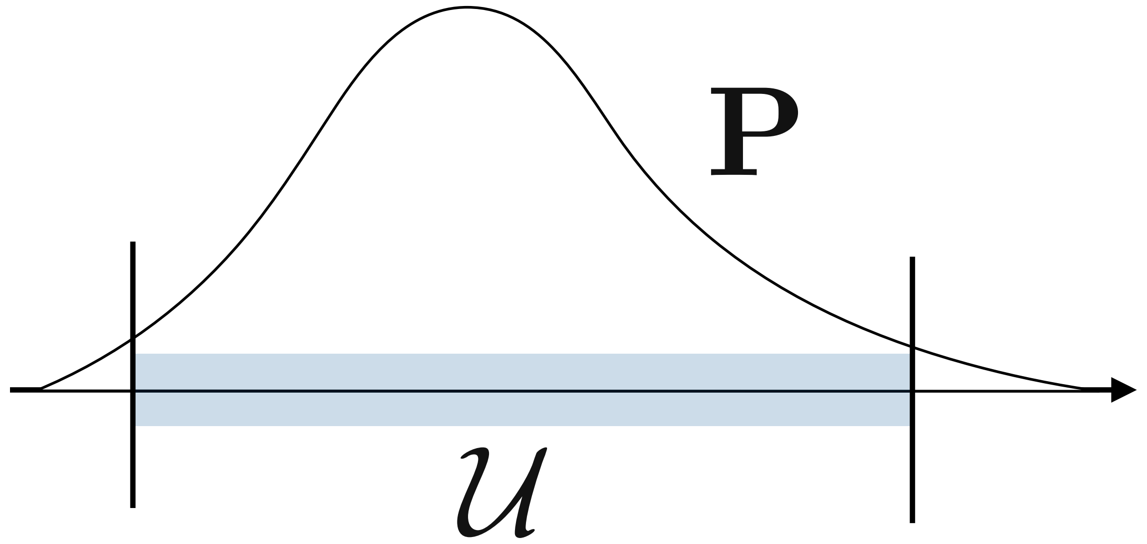
Bertsimas, Gupta, Kallus (2014)

# Constructing the uncertainty set is difficult

***data-free approaches***



Worst-case approach

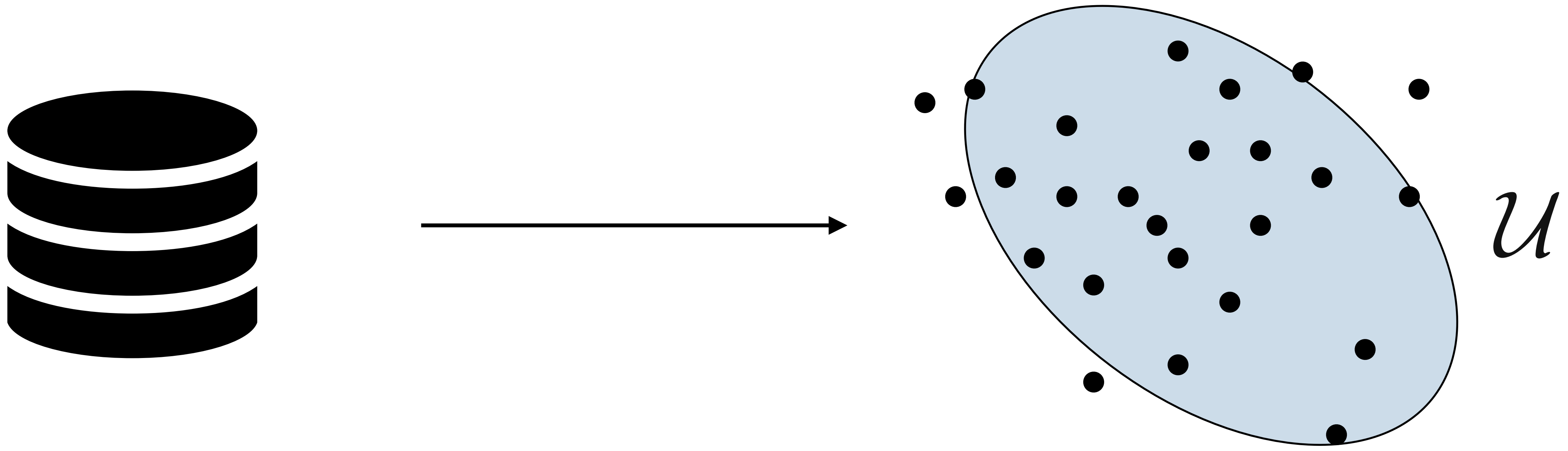


Probabilistic approach

**rely on *a-priori* pessimistic assumptions**

El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

***data-driven approaches***



**hypothesis testing**

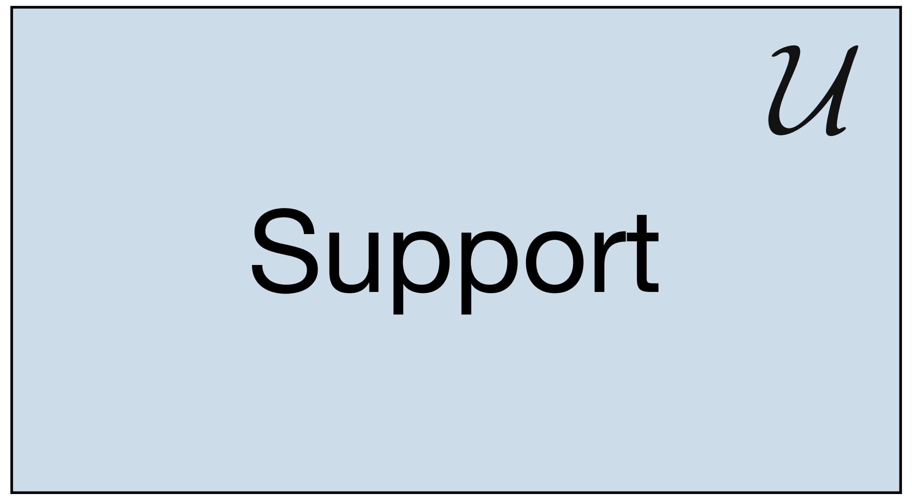
Bertsimas, Gupta, Kallus (2014)

**scenario approach**

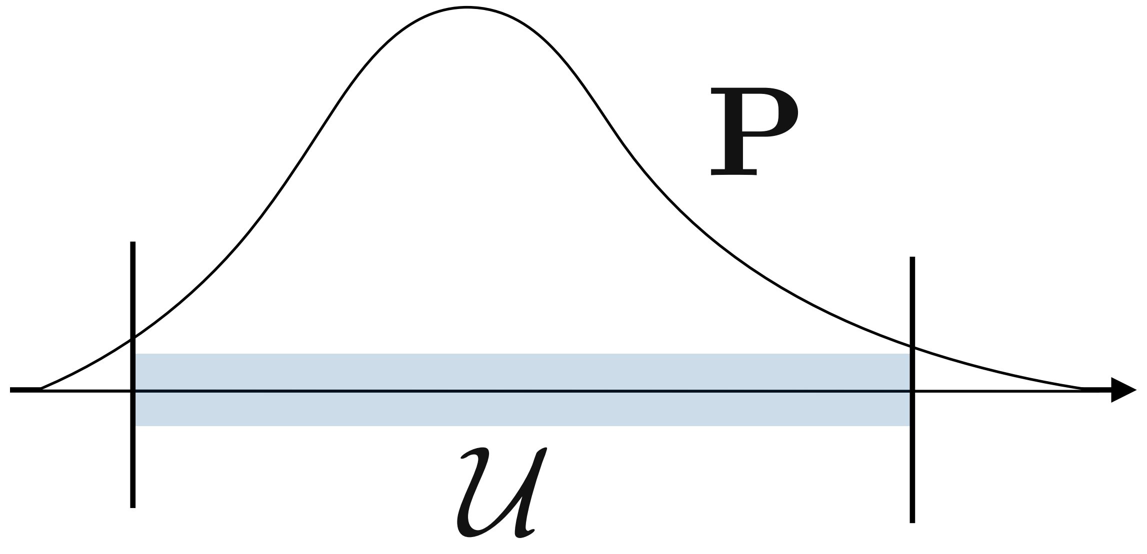
Calafiore and Campi (2006)

# Constructing the uncertainty set is difficult

*data-free approaches*



Worst-case approach

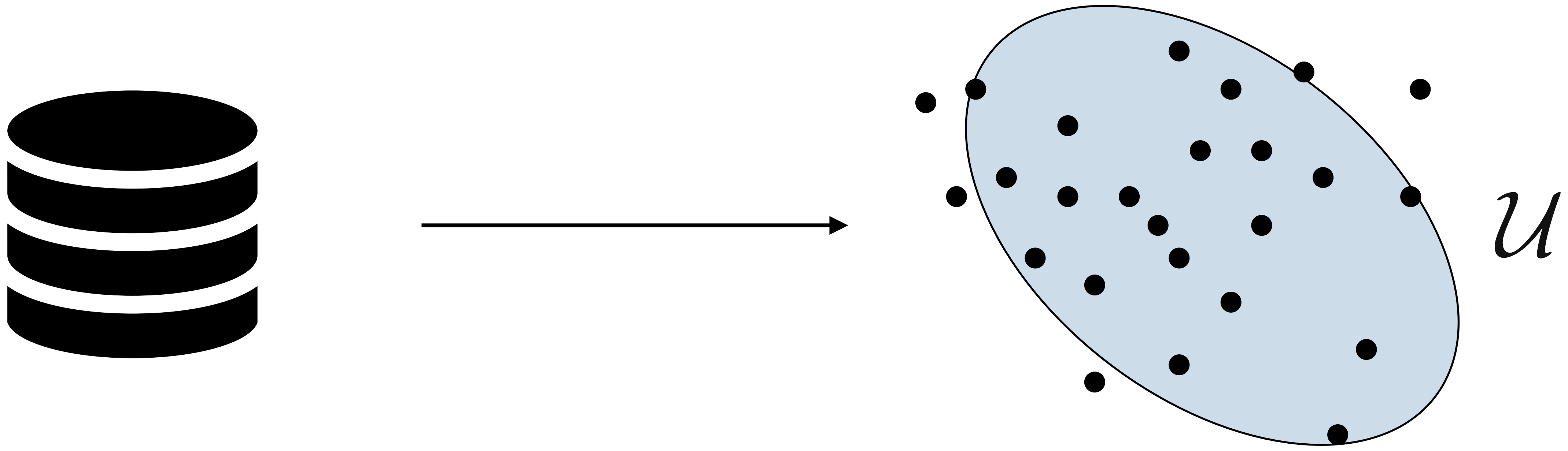


Probabilistic approach

**rely on *a-priori* pessimistic assumptions**

El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

*data-driven approaches*



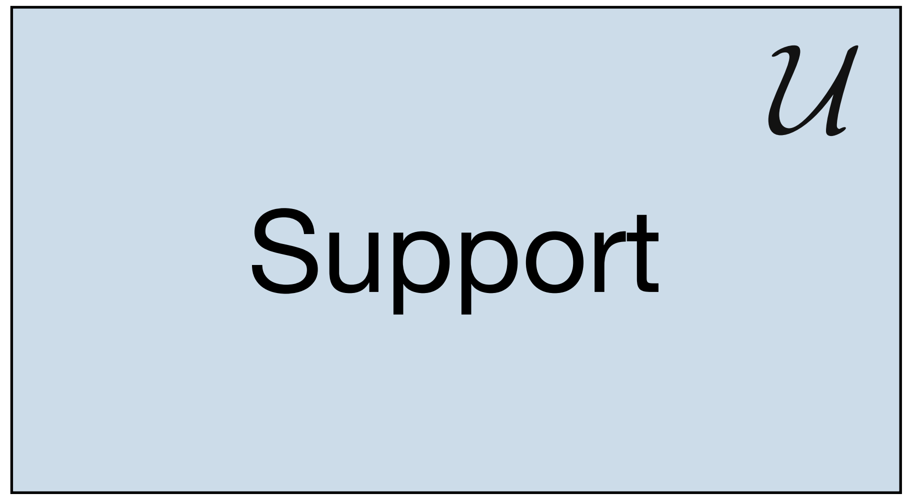
**hypothesis testing**
Bertsimas, Gupta, Kallus (2014)

**scenario approach**
Calafiore and Campi (2006)

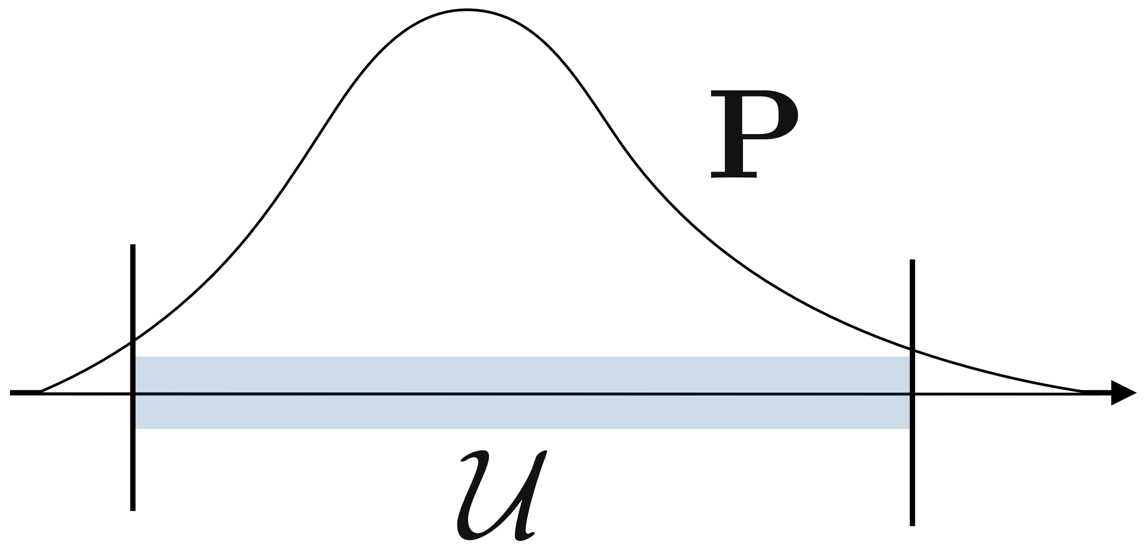
**quantile estimation**
Hong, Huang, Lam (2021)

# Constructing the uncertainty set is difficult

***data-free approaches***



Worst-case approach

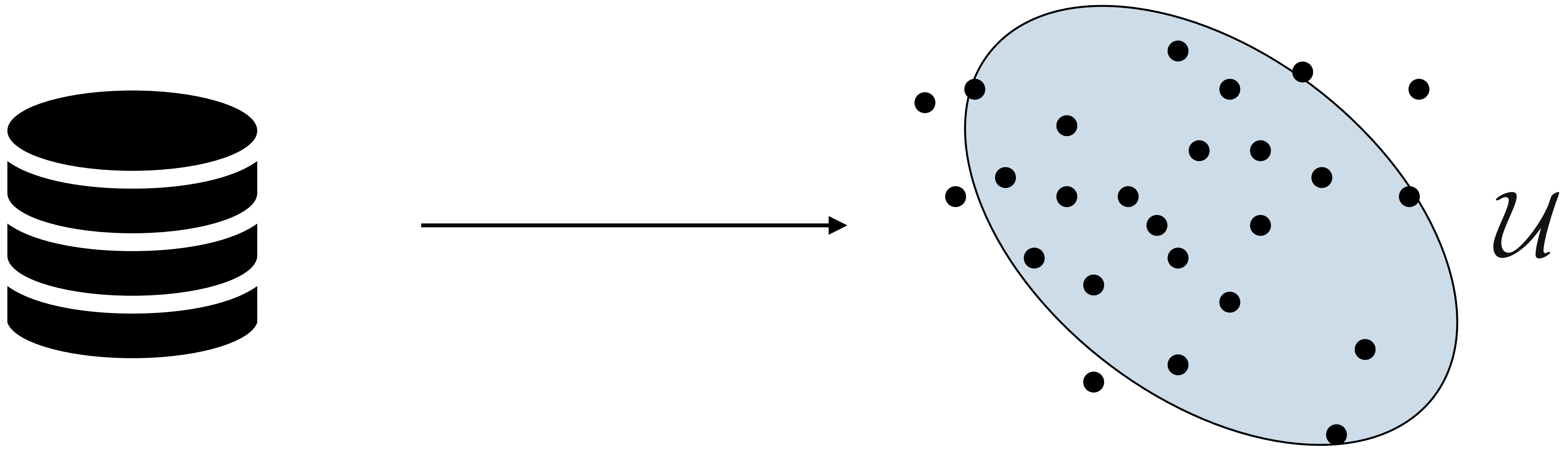


Probabilistic approach

**rely on *a-priori* pessimistic assumptions**

El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

***data-driven approaches***



**hypothesis testing**
Bertsimas, Gupta, Kallus (2014)

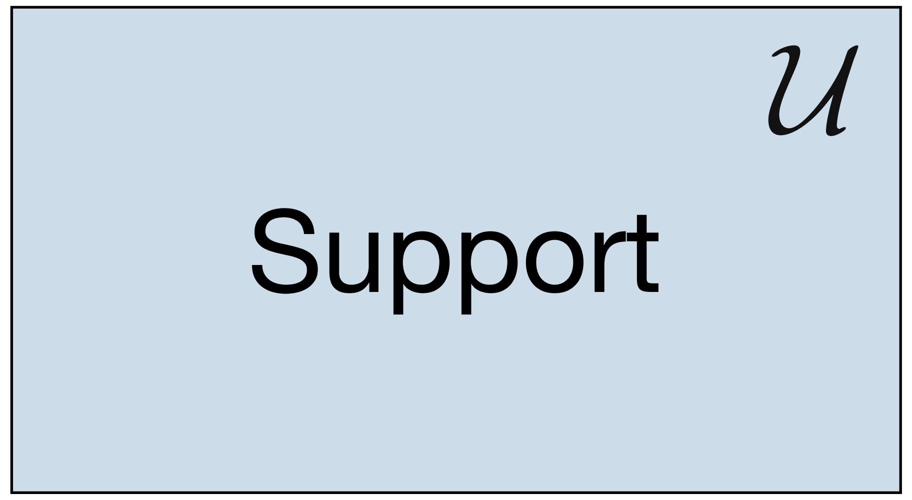
**scenario approach**
Calafiore and Campi (2006)

**quantile estimation**
Hong, Huang, Lam (2021)

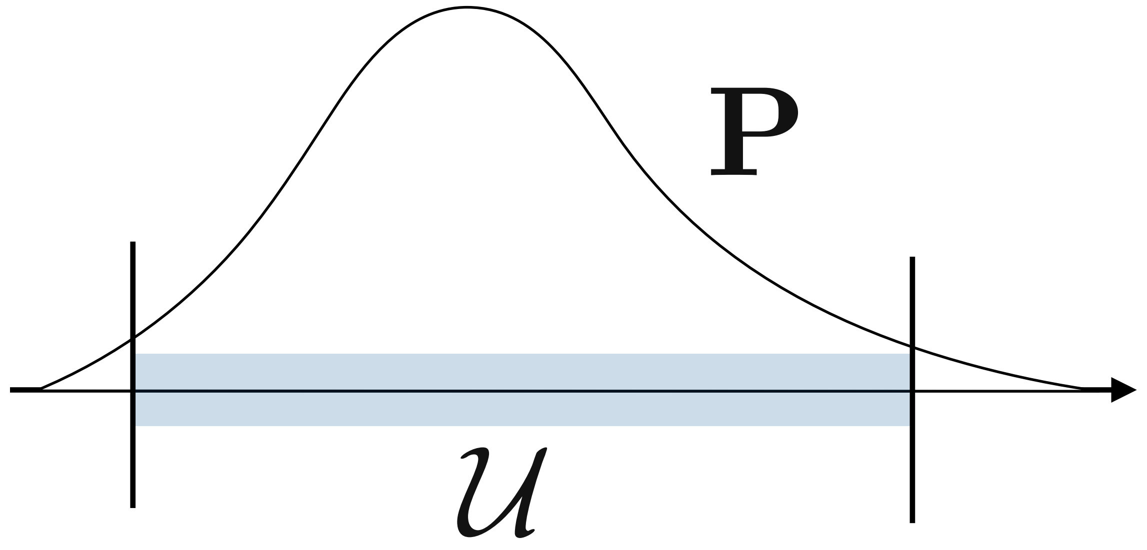
**deep learning**
Goerigk, Kurz (2025)

# Constructing the uncertainty set is difficult

*data-free approaches*



Worst-case approach

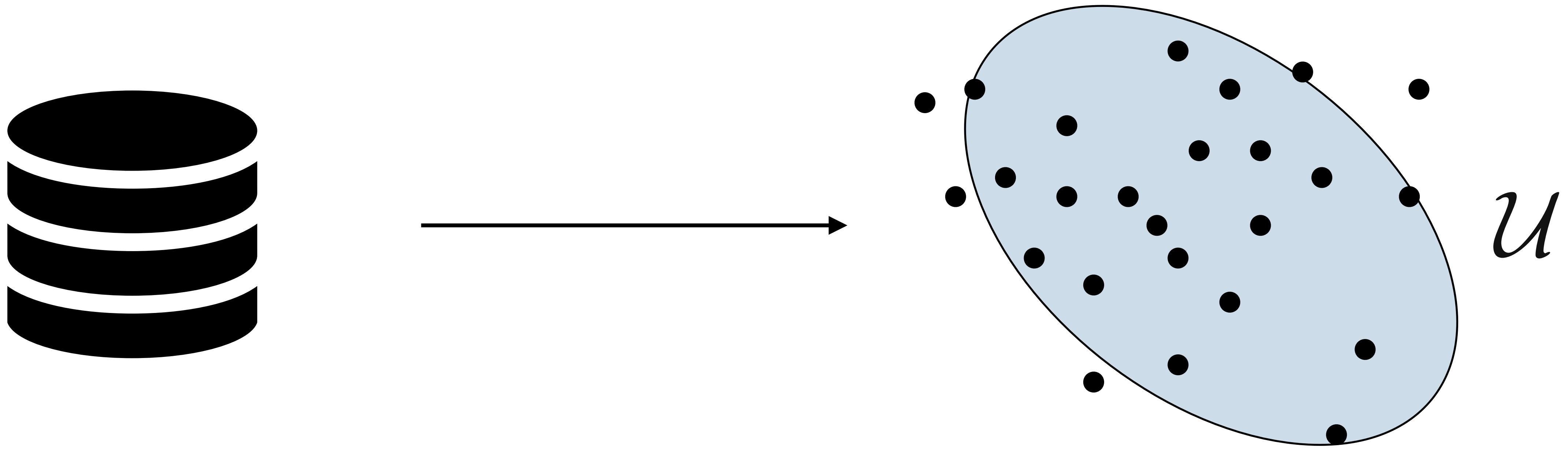


Probabilistic approach

**rely on *a-priori* pessimistic assumptions**

El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

*data-driven approaches*



**hypothesis testing**
Bertsimas, Gupta, Kallus (2014)

**scenario approach**
Calafiore and Campi (2006)

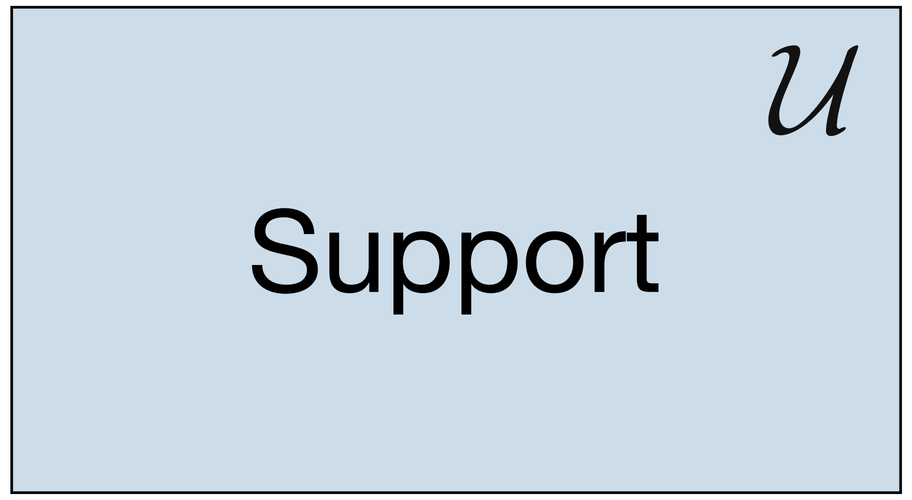
**quantile estimation**
Hong, Huang, Lam (2021)

**deep learning**
Goerigk, Kurz (2025)

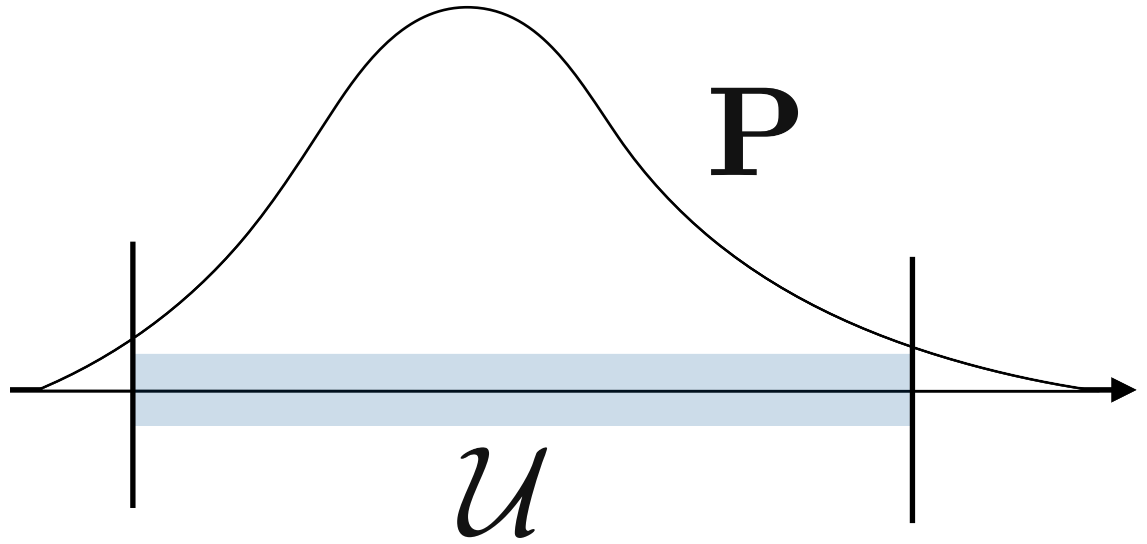
**end-to-end learning**
Chenreddy, Delage (2024)
Wang, Van Parys, Stellato (2025)

# Constructing the uncertainty set is difficult

***data-free approaches***



Worst-case approach

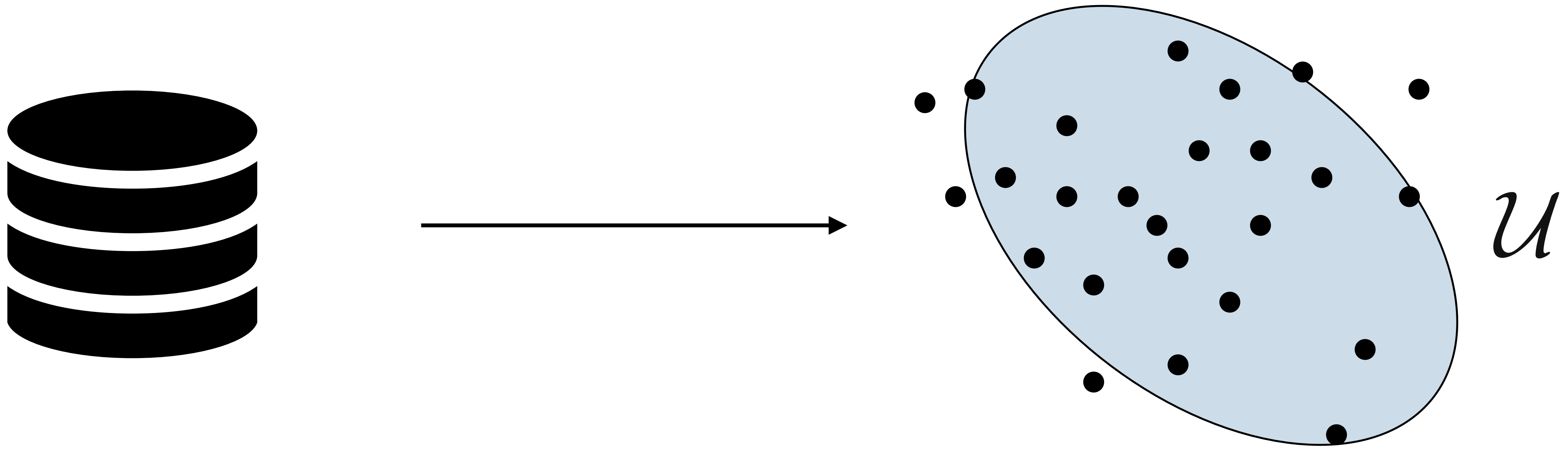


Probabilistic approach

**rely on *a-priori* pessimistic assumptions**

El Ghaoui, Ben-Tal, Bertsimas, Nemirovski, Sim, den Hertog…

***data-driven approaches***



**Irina Wang's Talk**
Jul 30, 2025, 12:20 PM
Caquot

**hypothesis testing**
Bertsimas, Gupta, Kallus (2014)

**scenario approach**
Calafiore and Campi (2006)

**quantile estimation**
Hong, Huang, Lam (2021)

**deep learning**
Goerigk, Kurz (2025)

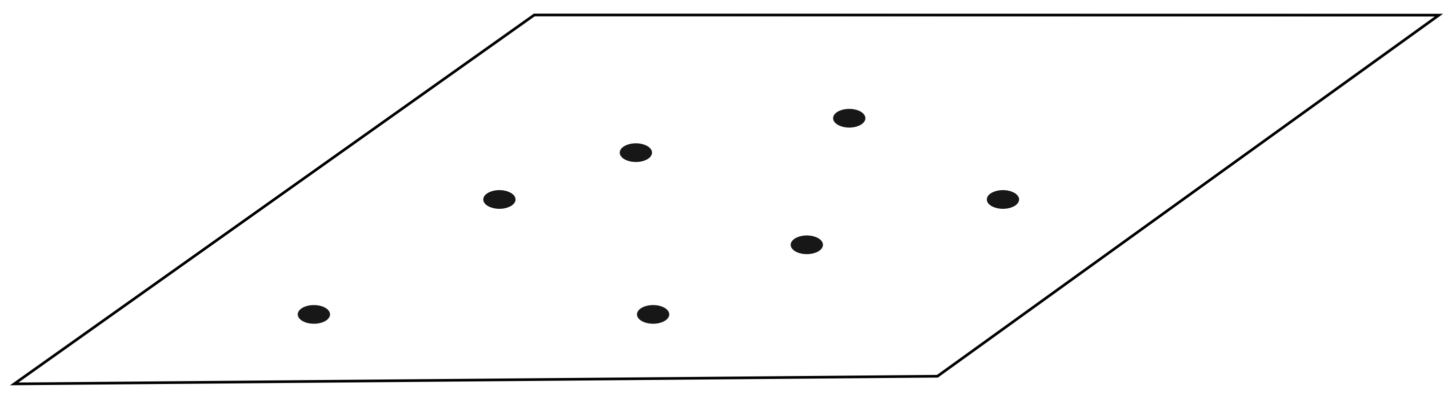
**end-to-end learning**
Chenreddy, Delage (2024)
Wang, Van Parys, Stellato (2025)

# The empirical distribution can help us build uncertainty sets

## Data-Driven Distributionally Robust Optimization

Data

$$\mathcal{D} = \{d_i\}_{i=1}^{N}$$



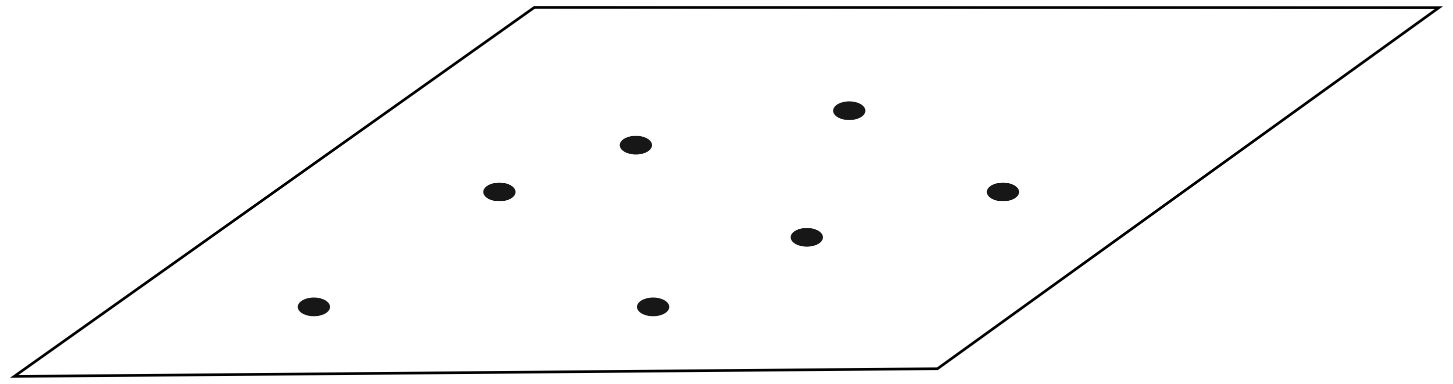
# The empirical distribution can help us build uncertainty sets

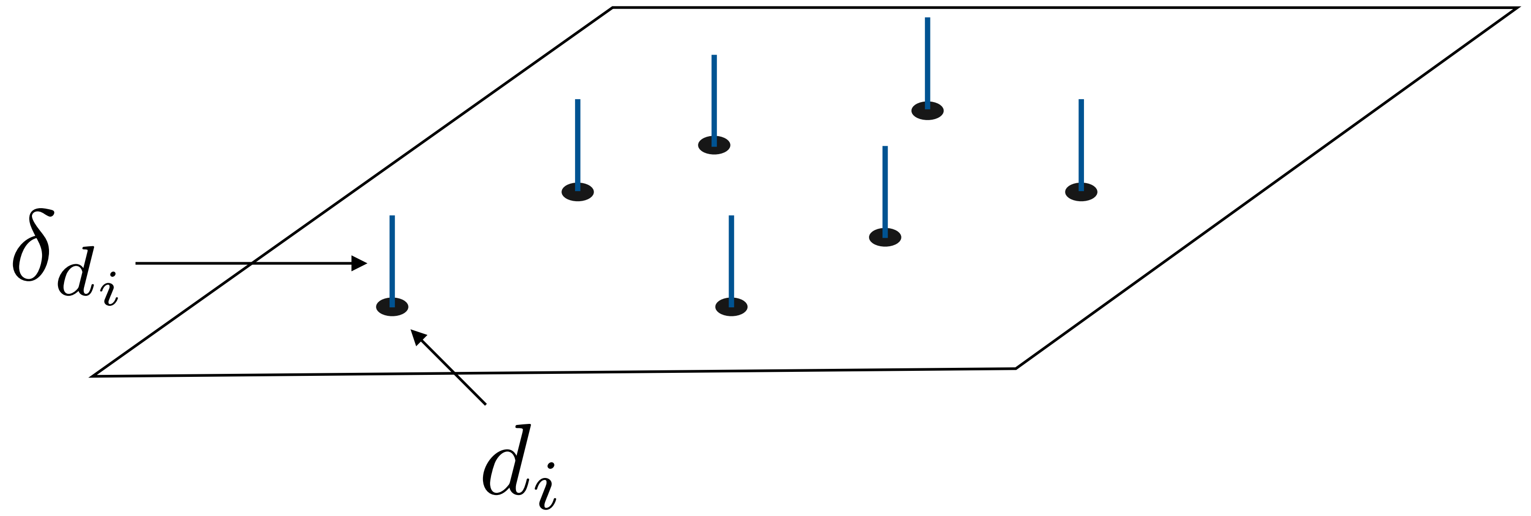
## Data-Driven Distributionally Robust Optimization

### Empirical Distribution

Data

$$\mathcal{D} = \{d_i\}_{i=1}^{N} \longrightarrow \hat{\mathbf{P}}^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{d_i}$$





# The empirical distribution can help us build uncertainty sets

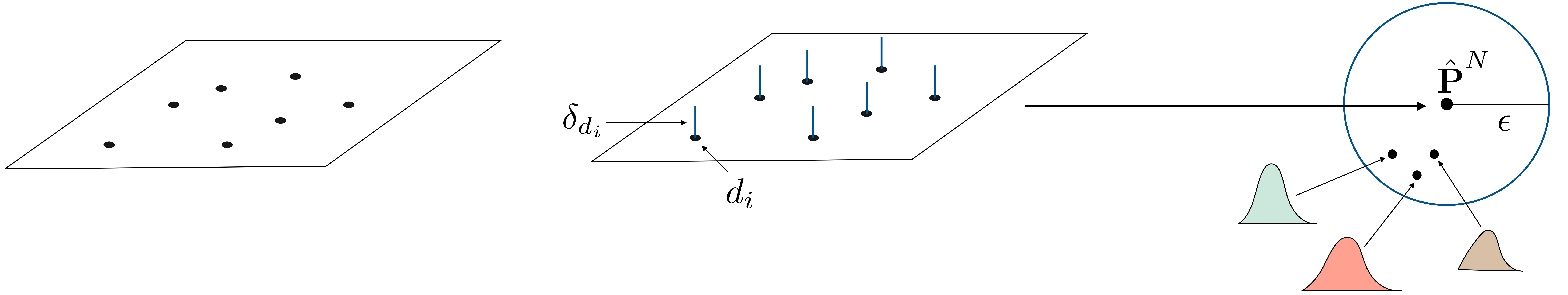
## Data-Driven Distributionally Robust Optimization

Data

**Empirical Distribution**

**Ambiguity Set**

$$\mathcal{D} = \{d_i\}_{i=1}^{N} \longrightarrow \hat{\mathbf{P}}^{N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{d_i} \longrightarrow \mathcal{P}_N = \left\{ W_p(\hat{\mathbf{P}}^{N}, \mathbf{P}) \leq \epsilon \right\}$$



# The empirical distribution can help us build uncertainty sets

## Data-Driven Distributionally Robust Optimization

### Data

$$\mathcal{D} = \{d_i\}_{i=1}^{N} \longrightarrow$$

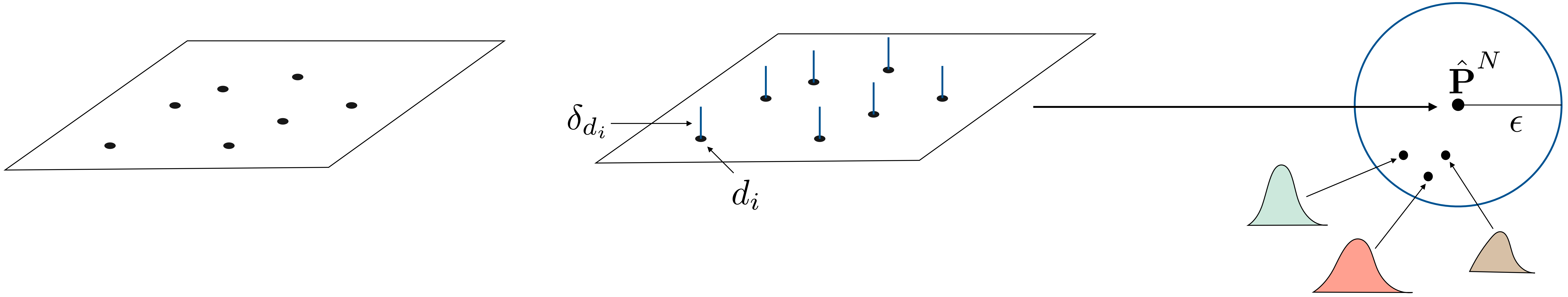
### Empirical Distribution

$$\hat{\mathbf{P}}^{N} = \frac{1}{N}\sum_{i=1}^{N}\delta_{d_i} \longrightarrow$$

### Ambiguity Set

$$\mathcal{P}_N = \left\{ W_p(\hat{\mathbf{P}}^{N}, \mathbf{P}) \leq \epsilon \right\}$$

Wasserstein distance

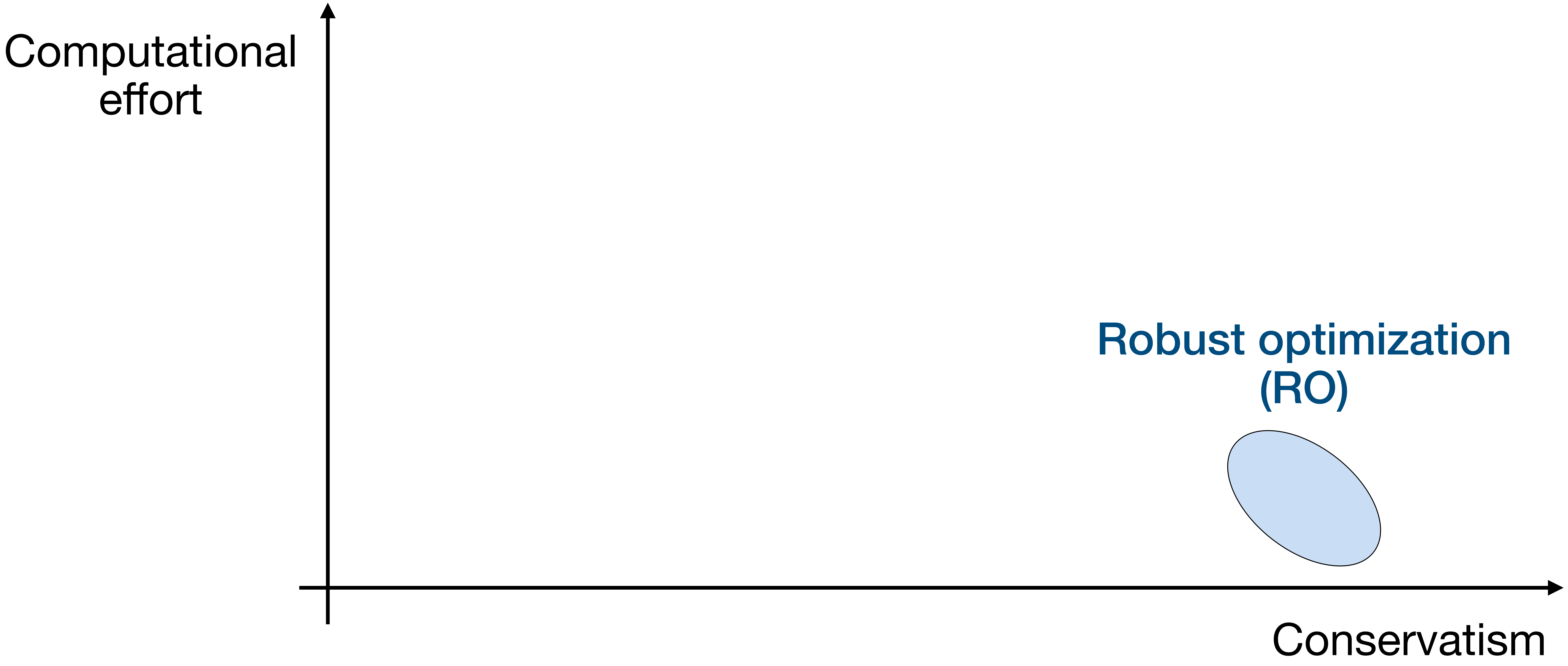


---

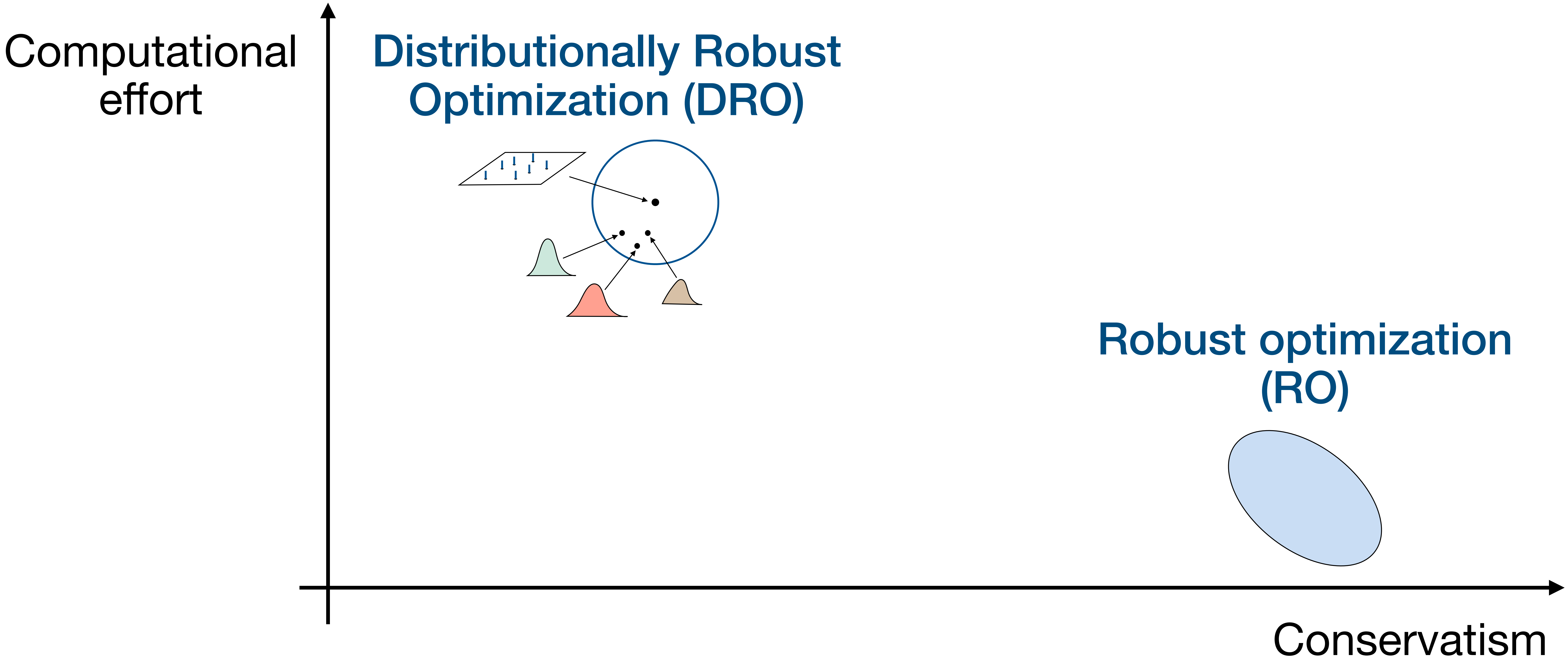
**Data-driven Wasserstein DRO.** D. Kuhn, P. Mohajerin Esfahani, B. Van Parys, R. Gao, S. Shafieezadeh Abadeh, W. Wiesemann, F. Dörfler, L. Aolaritei, A. Selvi….
**Optimal Transport.** G. Monge, L. Kantorovic, L. Wasserstein, C. Villani, …

# Robust vs Distributionally Robust Optimization

# Robust vs Distributionally Robust Optimization



# Robust vs Distributionally Robust Optimization



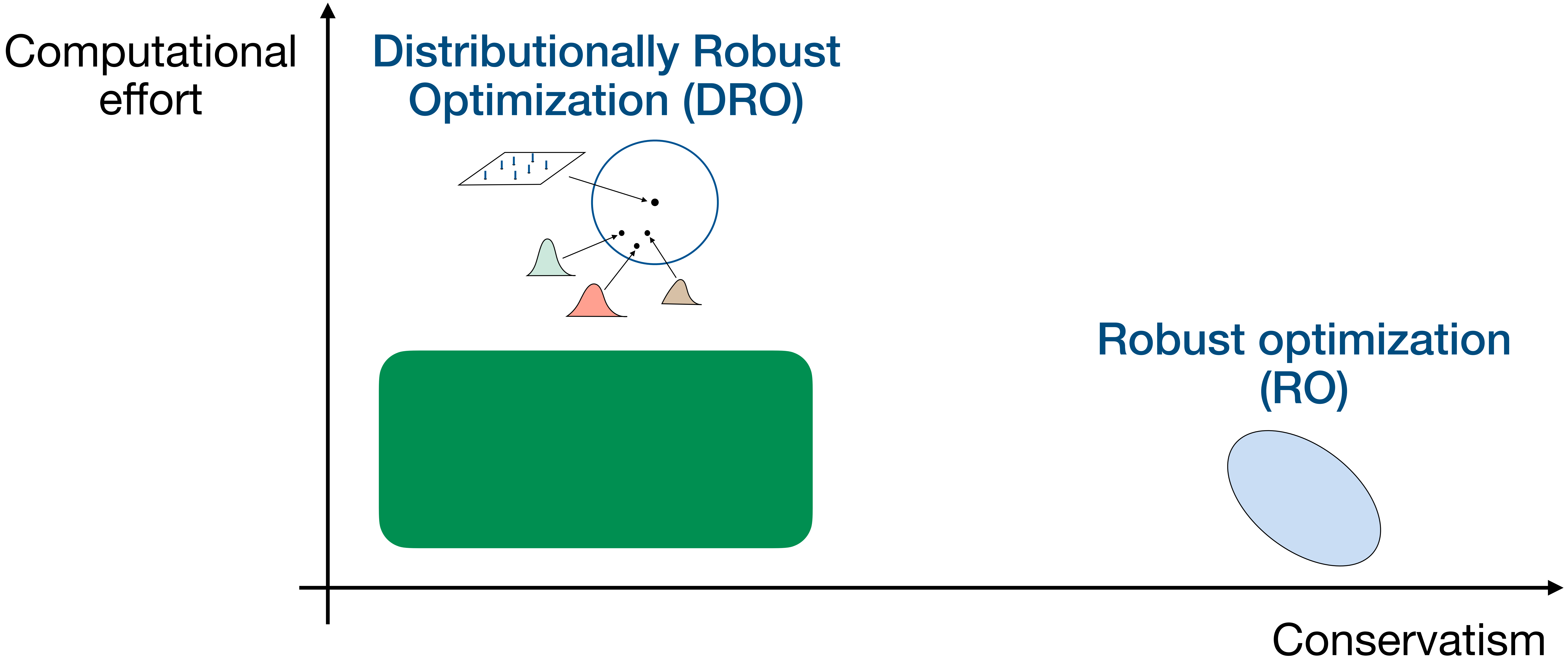
Computational effort

Distributionally Robust Optimization (DRO)

Robust optimization (RO)

Conservatism

# Robust vs Distributionally Robust Optimization



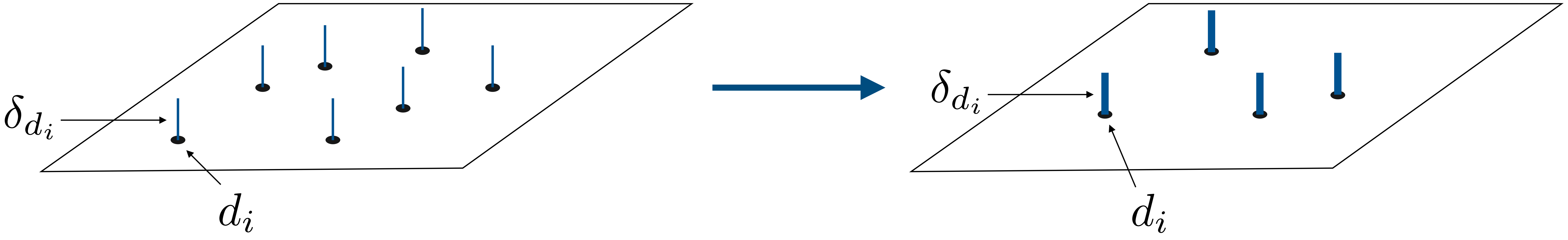
Can we get the best of both worlds?

# Mean Robust Optimization

# Data compression for decision-making under uncertainty

## scenario reduction

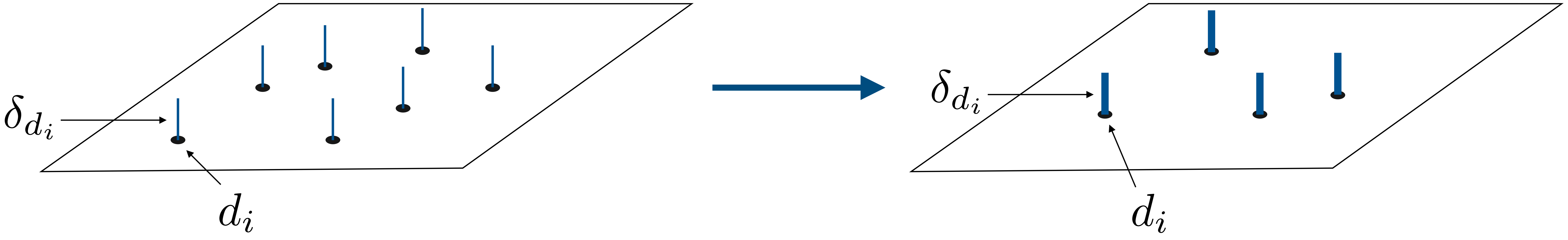
Dupačová et al. (2003), Beraldi and Bruni (2014),
Chen (2015), Emelogu et al. (2016), Jacobson et al. (2021)



# Data compression for decision-making under uncertainty

## scenario reduction

Dupačová et al. (2003), Beraldi and Bruni (2014),
Chen (2015), Emelogu et al. (2016), Jacobson et al. (2021)



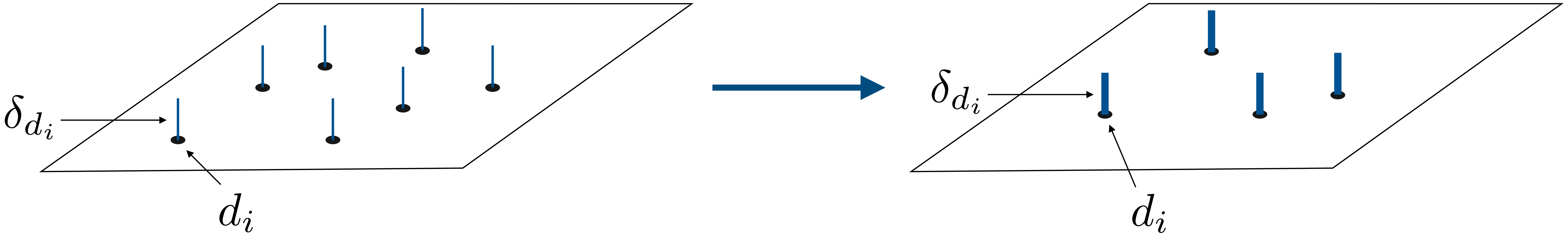
**approximation guarantees**

Rujeerapaiboon et al. (2022)

# Data compression for decision-making under uncertainty

## scenario reduction

Dupačová et al. (2003), Beraldi and Bruni (2014),
Chen (2015), Emelogu et al. (2016), Jacobson et al. (2021)



**approximation guarantees**

Rujeerapaiboon et al. (2022)

## dynamic problems

**two-stage optimization**

Bertsimas and Mundru (2022)
Wang et al. (2023)

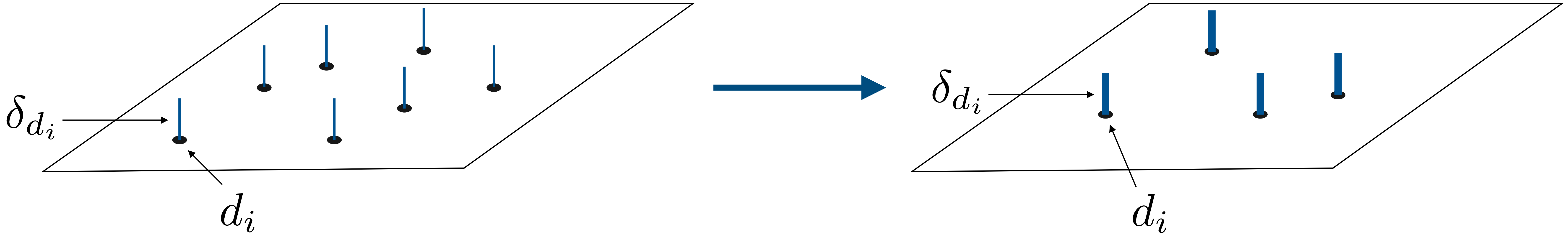
**optimal control**

Fabiani and Goulart (2021)

# Data compression for decision-making under uncertainty

## scenario reduction

Dupačová et al. (2003), Beraldi and Bruni (2014),
Chen (2015), Emelogu et al. (2016), Jacobson et al. (2021)



**approximation guarantees**
Rujeerapaiboon et al. (2022)

## dynamic problems

**two-stage optimization**
Bertsimas and Mundru (2022)
Wang et al. (2023)

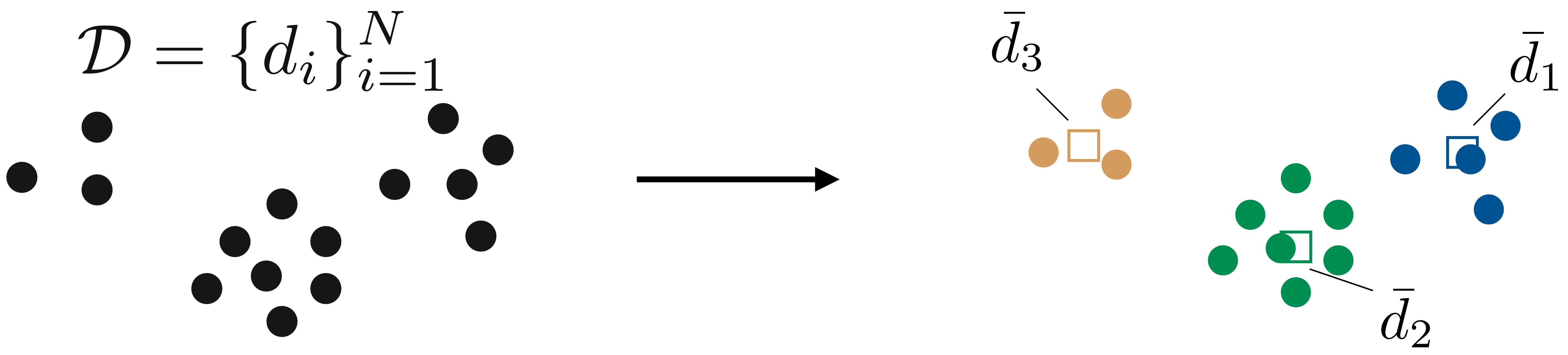
**optimal control**
Fabiani and Goulart (2021)

→ **tied to a specific clustering technique**

# We use clustering to reduce dimensionality and computation time

Data

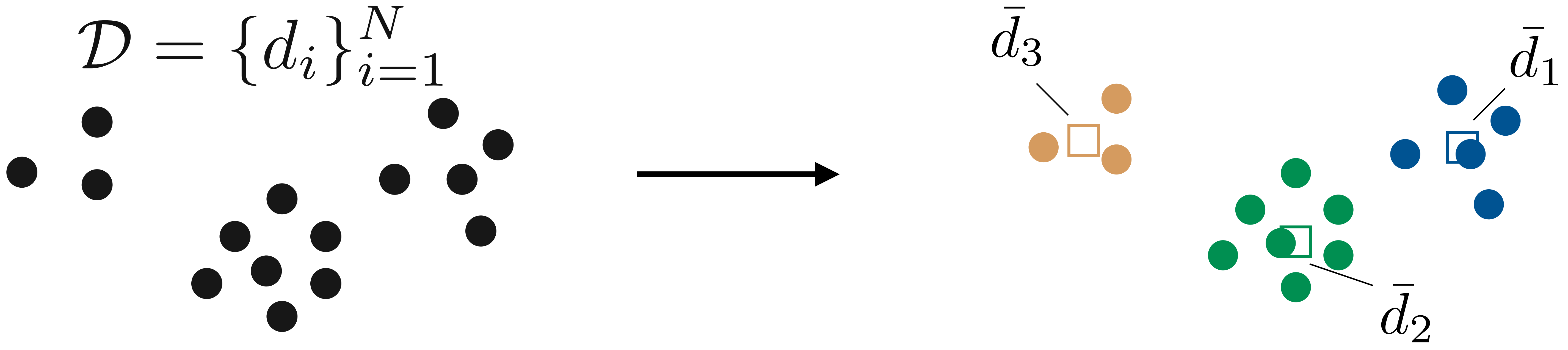
$\mathcal{D} = \{d_i\}_{i=1}^{N}$



# We use clustering to reduce dimensionality and computation time

Data

$\mathcal{D} = \{d_i\}_{i=1}^N$



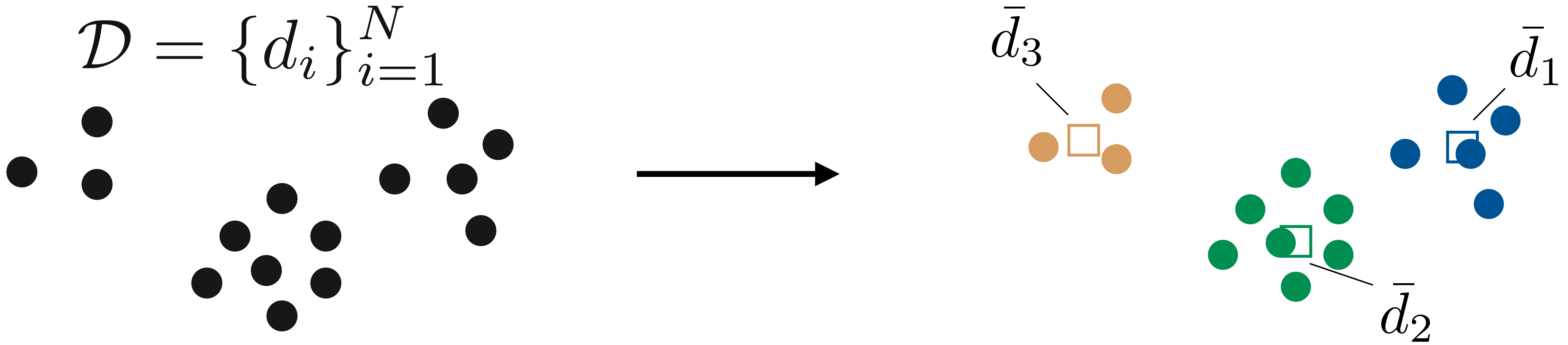
minimize $$\sum_{i=1}^{N} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^2$$

Jain et al. (1999)

# We use clustering to reduce dimensionality and computation time

Data

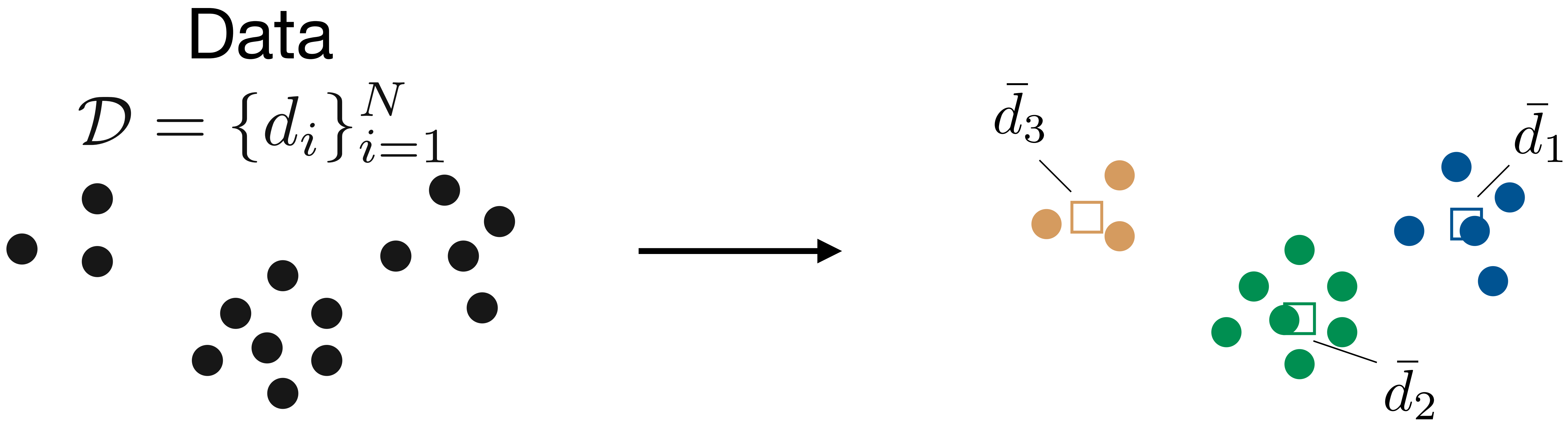
$\mathcal{D} = \{d_i\}_{i=1}^{N}$



minimize $$\sum_{i=1}^{N}\sum_{k=1}^{K} \underbrace{a_{ik}}_{\text{cluster assignments}} \|d_i - \bar{d}_k\|^2$$

Jain et al. (1999)

# We use clustering to reduce dimensionality and computation time


Data

$$\mathcal{D} = \{d_i\}_{i=1}^{N}$$

$$\text{minimize} \quad \sum_{i=1}^{N} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^2$$

cluster assignments

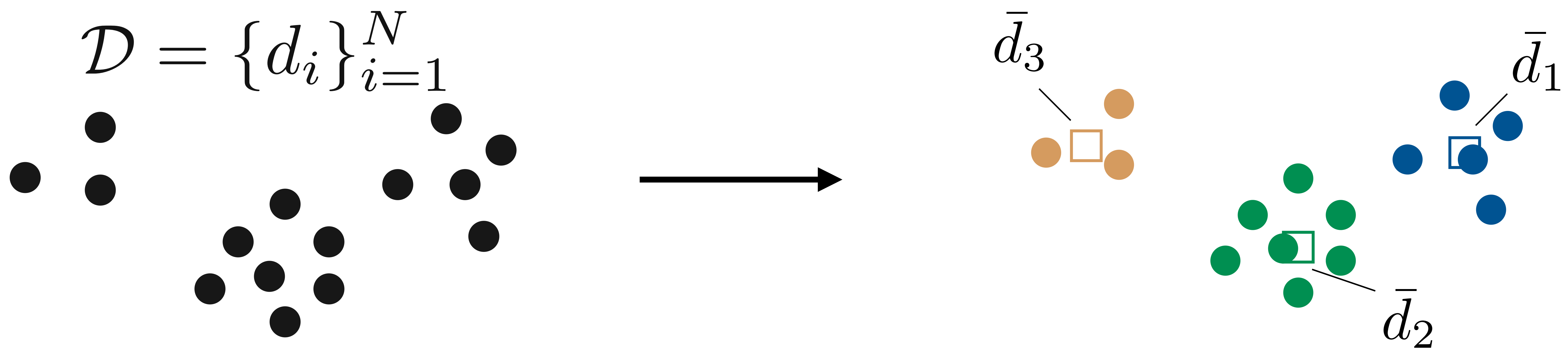
cluster centers

Jain et al. (1999)

# We use clustering to reduce dimensionality and computation time

Data

$\mathcal{D} = \{d_i\}_{i=1}^{N}$



minimize $$\sum_{i=1}^{N}\sum_{k=1}^{K} a_{ik}\|d_i - \bar{d}_k\|^2$$

cluster assignments

cluster centers

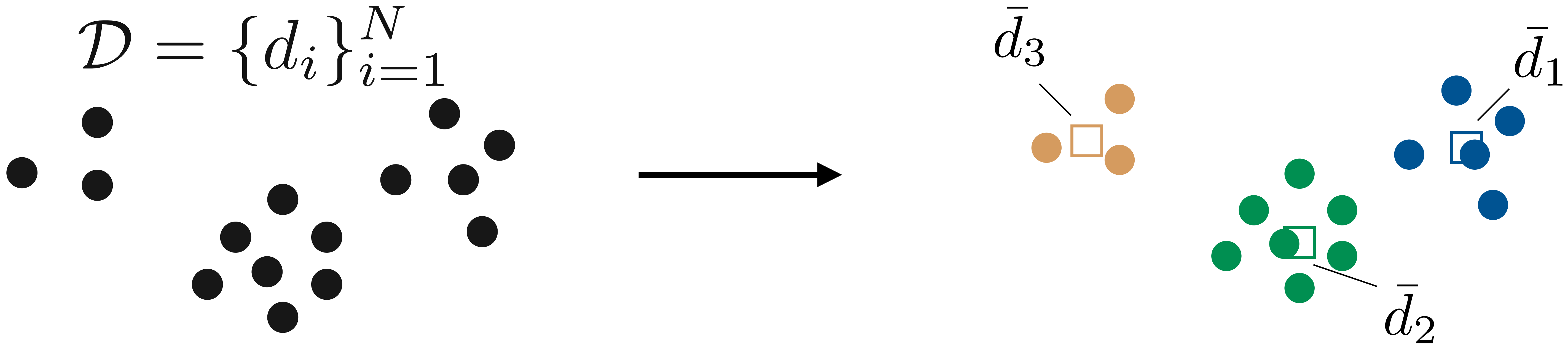
*we work with any clustering algorithm*

Jain et al. (1999)

# We use clustering to reduce dimensionality and computation time

Data

$\mathcal{D} = \{d_i\}_{i=1}^{N}$



minimize $$\sum_{i=1}^{N}\sum_{k=1}^{K} a_{ik}\|d_i - \bar{d}_k\|^2$$

cluster assignments

cluster centers

*we work with any clustering algorithm*

## Main idea

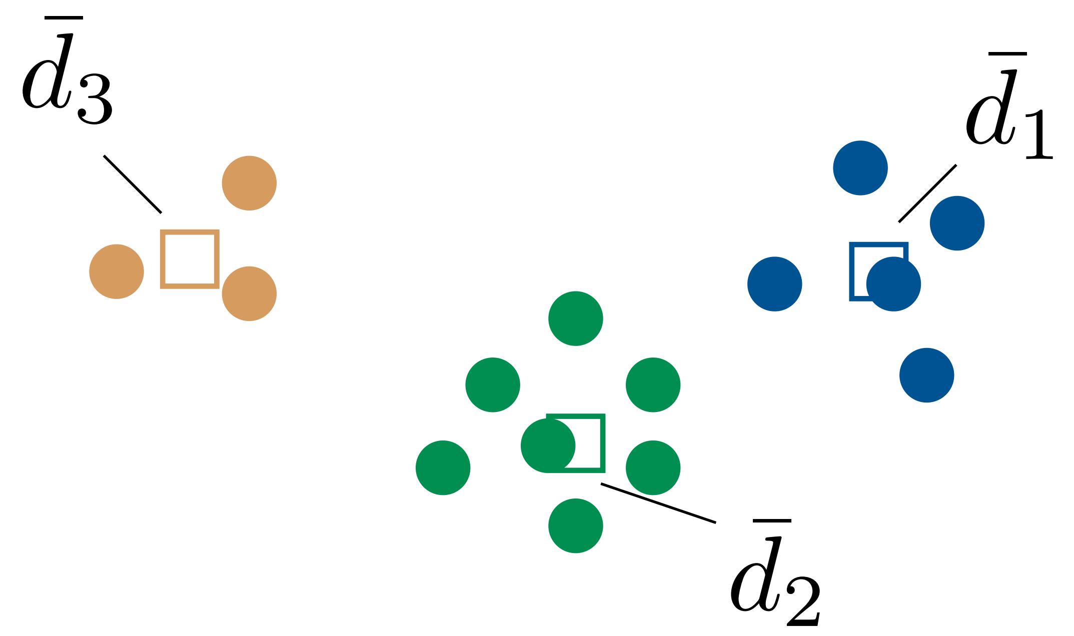
use cluster centers
instead of original data

Jain et al. (1999)

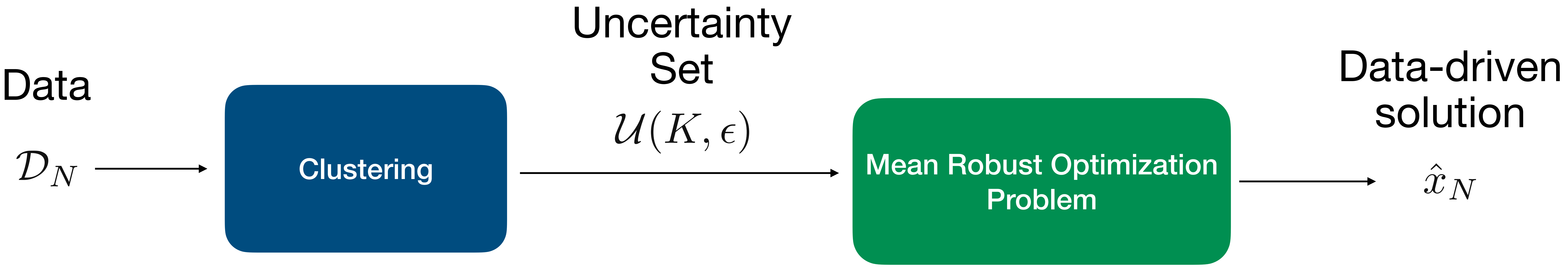
# Our approach: Mean Robust Optimization (MRO)

Data

$\mathcal{D}_N \longrightarrow$ **Clustering**



# Our approach: Mean Robust Optimization (MRO)



Data

$\mathcal{D}_N$

Clustering

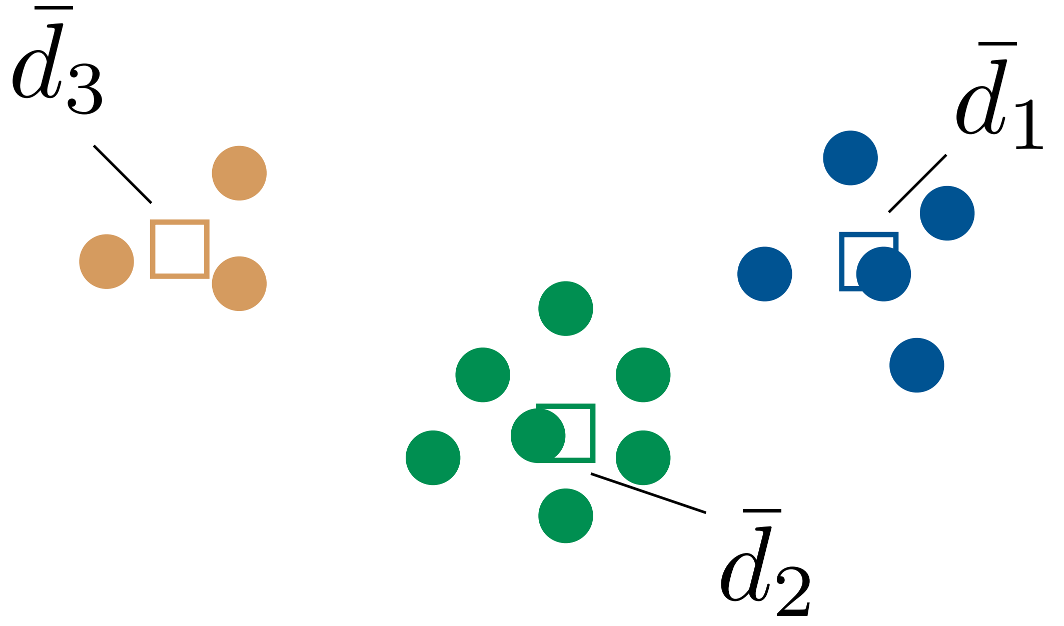
Uncertainty Set

$\mathcal{U}(K, \epsilon)$

Mean Robust Optimization Problem

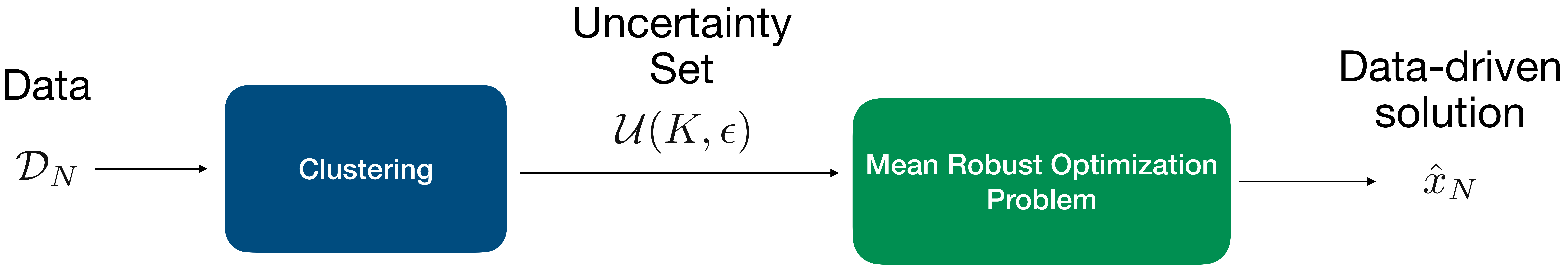
Data-driven solution

$\hat{x}_N$



$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \bar{g}(x, u) \leq 0 \quad \forall u \in \mathcal{U}(K, \epsilon) \end{array}$$

# Our approach: Mean Robust Optimization (MRO)



Data

$\mathcal{D}_N$

Clustering

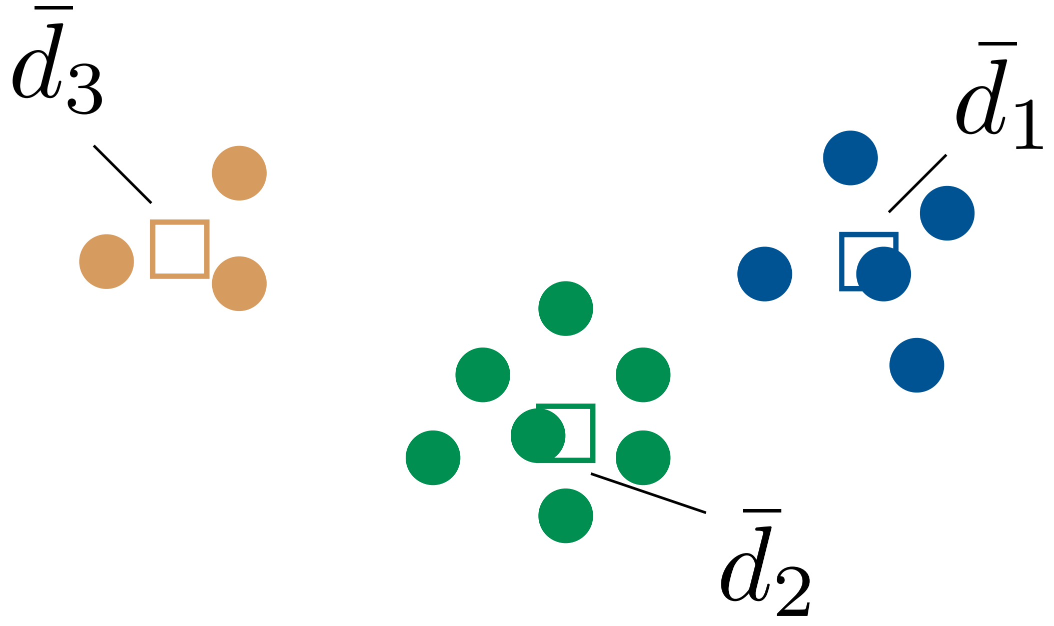
Uncertainty Set

$\mathcal{U}(K, \epsilon)$

Mean Robust Optimization Problem

Data-driven solution

$\hat{x}_N$

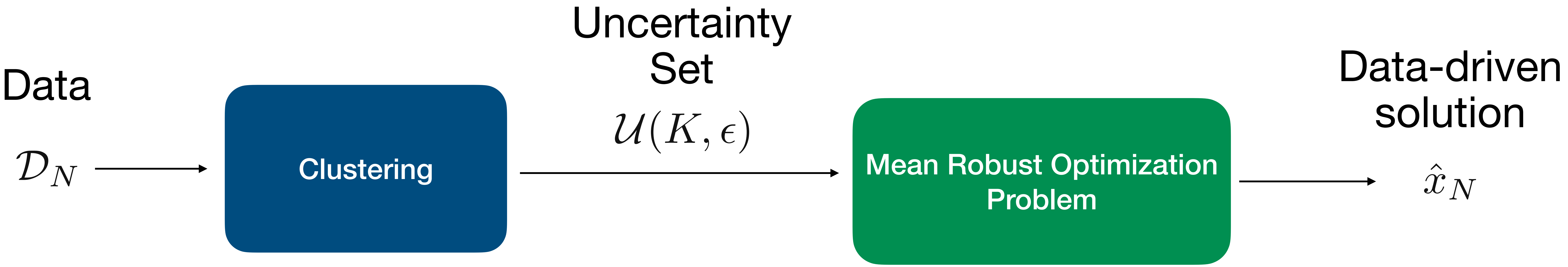


$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \bar{g}(x,u) \leq 0 \quad \forall u \in \mathcal{U}(K, \epsilon) \end{array}$$

$\uparrow$

constraint function

# Our approach: Mean Robust Optimization (MRO)



Data

$\mathcal{D}_N$

Clustering

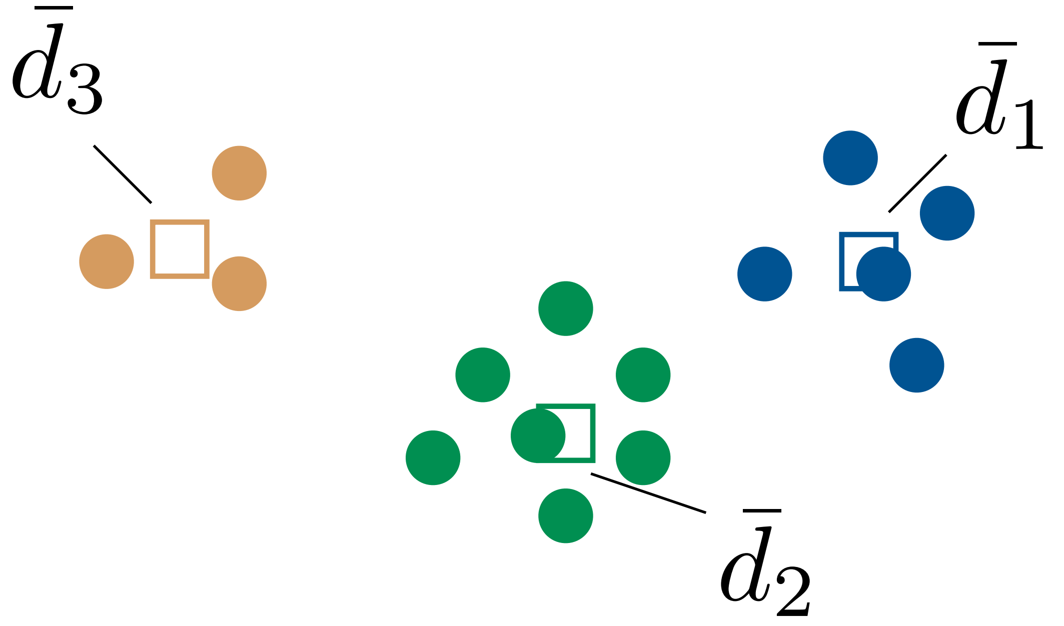
Uncertainty Set

$\mathcal{U}(K, \epsilon)$

Mean Robust Optimization Problem

Data-driven solution

$\hat{x}_N$

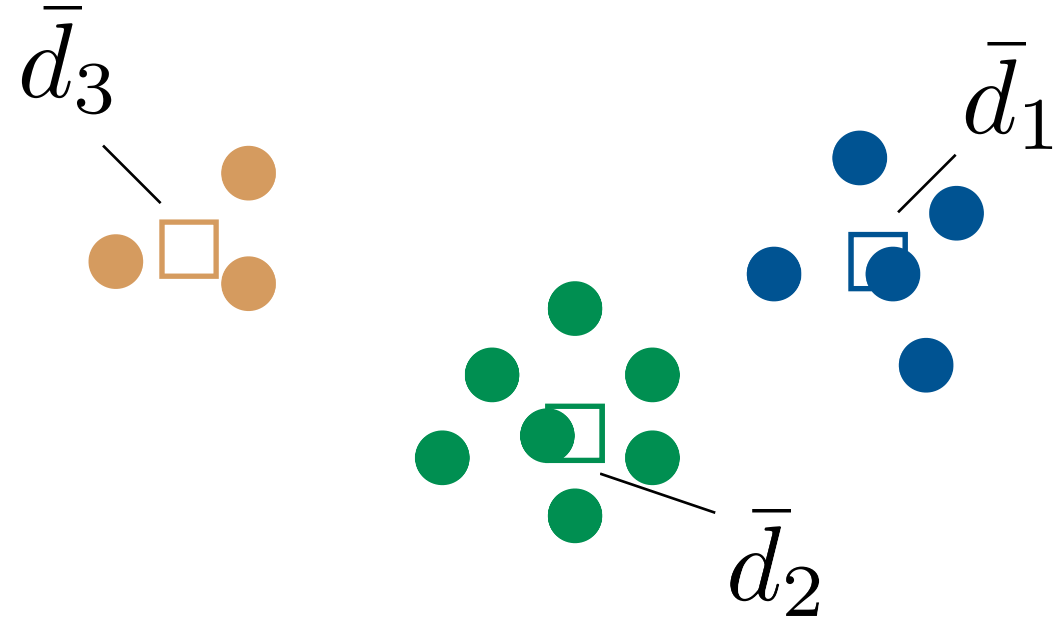


$$\begin{array}{ll}\text{minimize} & f(x)\\ \text{subject to} & \bar{g}(x, u) \leq 0 \quad \forall u \in \mathcal{U}(K, \epsilon)\end{array}$$

constraint function

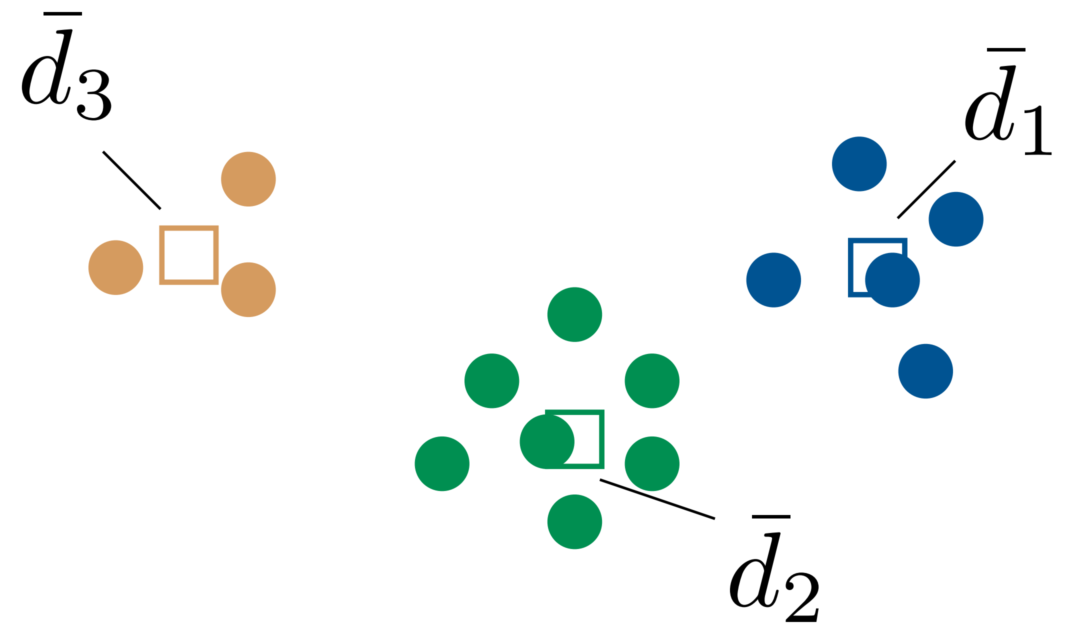
uncertainty set

# Uncertainty set



$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1,\ldots,v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \le \epsilon^p \right\}$$

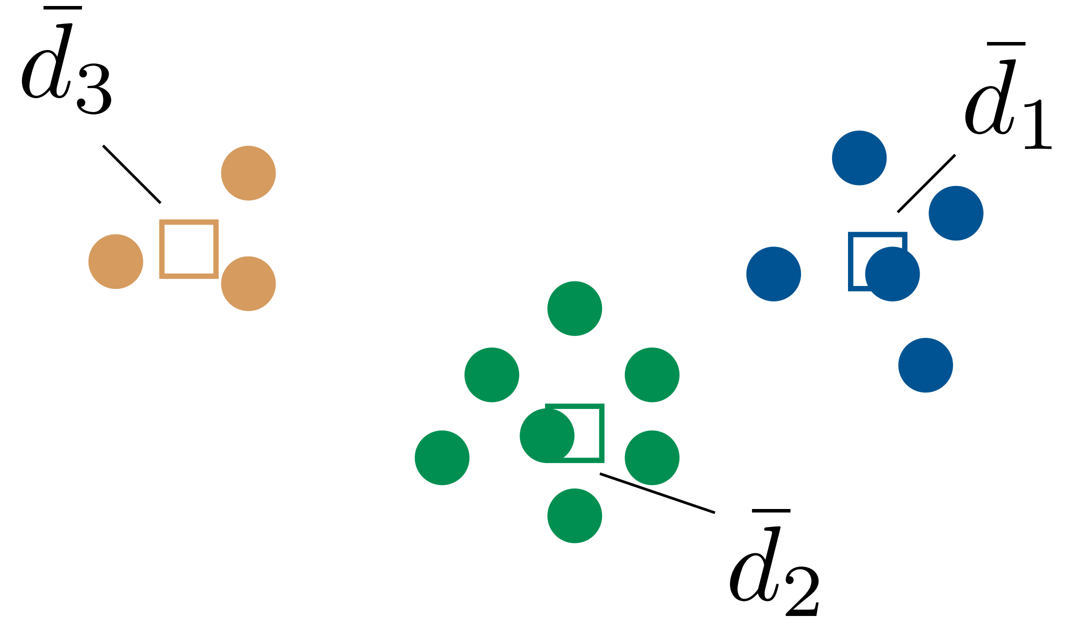
# Uncertainty set



cluster weights

$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1, \ldots, v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \leq \epsilon^p \right\}$$

# Uncertainty set

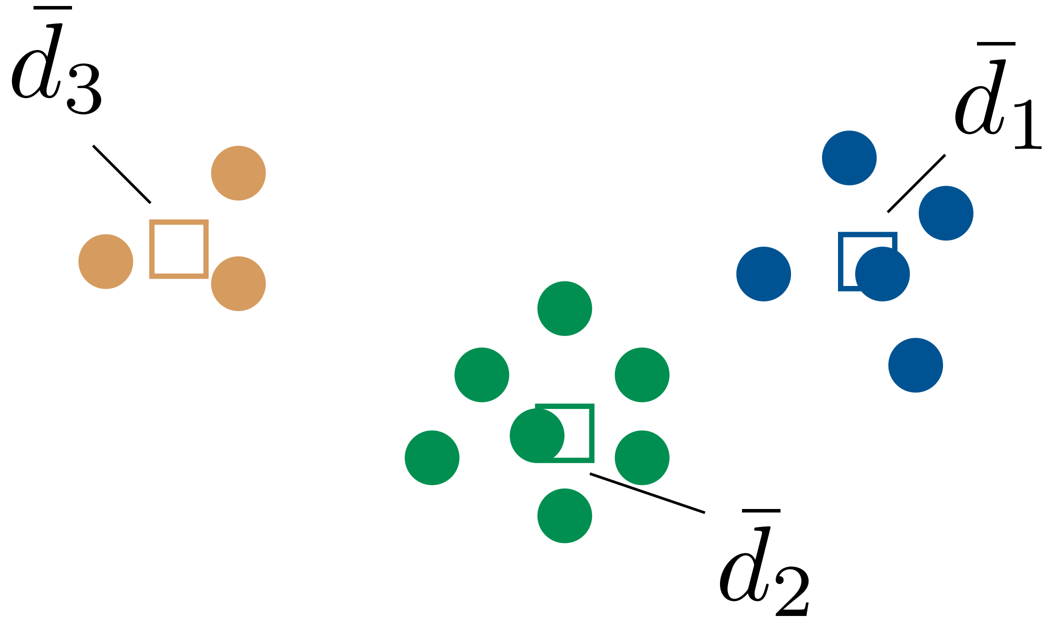


cluster weights

$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1,\dots,v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \le \epsilon^p \right\}$$

cluster centers

# Uncertainty set



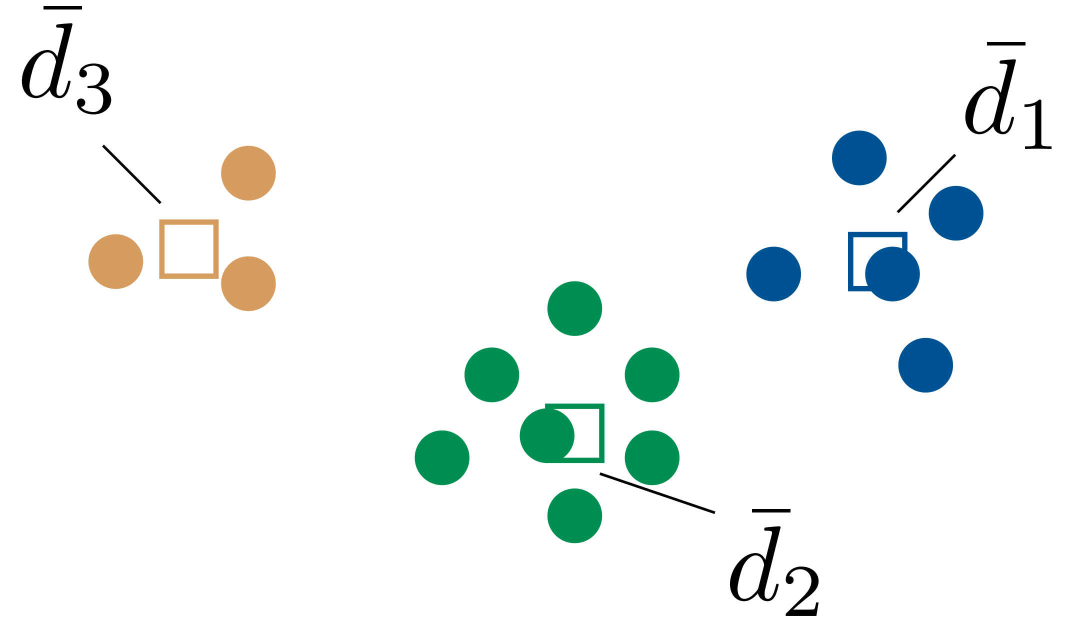
cluster weights

order

$$\mathcal{U}(K, \epsilon) = \left\{ u = (v_1, \dots, v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \le \epsilon^p \right\}$$

cluster centers

# Uncertainty set



cluster weights

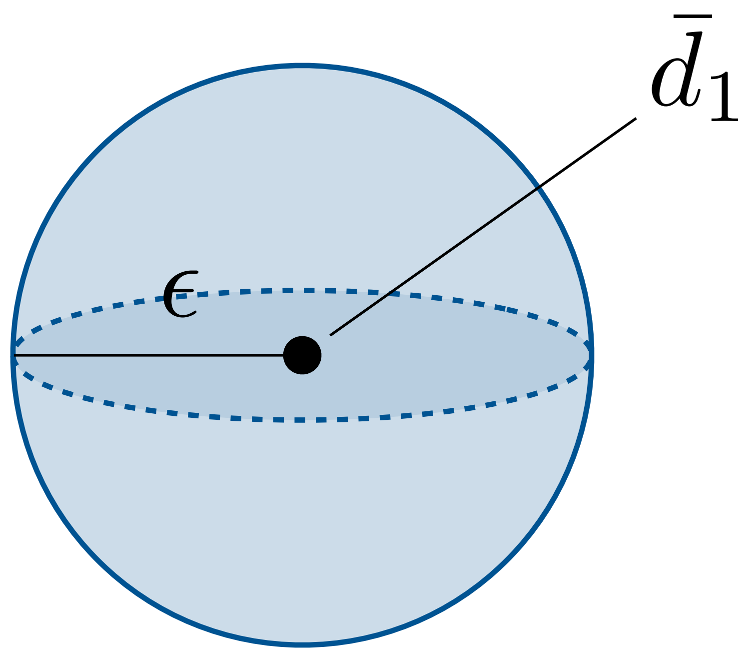
order

$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1,\dots,v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \le \epsilon^p \right\}$$

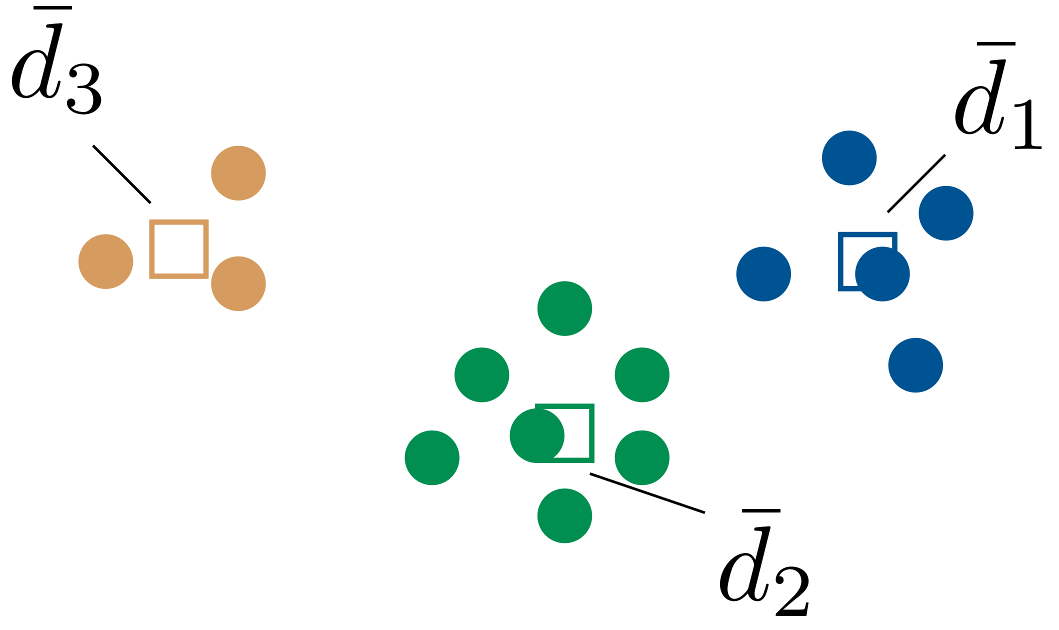
cluster centers

**Examples**

$K = 1$



# Uncertainty set



cluster weights

order

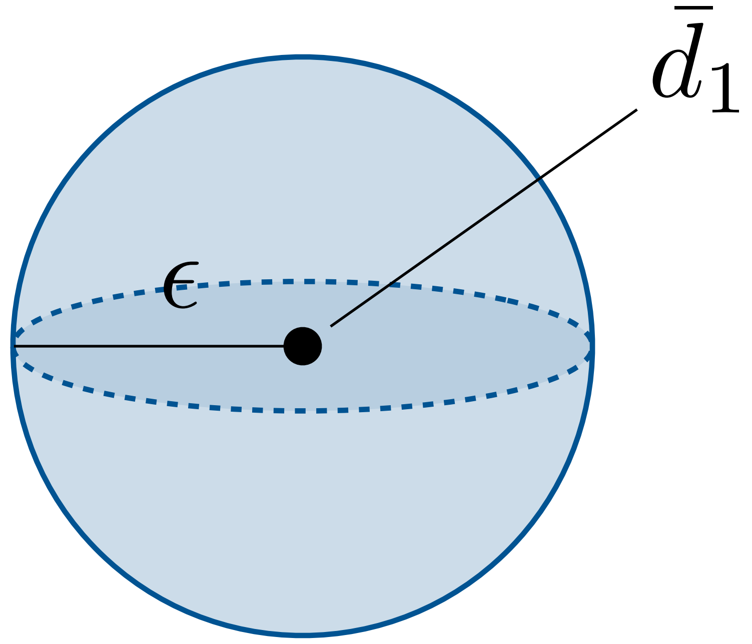
$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1,\ldots,v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \leq \epsilon^p \right\}$$

cluster centers

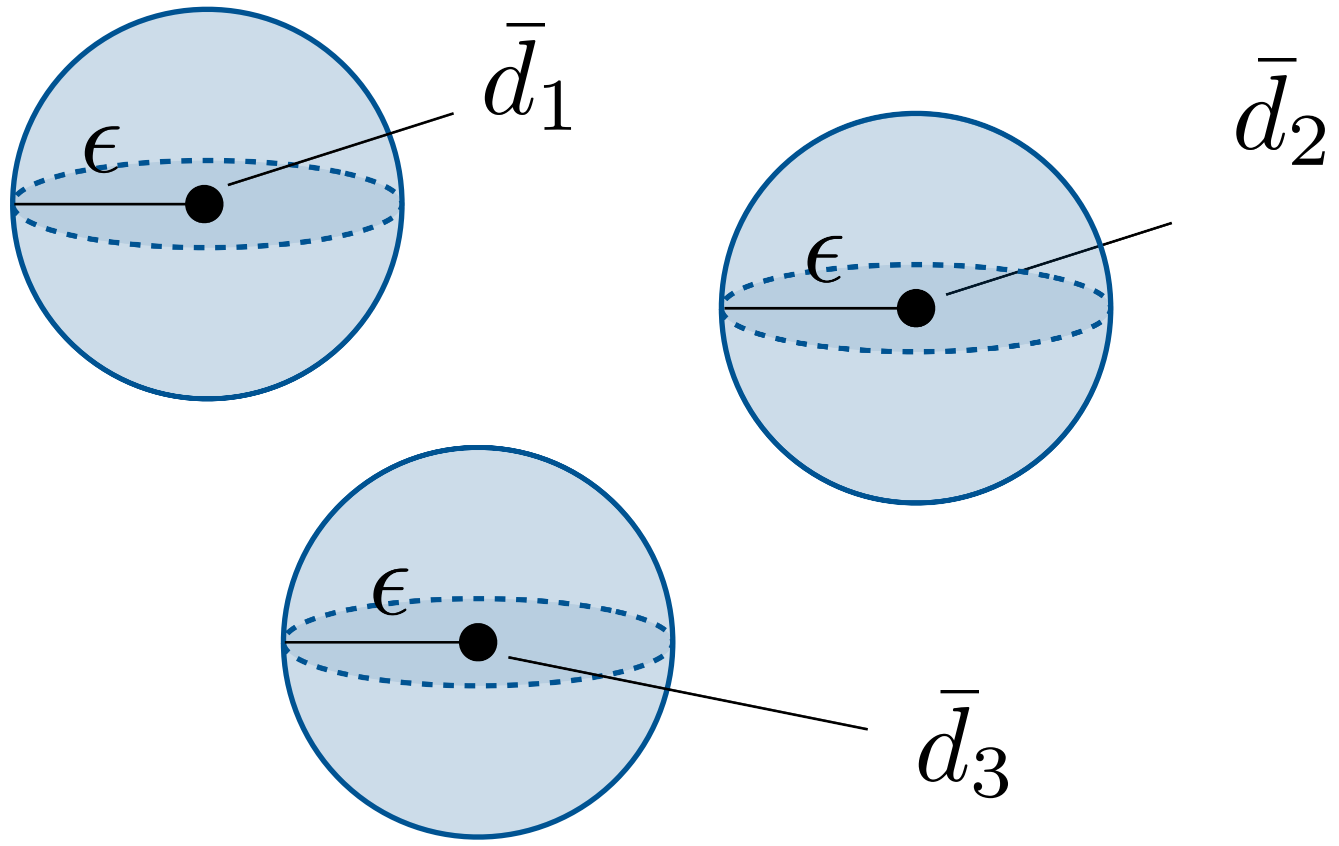
**Examples**

$K$

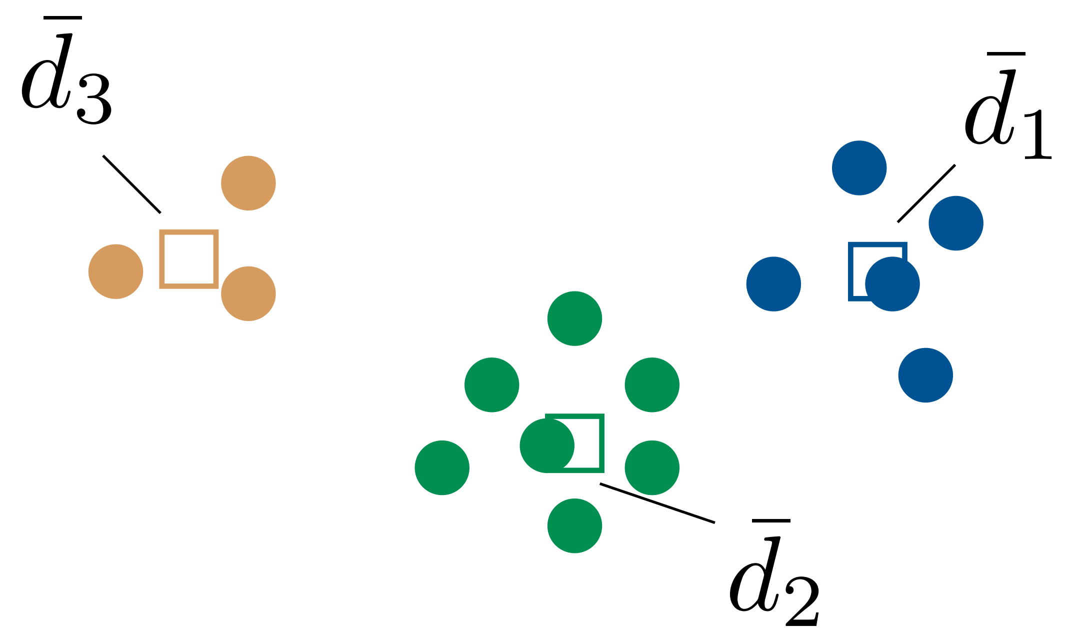
$K = 1$



$K = 3, p = \infty$



# Uncertainty set



cluster weights

order

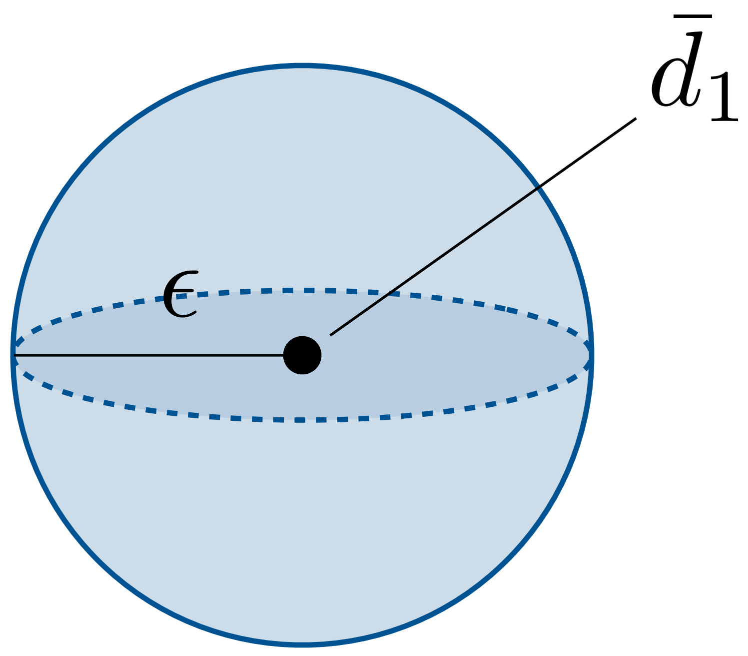
$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1, \dots, v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \leq \epsilon^p \right\}$$

cluster centers

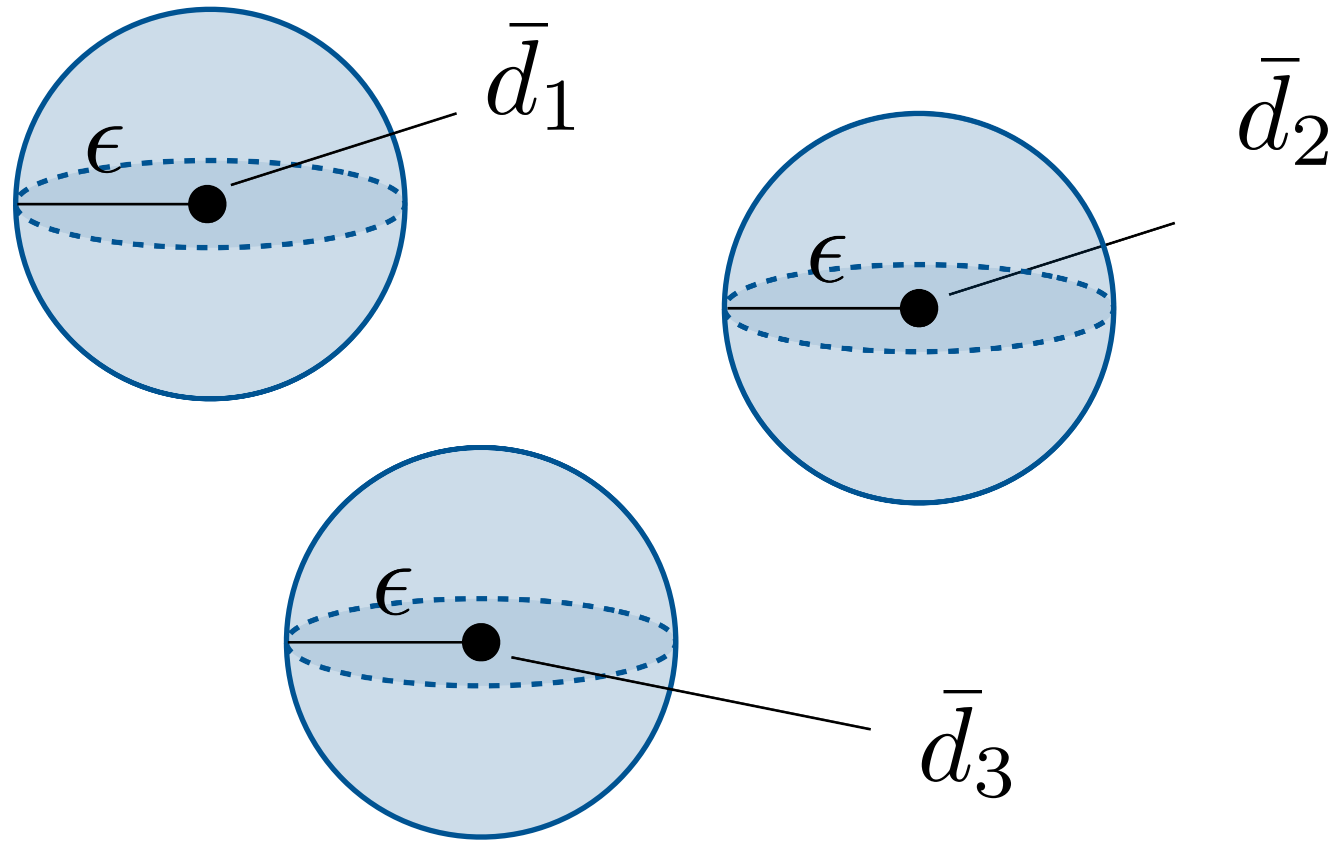
## Examples

$K$

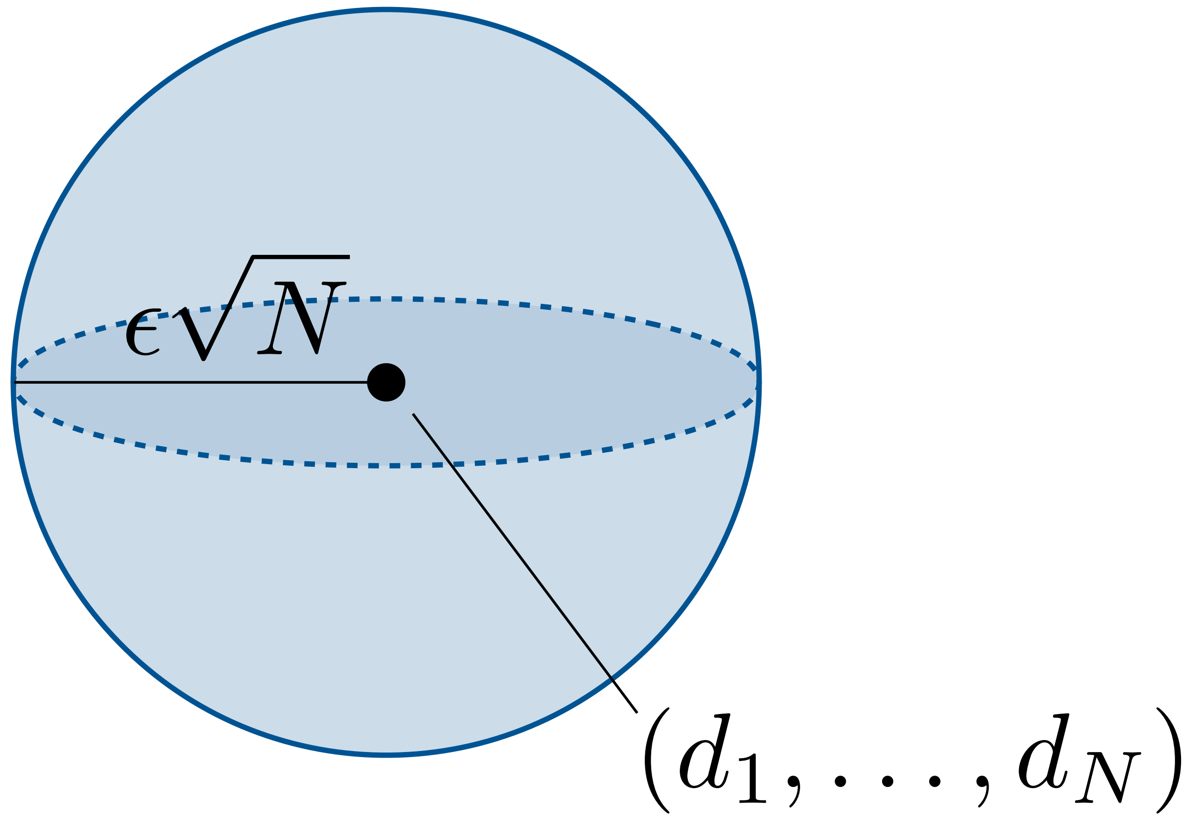
$K = 1$



$K = 3, p = \infty$



$K = N, p = 2$



# Back to the Mean Robust Optimization problem

Uncertain variable lifting

$u = (v_1, \dots, v_K)$

$$\longrightarrow$$

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \bar{g}(x,u) \le 0 \quad \forall u \in \mathcal{U}(K,\epsilon) \end{array}$$

# Back to the Mean Robust Optimization problem

uncertainty set

$$\left\{ \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \leq \epsilon^p \right\}$$

Uncertain variable lifting

$$u = (v_1, \ldots, v_K) \longrightarrow \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \bar{g}(x, u) \leq 0 \quad \forall u \in \mathcal{U}(K, \epsilon) \end{array}$$

# Back to the Mean Robust Optimization problem

Uncertain variable lifting

$u = (v_1, \dots, v_K)$

$\longrightarrow$

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{subject to} & \bar{g}(x, u) \le 0 \quad \forall u \in \mathcal{U}(K, \epsilon)
\end{array}
$$

uncertainty set

$$
\left\{ \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \le \epsilon^p \right\}
$$

constraint function

$$
\sum_{k=1}^{K} w_k g(x, v_k)
$$

# Solving the MRO problem

dualize constraint

$$\bar{g}(x,u) \leq 0, \ \forall u \in \mathcal{U}(K,\epsilon)$$

# Solving the MRO problem

dualize constraint

$$\bar{g}(x, u) \leq 0, \; \forall u \in \mathcal{U}(K, \epsilon)$$

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \sum_{k=1}^{K} w_k s_k \leq 0 \\ & [-g]^*(x, z_k) - z_k^T \bar{d}_k + \phi(p)\lambda \|z_k/\lambda\|_*^{p/(p-1)} + \lambda \epsilon^p \leq s_k, \quad k = 1, \dots, K \\ & \lambda \geq 0 \end{array}$$

conjugate function

cluster centers

function of $p \geq 1$
$\phi(p) \rightarrow 1$ as $p \rightarrow \infty$
$\phi(1) = 0$

# Solving the MRO problem

dualize constraint

$$\bar{g}(x,u) \leq 0,\ \forall u \in \mathcal{U}(K,\epsilon)$$

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \sum_{k=1}^{K} w_k s_k \leq 0 \\ & [-g]^*(x,z_k) - z_k^T \bar{d}_k + \phi(p)\lambda\|z_k/\lambda\|_*^{p/(p-1)} + \lambda\epsilon^p \leq s_k, \quad k=1,\dots,K \\ & \lambda \geq 0 \end{array}$$

conjugate function

cluster centers

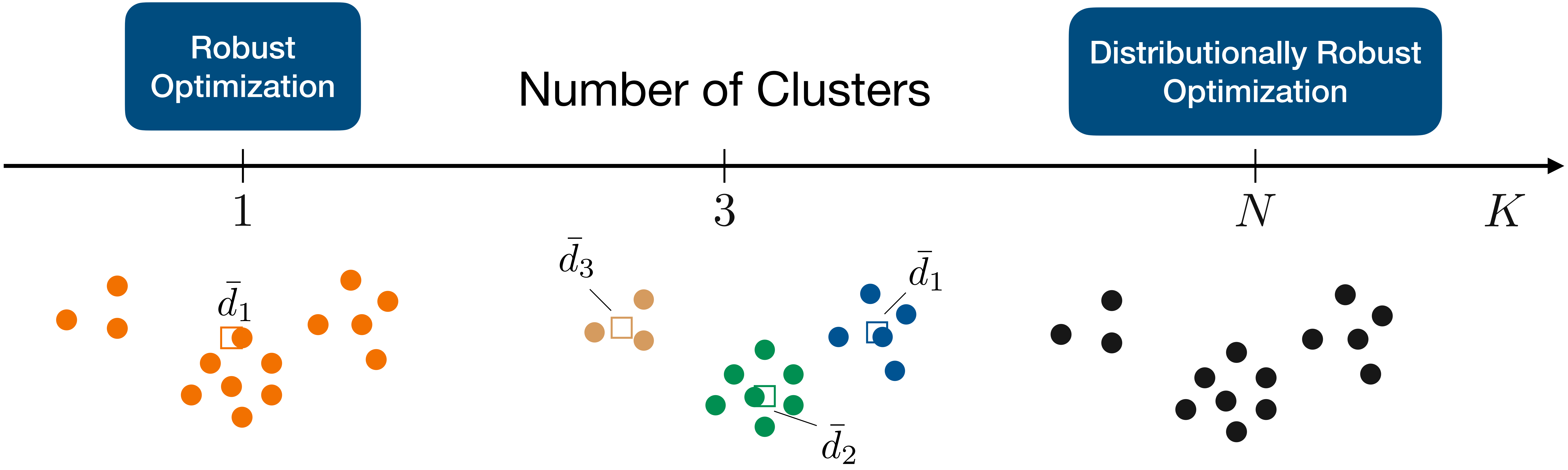
function of $p \geq 1$

$\phi(p) \to 1$ as $p \to \infty$

$\phi(1) = 0$

It can be very expensive when $K$ is large (e.g., $K = N$)

# MRO bridges RO and DRO



**Robust Optimization**

Number of Clusters

**Distributionally Robust Optimization**

Kuhn, Mohajerin Esfahani, Nguyen, Shafieezadeh-Abadeh (2019)

# Probabilistic guarantees in MRO

$$\mathbf{P}^N\left(\mathbf{E}(g(\hat{x}_N, u)) \leq 0\right) \geq 1 - \beta$$

# Probabilistic guarantees in MRO

probability of
constraint
satisfaction

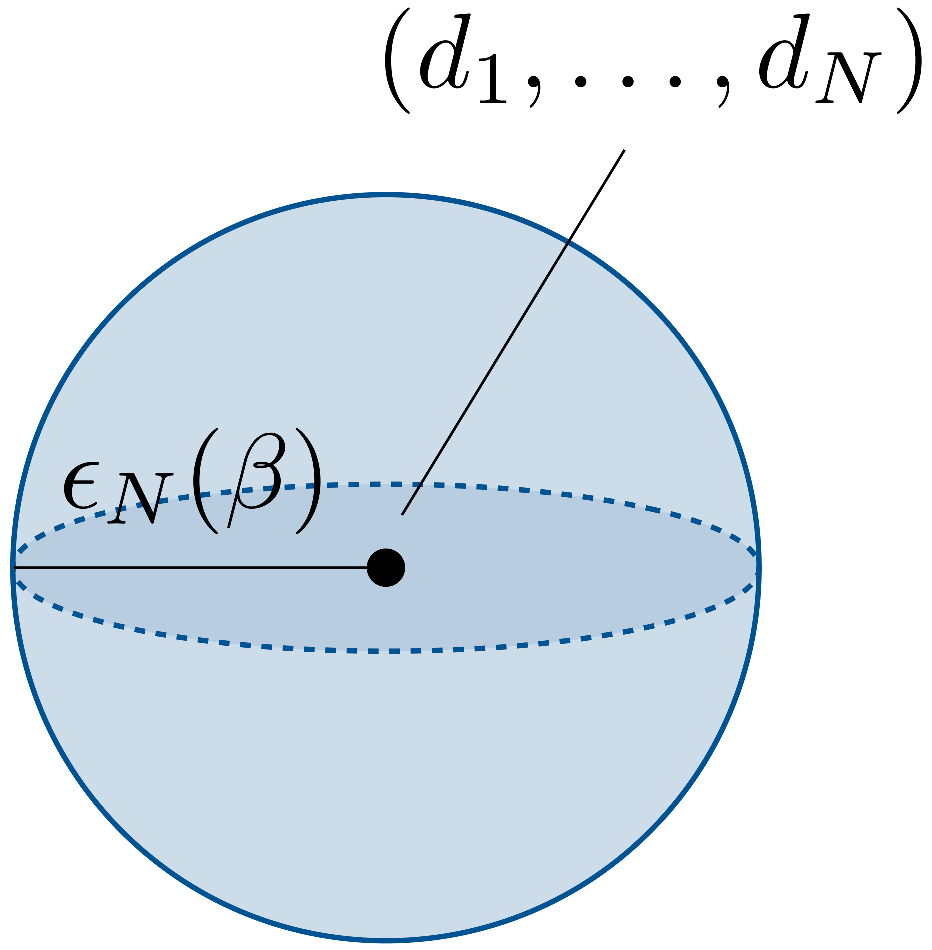
$$\mathbf{P}^N\left(\mathbf{E}(g(\hat{x}_N, u)) \leq 0\right) \geq \boxed{1-\beta}$$

# Probabilistic guarantees in MRO

probability of constraint satisfaction

uncertainty set radius

$$\mathbf{P}^N\left(\mathbf{E}(g(\hat{x}_N,u)) \le 0\right) \ge 1-\beta \quad \xrightarrow[\text{Fournier and Guillin (2013)}]{\text{light-tailed}} \quad \mathcal{U}(N,\epsilon_N(\beta))$$

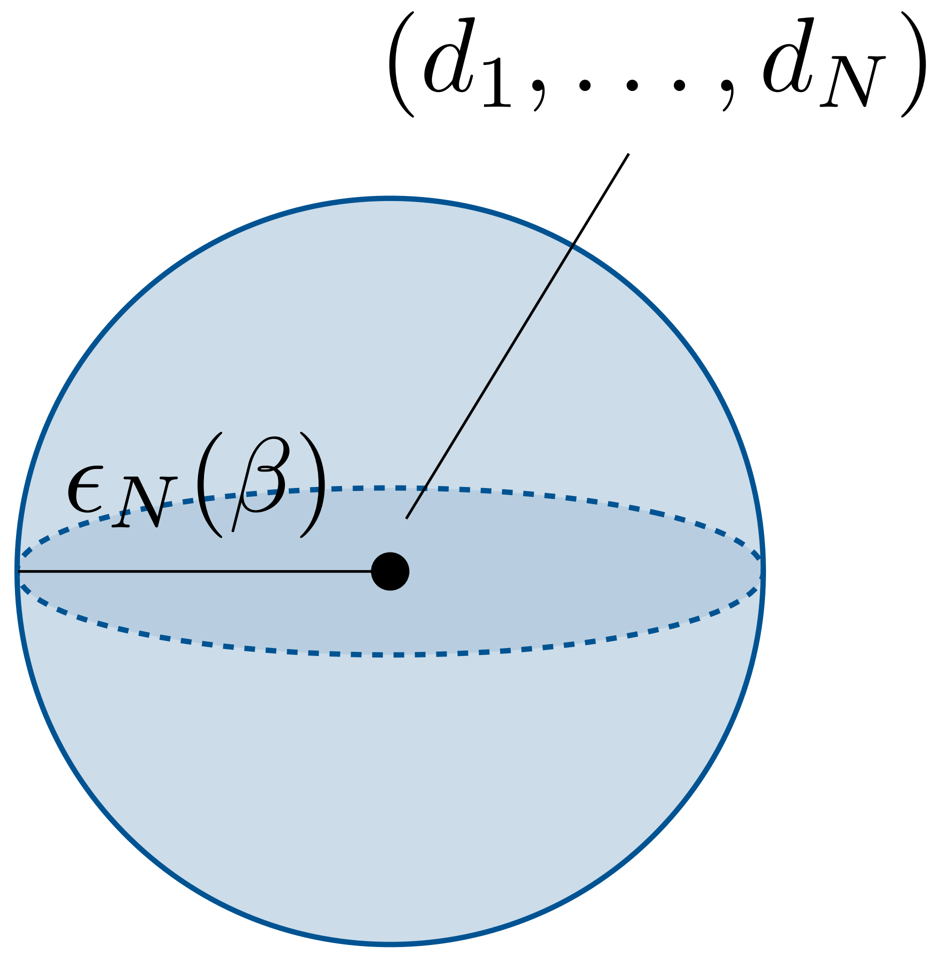


# Probabilistic guarantees in MRO

probability of constraint satisfaction

uncertainty set radius

$$\mathbf{P}^N\left(\mathbf{E}(g(\hat{x}_N,u))\le 0\right)\ge 1-\beta \xrightarrow[\text{Fournier and Guillin (2013)}]{\text{light-tailed}} \mathcal{U}(N,\epsilon_N(\beta))$$

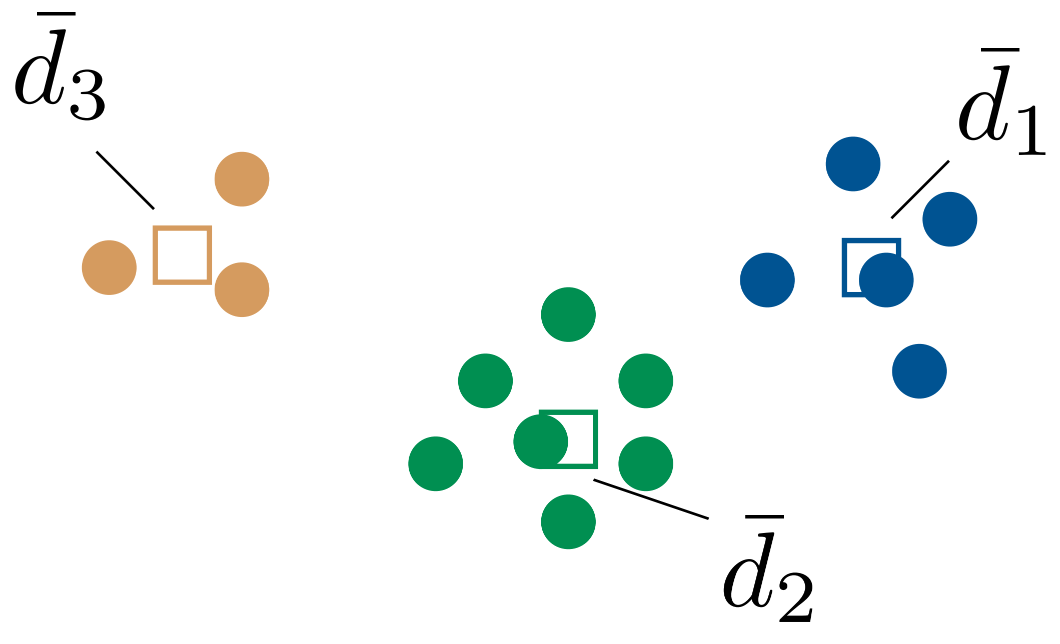


**MRO clustering**

$$\mathcal{U}(K,\epsilon_N(\beta)+\eta_K)$$

clustering error

$$\sum_{i=1}^{N}\sum_{k=1}^{K} a_{ik}\|d_i-\bar{d}_k\|^p$$

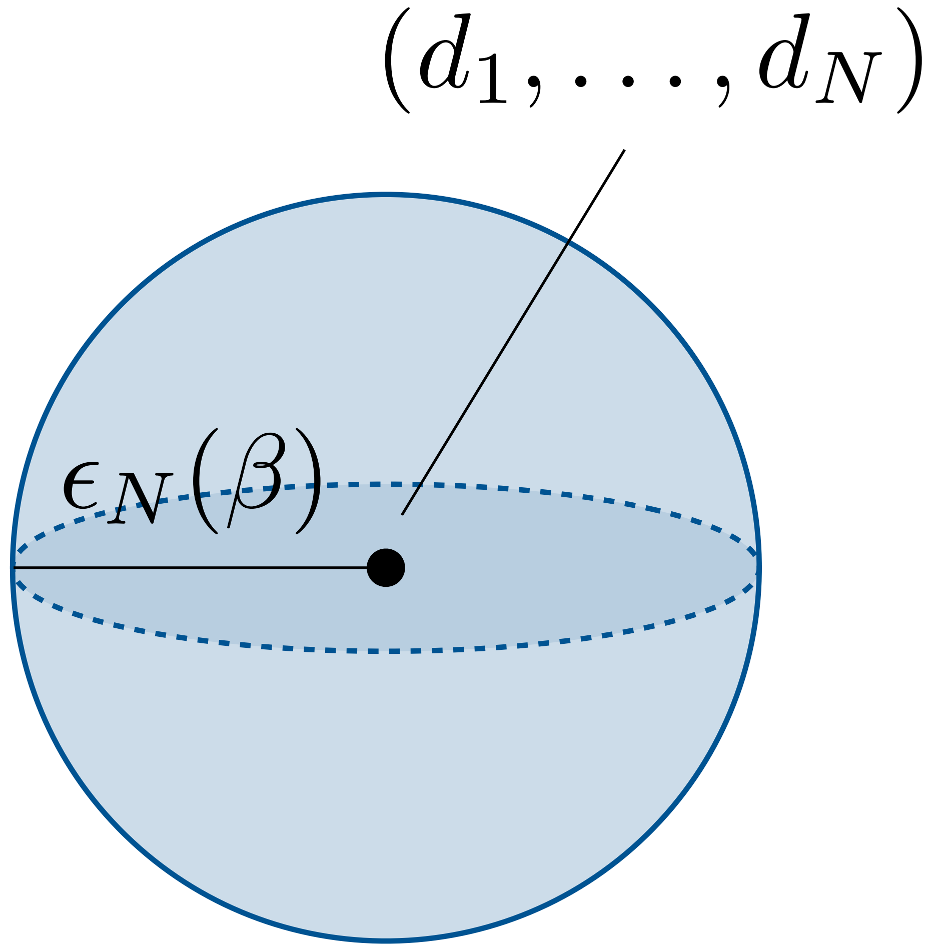


# Probabilistic guarantees in MRO

probability of constraint satisfaction

uncertainty set radius

$$\mathbf{P}^N\left(\mathbf{E}(g(\hat{x}_N, u)) \leq 0\right) \geq 1 - \beta \xrightarrow[\text{Fournier and Guillin (2013)}]{\text{light-tailed}} \mathcal{U}(N, \epsilon_N(\beta))$$

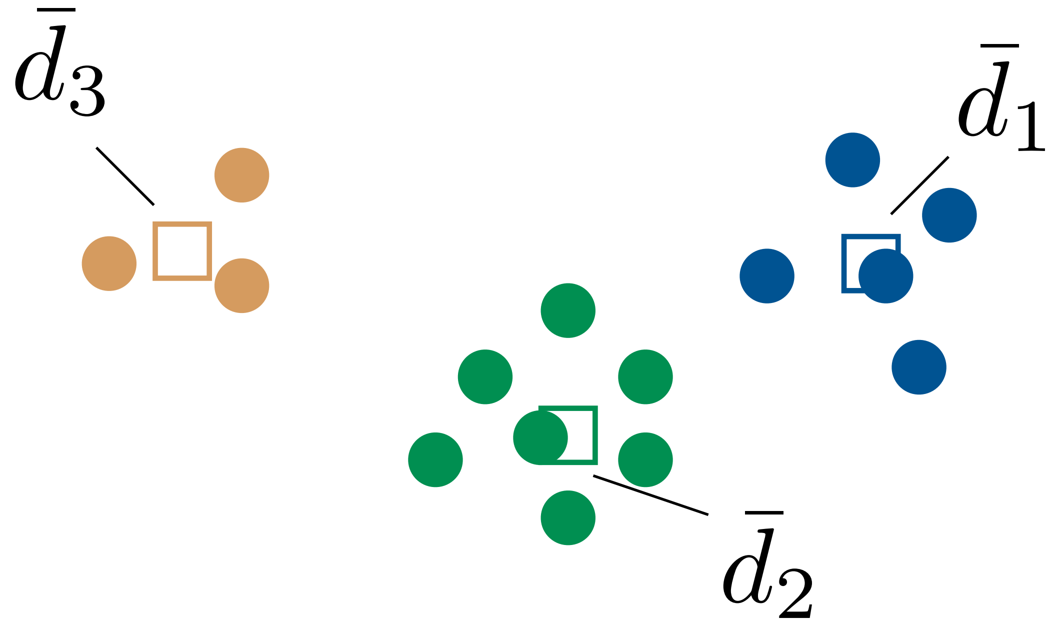


## MRO clustering

$$\mathcal{U}(K, \epsilon_N(\beta) + \eta_K)$$

clustering error

$$\sum_{i=1}^{N} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^p$$



Quite conservative bounds… can we do better?

# Bounding the conservatism

## MRO constraint

$$\bar{g}(x,u) \leq 0 \quad \forall u \in \mathcal{U}(K,\epsilon)$$

## Worst-case values

$$\bar{g}^{N}(x) = \underset{u \in \mathcal{U}(N,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

$$\bar{g}^{K}(x) = \underset{u \in \mathcal{U}(K,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

# Bounding the conservatism

## MRO constraint

$$\bar{g}(x,u) \leq 0 \quad \forall u \in \mathcal{U}(K,\epsilon)$$

## Worst-case values

$$\bar{g}^{N}(x) = \underset{u \in \mathcal{U}(N,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

$$\bar{g}^{K}(x) = \underset{u \in \mathcal{U}(K,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

### Theorem

If $-g$ is $L$-smooth in $u$, we have

$$\bar{g}^{N}(x) \leq \bar{g}^{K}(x) \leq \bar{g}^{N}(x) + \frac{L}{2} D(K)$$

# Bounding the conservatism

## MRO constraint

$$\bar{g}(x,u) \leq 0 \quad \forall u \in \mathcal{U}(K,\epsilon)$$

## Worst-case values

$$\bar{g}^{N}(x) = \underset{u\in\mathcal{U}(N,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

$$\bar{g}^{K}(x) = \underset{u\in\mathcal{U}(K,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

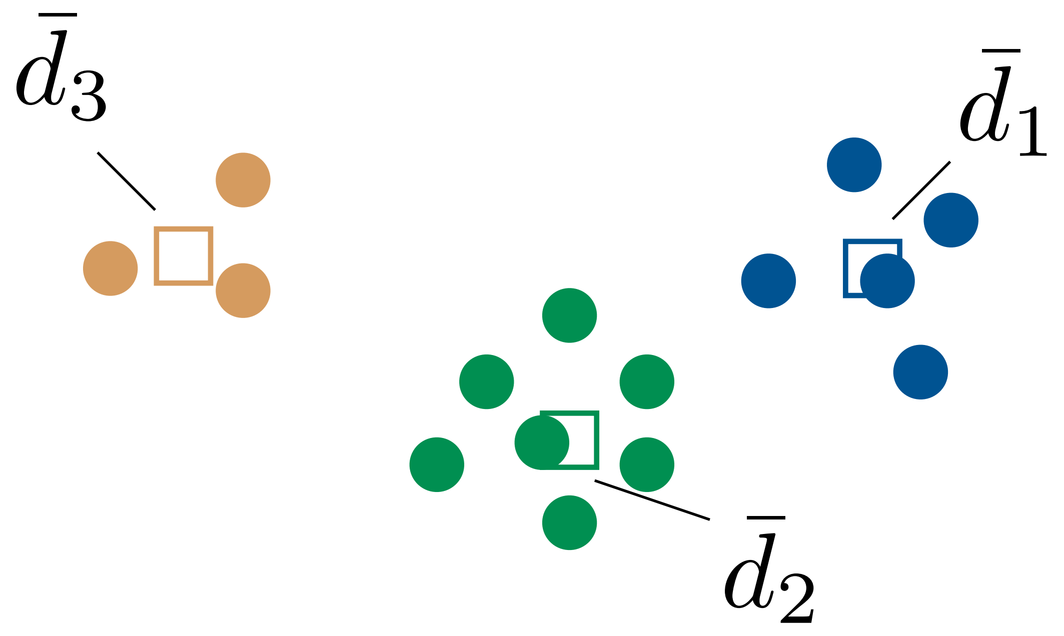
## Theorem

If $-g$ is $L$-smooth in $u$, we have

$$\bar{g}^{N}(x) \leq \bar{g}^{K}(x) \leq \bar{g}^{N}(x) + \frac{L}{2}D(K)$$

## clustering value

$$\sum_{i=1}^{N}\sum_{k=1}^{K} a_{ik}\|d_i - \bar{d}_k\|^2$$



# Bounding the conservatism

**Worst-case values**

$$\bar{g}^{N}(x) = \underset{u \in \mathcal{U}(N,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

$$\bar{g}^{K}(x) = \underset{u \in \mathcal{U}(K,\epsilon)}{\text{maximize}} \quad \bar{g}(x,u)$$

**MRO constraint**

$$\bar{g}(x,u) \le 0 \quad \forall u \in \mathcal{U}(K,\epsilon)$$

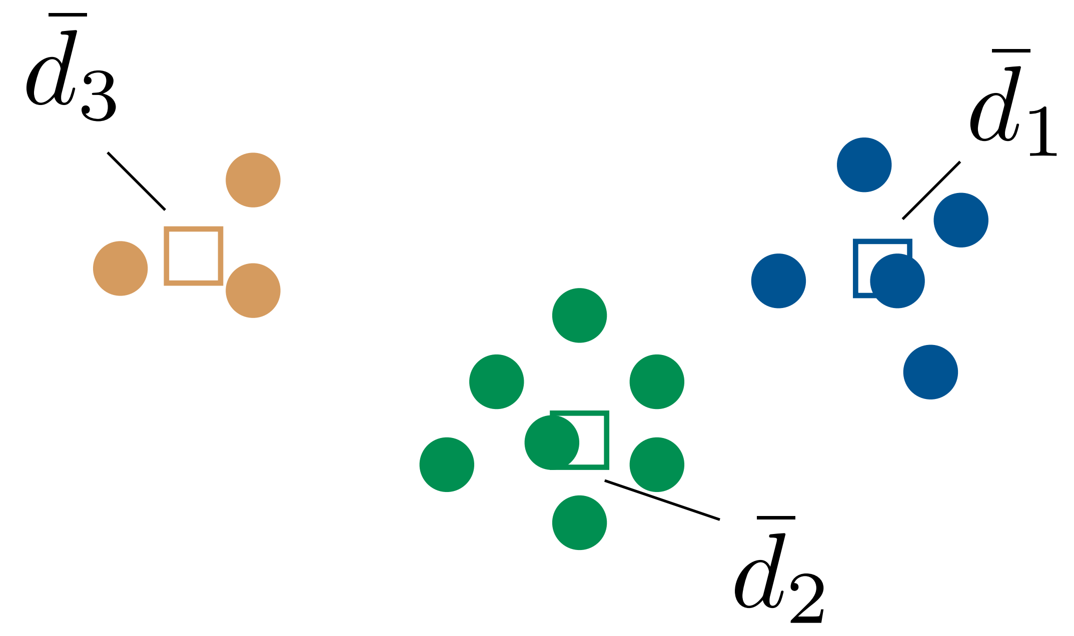
**Theorem**

If $-g$ is $L$-smooth in $u$, we have

$$\bar{g}^{N}(x) \le \bar{g}^{K}(x) \le \bar{g}^{N}(x) + \frac{L}{2} D(K) \leftarrow$$

clustering value

$$\sum_{i=1}^{N} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^2$$



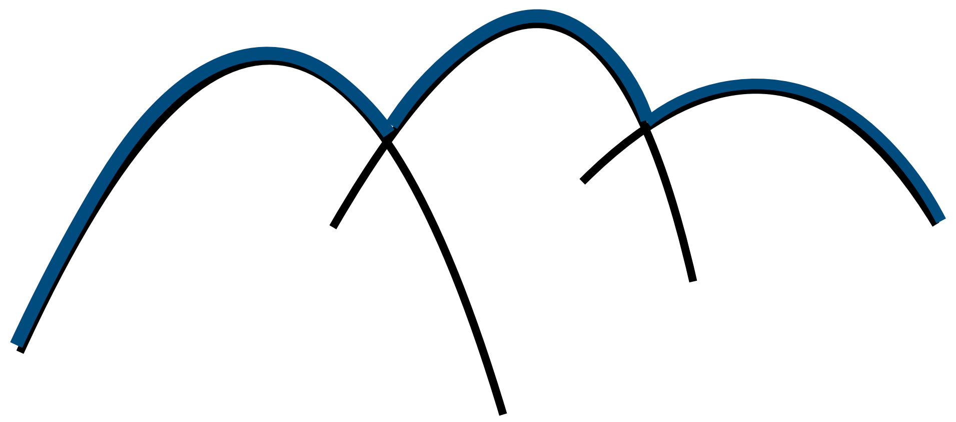
**When $g$ is affine in $u$ ($L = 0$), clustering makes no difference to the optimal value or optimal solution**

# We can also model maximum-of-concave constraints

$$g(x, u) = \max_{j \leq J} g_j(x, u)$$

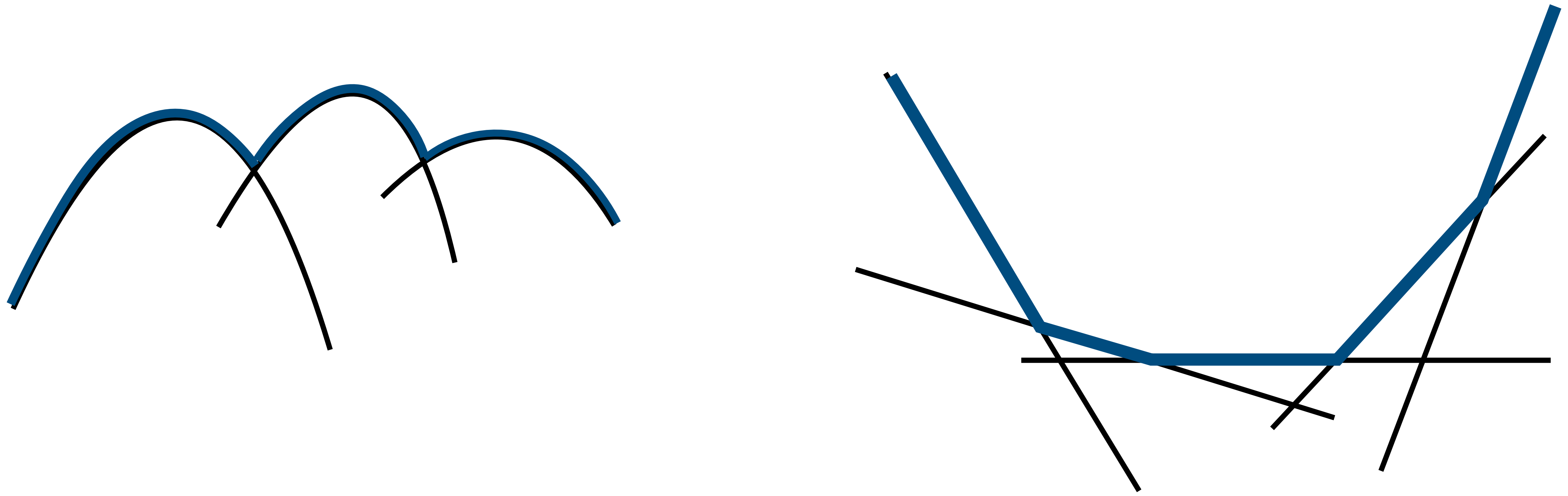
# We can also model maximum-of-concave constraints

$$g(x,u) = \max_{j \leq J} g_j(x,u)$$



# We can also model maximum-of-concave constraints

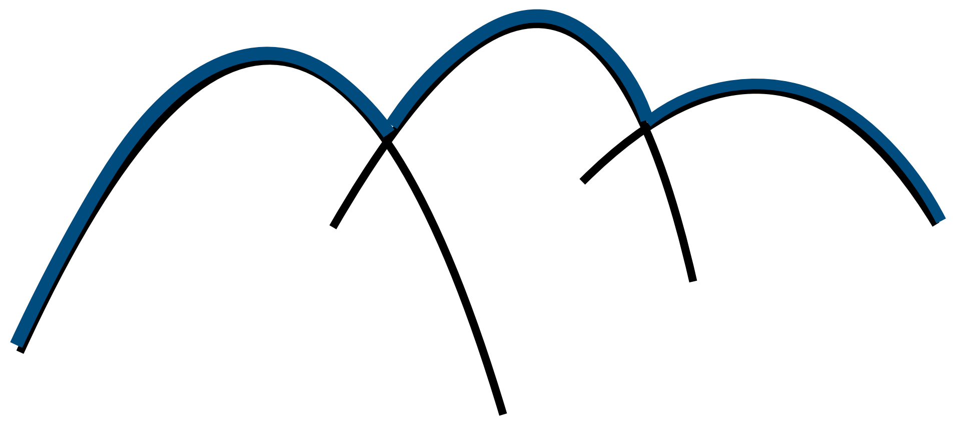
$$g(x,u) = \max_{j \leq J} g_j(x,u)$$

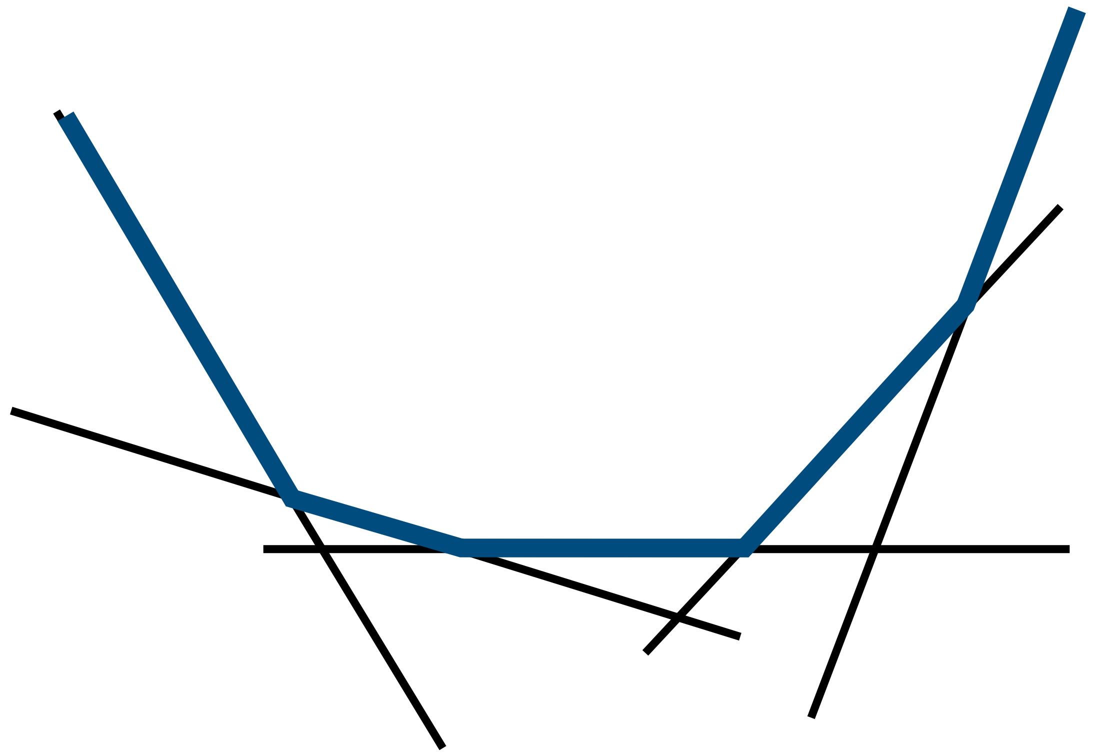


# We can also model maximum-of-concave constraints

$$g(x, u) = \max_{j \leq J} g_j(x, u)$$

- Joint uncertain constraints
- Conditional Value at Risk

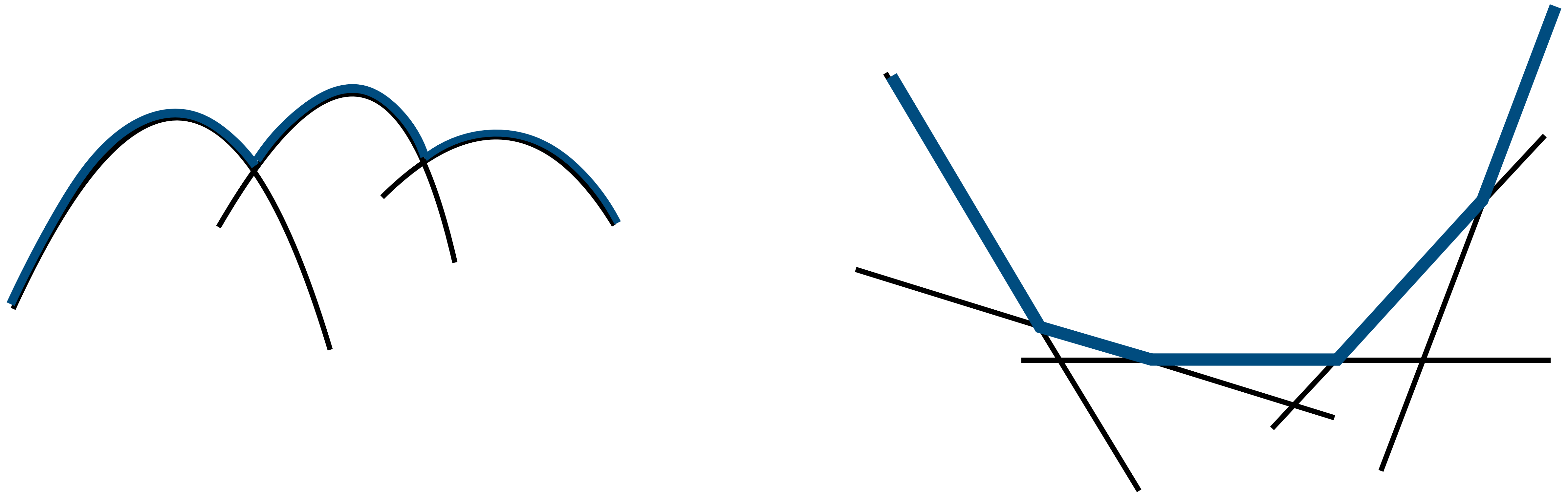




# We can also model maximum-of-concave constraints

$$g(x,u) = \max_{j \le J} g_j(x,u)$$

- Joint uncertain constraints
- Conditional Value at Risk



## MRO uncertainty set

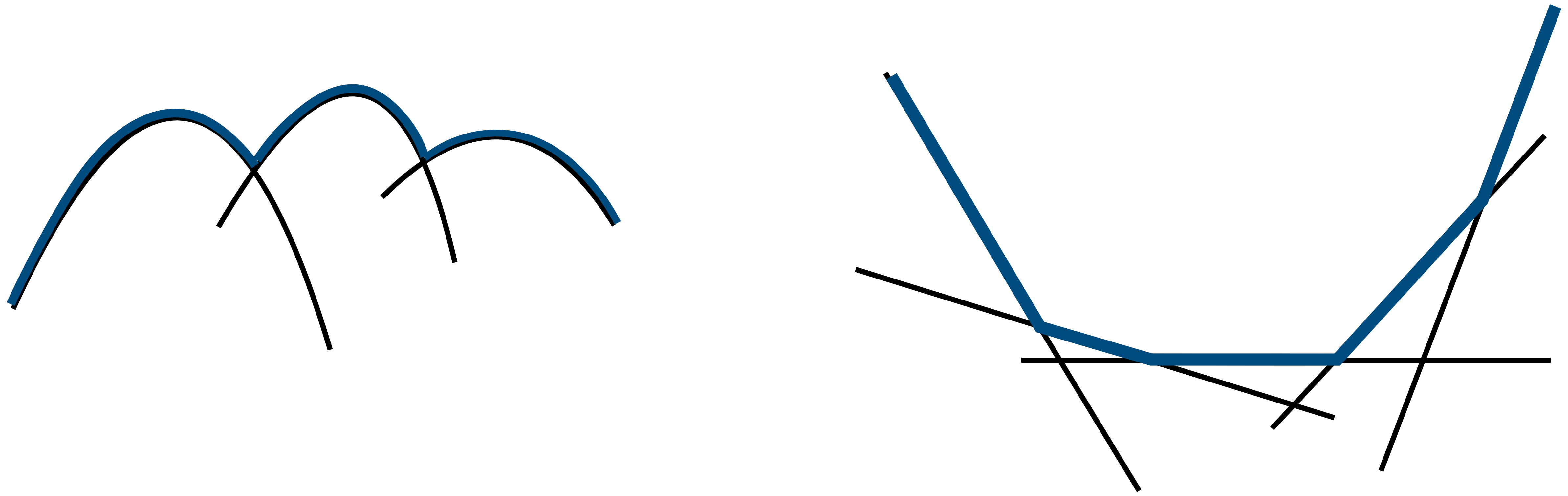
$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_{11},\dots,v_{JK}) \in \mathbf{R}^{J\times K} \;\middle|\; \exists \alpha \in \Gamma,\;\; \sum_{k=1}^{K}\sum_{j=1}^{J} \alpha_{jk}\|v_{jk} - \bar{d}_k\|^p \le \epsilon^p \right\}$$

$$\Gamma = \left\{ \alpha \;\middle|\; \sum_{j=1}^{J} \alpha_{jk} = w_k,\;\; \alpha_{jk} \ge 0 \quad \forall k, j \right\}$$

# We can also model maximum-of-concave constraints

$$g(x,u) = \max_{j \le J} g_j(x,u)$$

- Joint uncertain constraints
- Conditional Value at Risk



## MRO uncertainty set

$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_{11},\dots,v_{JK}) \in \mathbf{R}^{J\times K} \;\middle|\; \exists \alpha \in \Gamma,\ \sum_{k=1}^{K}\sum_{j=1}^{J} \alpha_{jk} \|v_{jk} - \bar{d}_k\|^p \le \epsilon^p \right\}$$

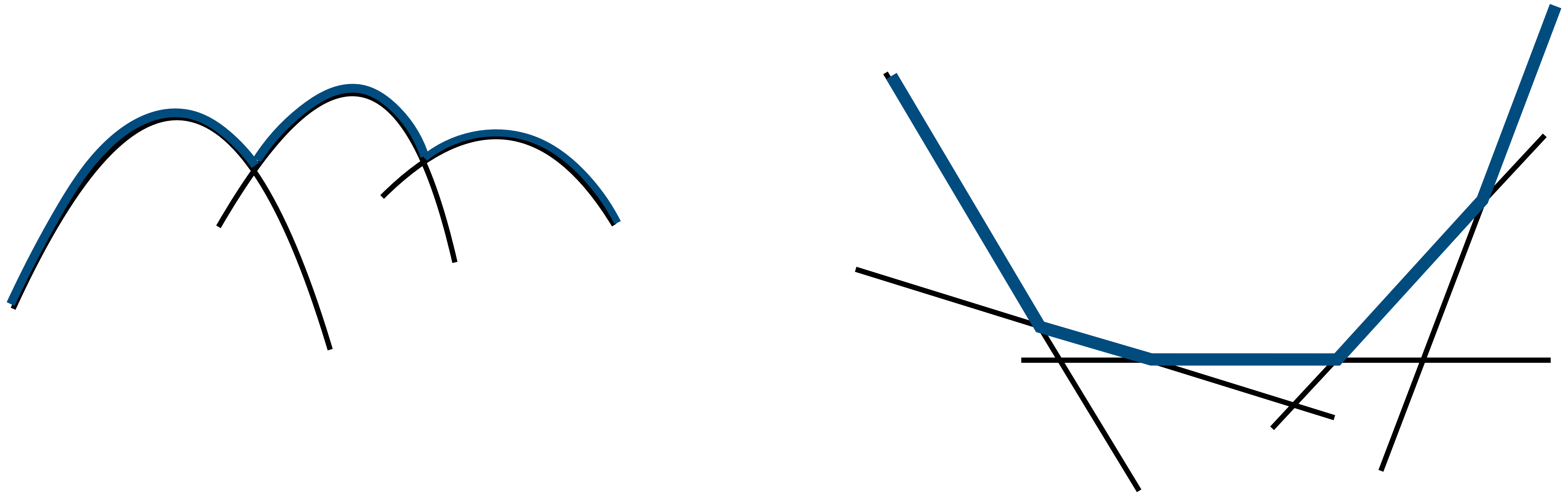
$$\Gamma = \left\{ \alpha \;\middle|\; \sum_{j=1}^{J} \alpha_{jk} = w_k,\ \ \alpha_{jk} \ge 0 \ \ \forall k, j \right\}$$

perturbations split across function pieces

# We can also model maximum-of-concave constraints

$$g(x,u) = \max_{j \leq J} g_j(x,u)$$

- Joint uncertain constraints
- Conditional Value at Risk



## MRO uncertainty set

$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_{11},\ldots,v_{JK}) \in \mathbf{R}^{J\times K} \;\middle|\; \exists \alpha \in \Gamma, \quad \sum_{k=1}^{K}\sum_{j=1}^{J} \alpha_{jk}\|v_{jk} - \bar{d}_k\|^p \leq \epsilon^p \right\}$$

$$\Gamma = \left\{ \alpha \;\middle|\; \sum_{j=1}^{J} \alpha_{jk} = w_k, \quad \alpha_{jk} \geq 0 \quad \forall k,j \right\}$$

perturbations split across function pieces

cluster weights split across pieces

# Worst-case values with maximum-of-concave constraints

$$g(x, u) = \max_{j \leq J} g_j(x, u)$$

# Worst-case values with maximum-of-concave constraints

$$g(x,u) = \max_{j \leq J} g_j(x,u)$$

**Theorem** (max-of-concave function)

If each $-g_j$ is $L_j$-smooth in $u$, we have

$$\bar{g}^N(x) - \Phi^K \leq \bar{g}^K(x) \leq \bar{g}^N(x) + \frac{\max_{j \leq J} L_j}{2} D(K)$$

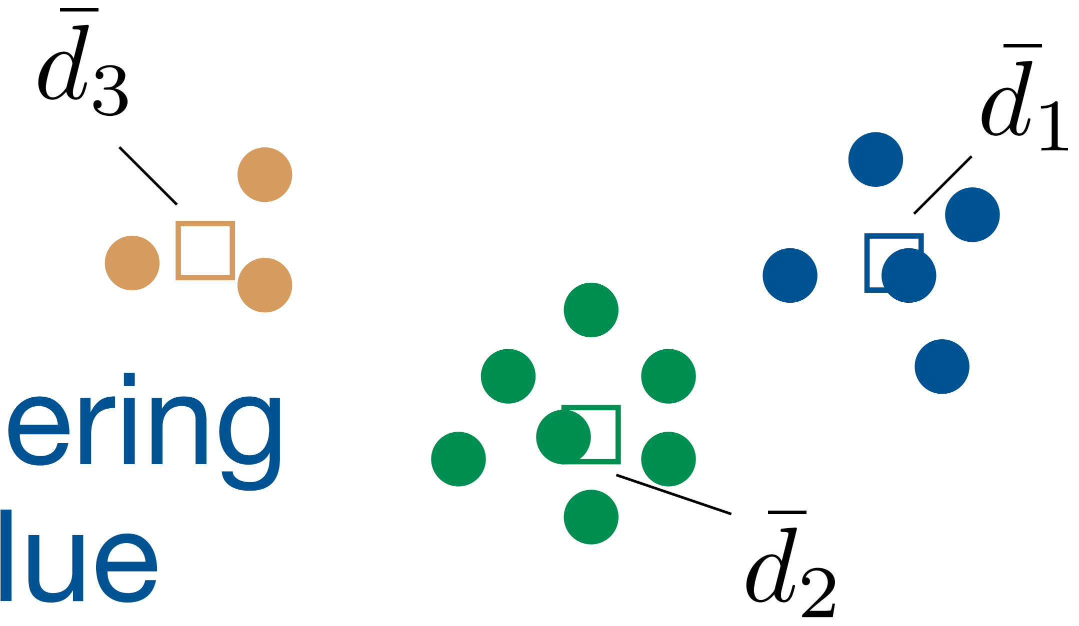
# Worst-case values with maximum-of-concave constraints

$$g(x,u) = \max_{j \le J} g_j(x,u)$$

**Theorem** (max-of-concave function)

If each $-g_j$ is $L_j$-smooth in $u$, we have

$$\bar{g}^N(x) - \Phi^K \le \bar{g}^K(x) \le \bar{g}^N(x) + \frac{\max_{j \le J} L_j}{2} D(K)$$



clustering value

$$\sum_{i=1}^{N} \sum_{k=1}^{K} a_{ik} \| d_i - \bar{d}_k \|^2$$

# Worst-case values with maximum-of-concave constraints

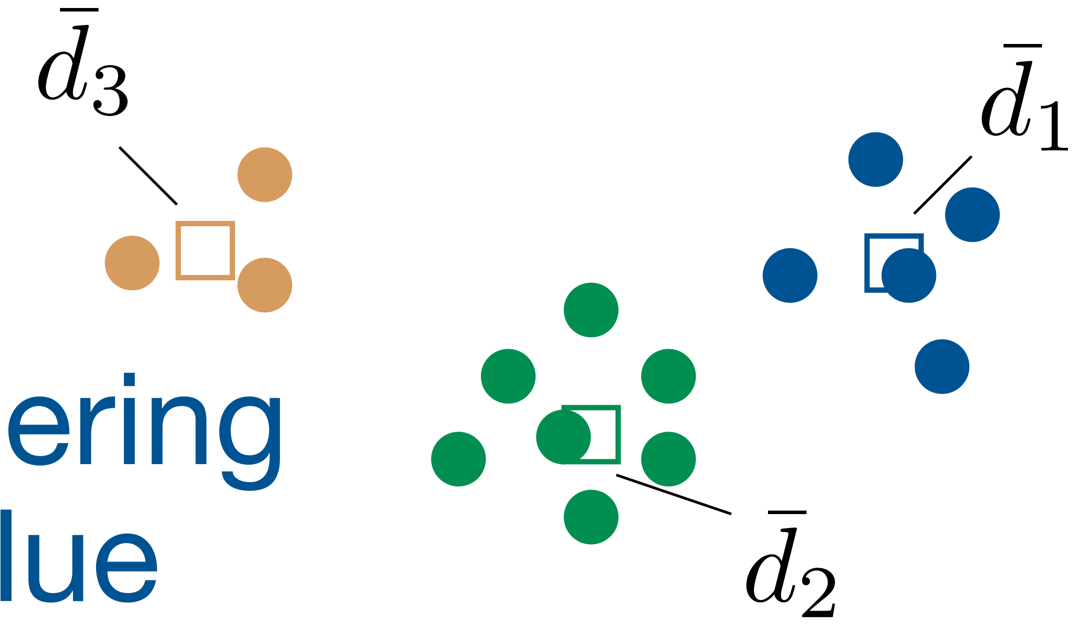
$$g(x,u)=\max_{j\le J} g_j(x,u)$$

**Theorem** (max-of-concave function)

If each $-g_j$ is $L_j$-smooth in $u$, we have

$$\bar{g}^N(x)-\Phi^K \le \bar{g}^K(x) \le \bar{g}^N(x)+\frac{\max_{j\le J} L_j}{2} D(K) \longleftarrow \sum_{i=1}^{N}\sum_{k=1}^{K} a_{ik}\|d_i-\bar{d}_k\|^2$$

maximum of smoothness constants



clustering value

# Worst-case values with maximum-of-concave constraints

$$g(x,u) = \max_{j \leq J} g_j(x,u)$$

**Theorem** (max-of-concave function)

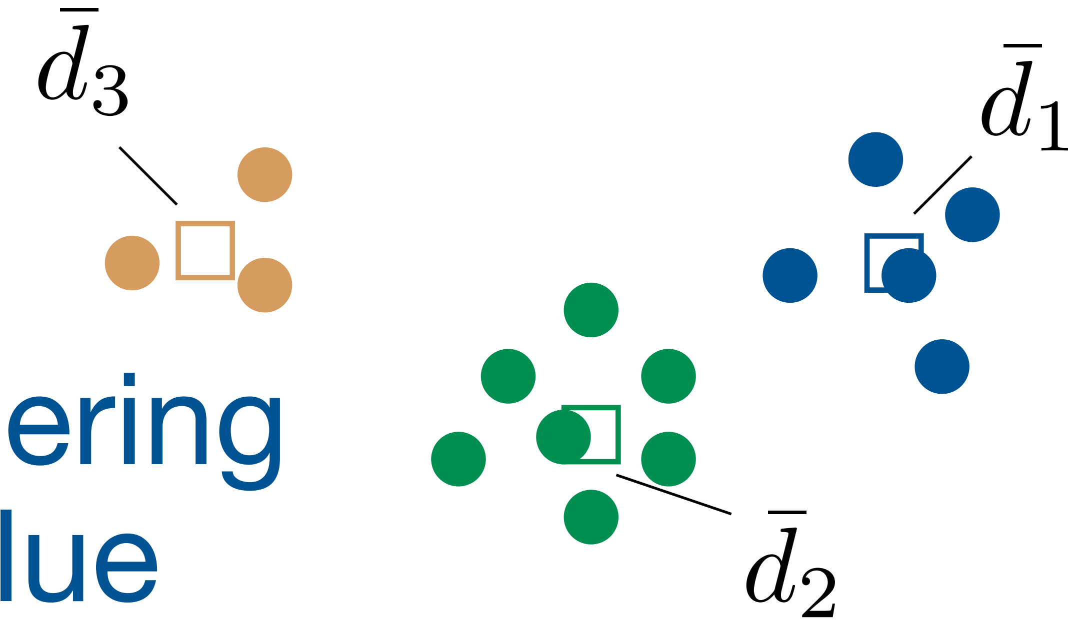
If each $-g_j$ is $L_j$-smooth in $u$, we have

$$\bar{g}^N(x) - \Phi^K \leq \bar{g}^K(x) \leq \bar{g}^N(x) + \frac{\max_{j \leq J} L_j}{2} D(K) \longleftarrow \sum_{i=1}^{N} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^2$$

bound based on dual variables

maximum of smoothness constants

clustering value



# Computational speedups on sparse portfolio optimization

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

# Computational speedups on sparse portfolio optimization

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

← conditional value-at-risk

# Computational speedups on sparse portfolio optimization

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

conditional value-at-risk

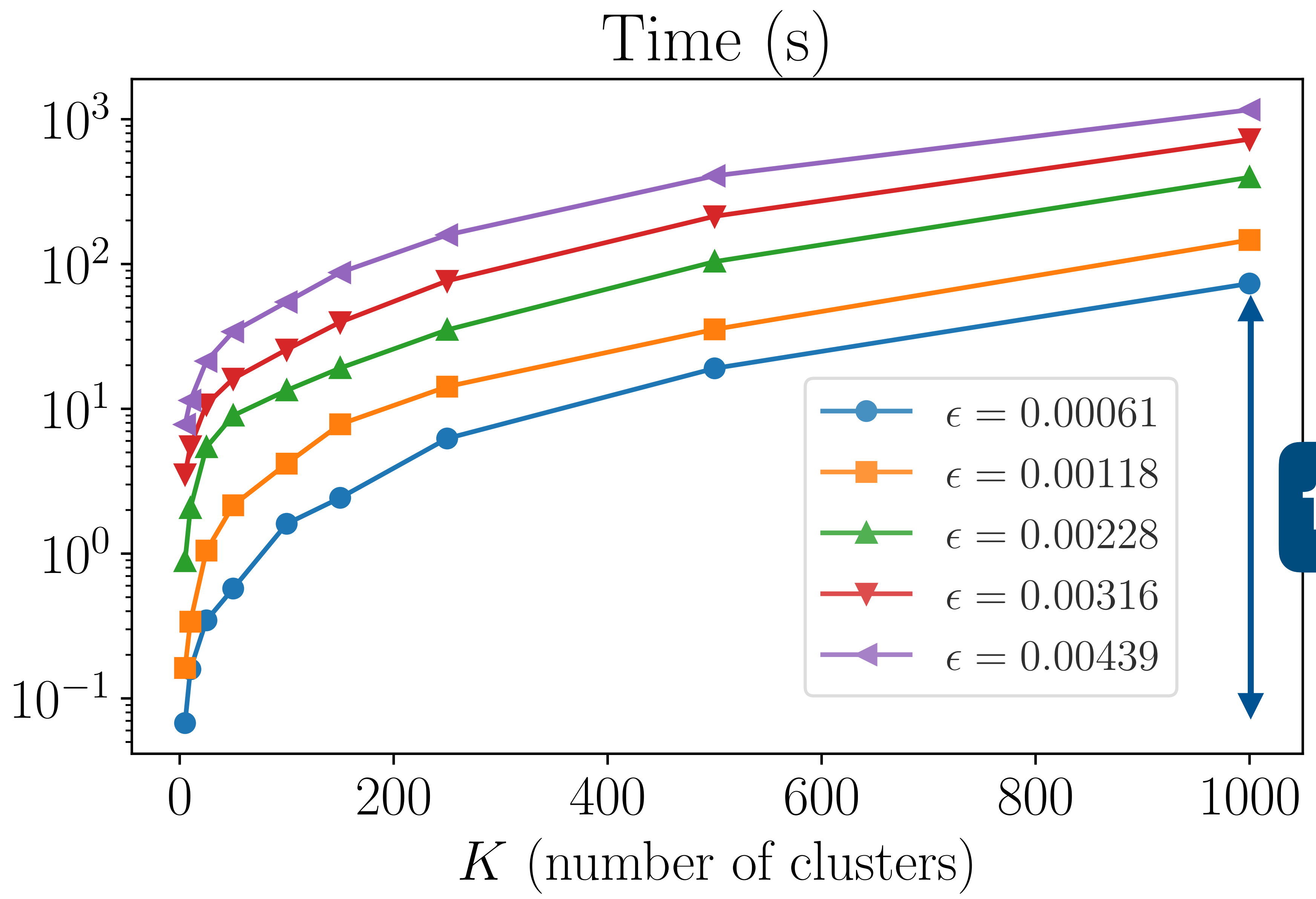
cardinality constraint

# Computational speedups on sparse portfolio optimization

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

conditional value-at-risk

cardinality constraint



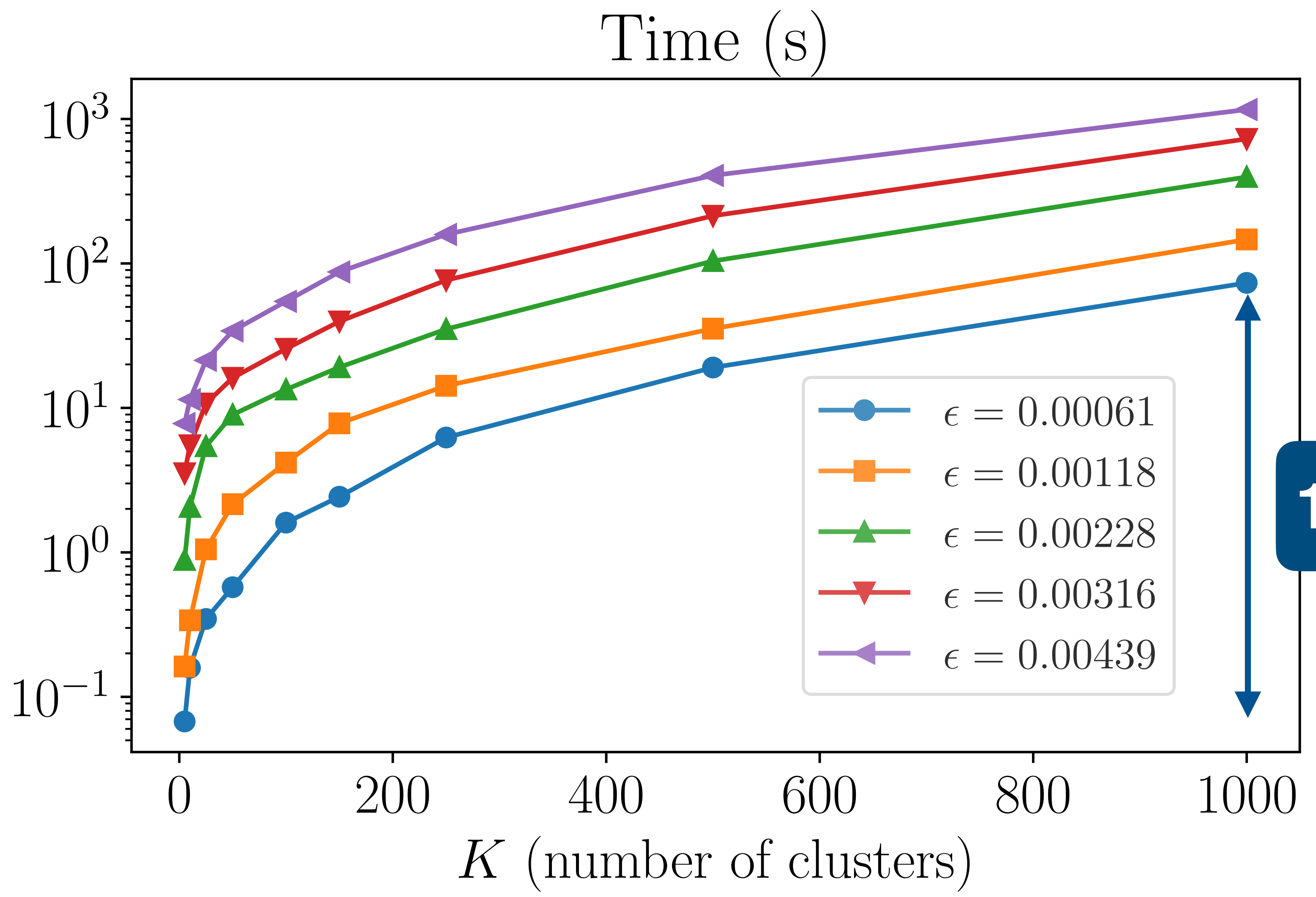
1000x

# Computational speedups on sparse portfolio optimization

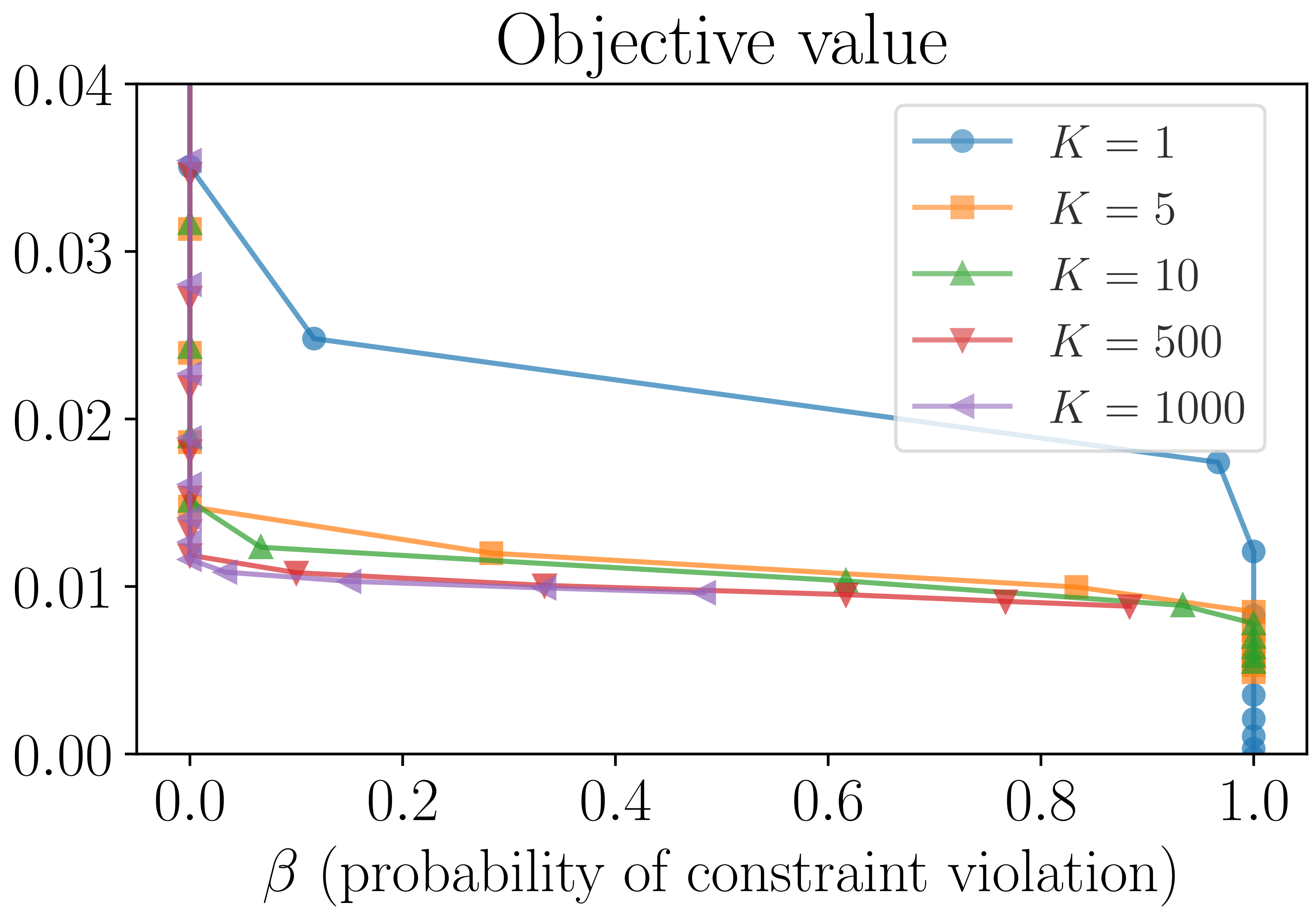
$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

conditional value-at-risk

cardinality constraint



1000x



near-optimal performance with 5 clusters

# Computational speedups on capital budgeting example

$$\begin{array}{ll} \text{maximize} & \eta(u)^T x \\ \text{subject to} & a^T x \leq b \\ & x \in \{0, 1\}^n \end{array}$$

# Computational speedups on capital budgeting example

total
net present value (NPV)

$$
\begin{array}{ll}
\text{maximize} & \eta(u)^T x \\
\text{subject to} & a^T x \leq b \\
& x \in \{0,1\}^n
\end{array}
$$

# Computational speedups on capital budgeting example

total
net present value (NPV)

$$\begin{array}{ll} \text{maximize} & \eta(u)^T x \\ \text{subject to} & a^T x \leq b \\ & x \in \{0,1\}^n \end{array}$$

budget
constraint

# Computational speedups on capital budgeting example

total
net present value (NPV)

$$\begin{array}{ll} \text{maximize} & \eta(u)^T x \\ \text{subject to} & a^T x \leq b \\ & x \in \{0,1\}^n \end{array}$$

budget
constraint

NPV of project $j$

$$\eta_j(u) = \sum_{t=1}^{T} \frac{F_{jt}}{(1+u_j)^t}$$

cash
flow

discount
rate

# Computational speedups on capital budgeting example

total
net present value (NPV)

$$\begin{array}{ll} \text{maximize} & \eta(u)^T x \\ \text{subject to} & a^T x \leq b \\ & x \in \{0,1\}^n \end{array}$$

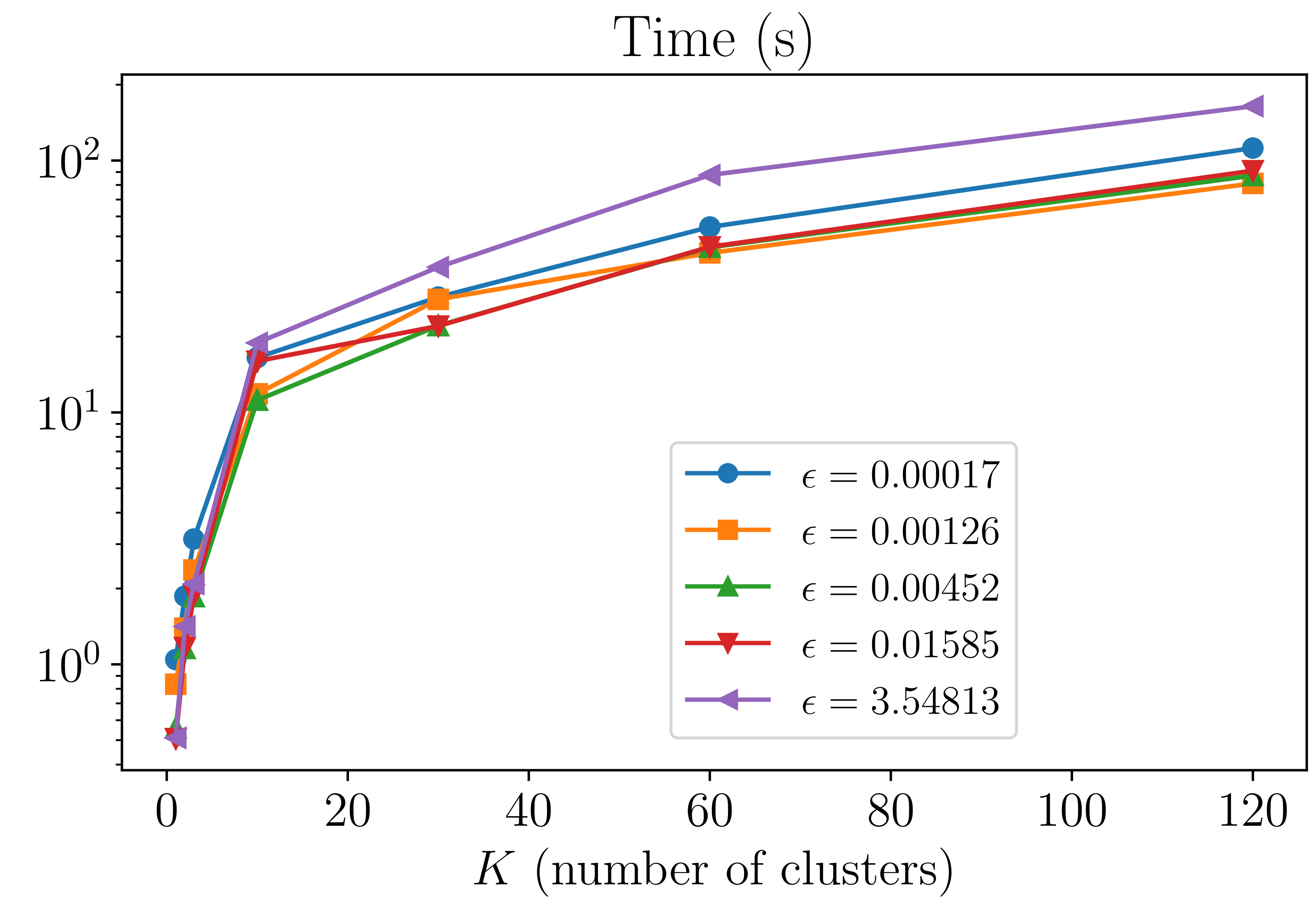
budget
constraint

NPV of project $j$

$$\eta_j(u) = \sum_{t=1}^{T} \frac{F_{jt}}{(1+u_j)^t}$$

cash
flow

discount
rate



# Computational speedups on capital budgeting example

total
net present value (NPV)

$$\begin{aligned} \text{maximize} \quad & \eta(u)^T x \\ \text{subject to} \quad & a^T x \leq b \\ & x \in \{0,1\}^n \end{aligned}$$

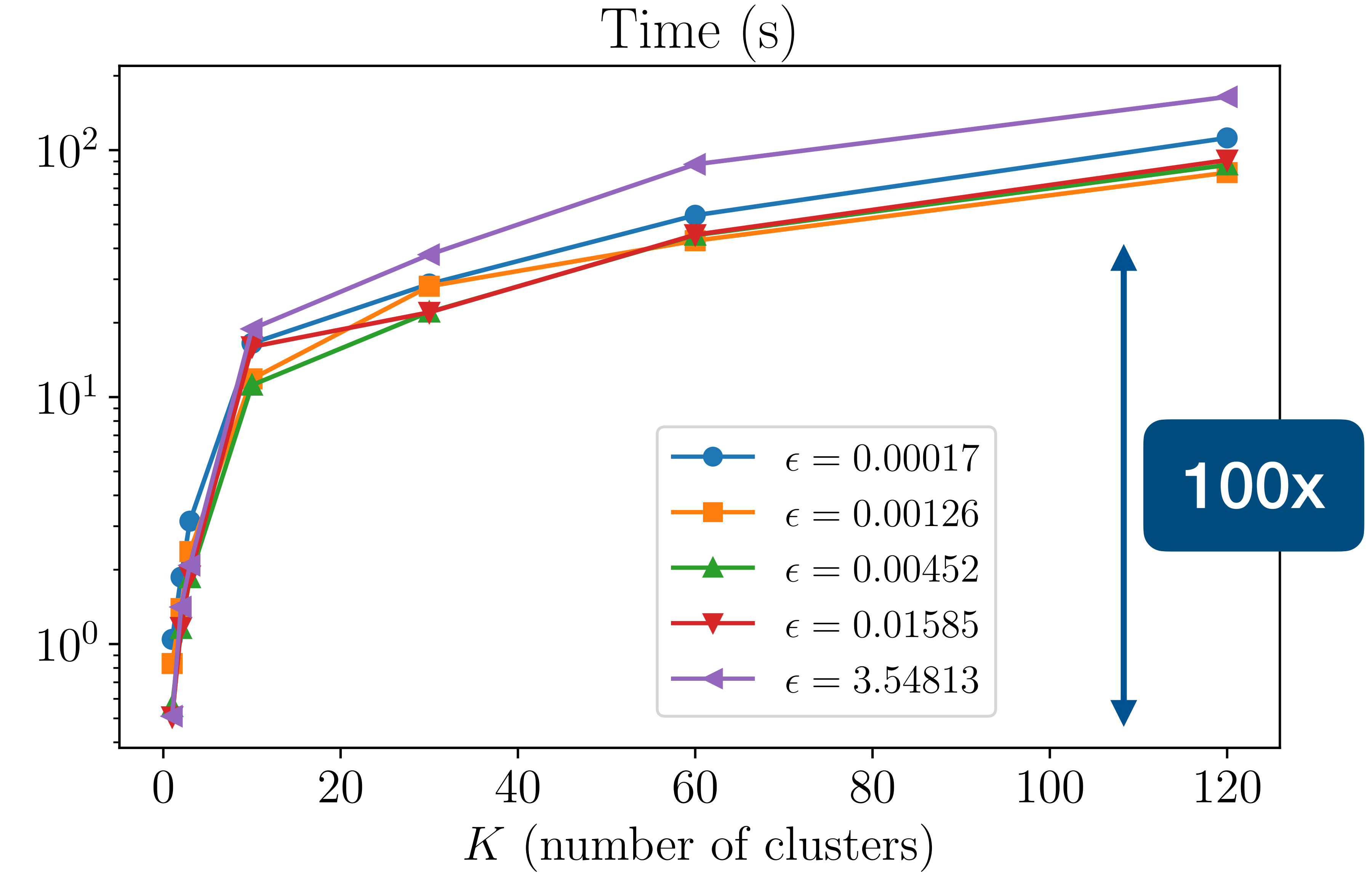
budget
constraint

NPV of project $j$

$$\eta_j(u) = \sum_{t=1}^{T} \frac{F_{jt}}{(1+u_j)^t}$$

cash
flow

discount
rate



# Computational speedups on capital budgeting example

total
net present value (NPV)

$$\begin{aligned} &\text{maximize} && \eta(u)^T x \\ &\text{subject to} && a^T x \le b \\ & && x \in \{0,1\}^n \end{aligned}$$

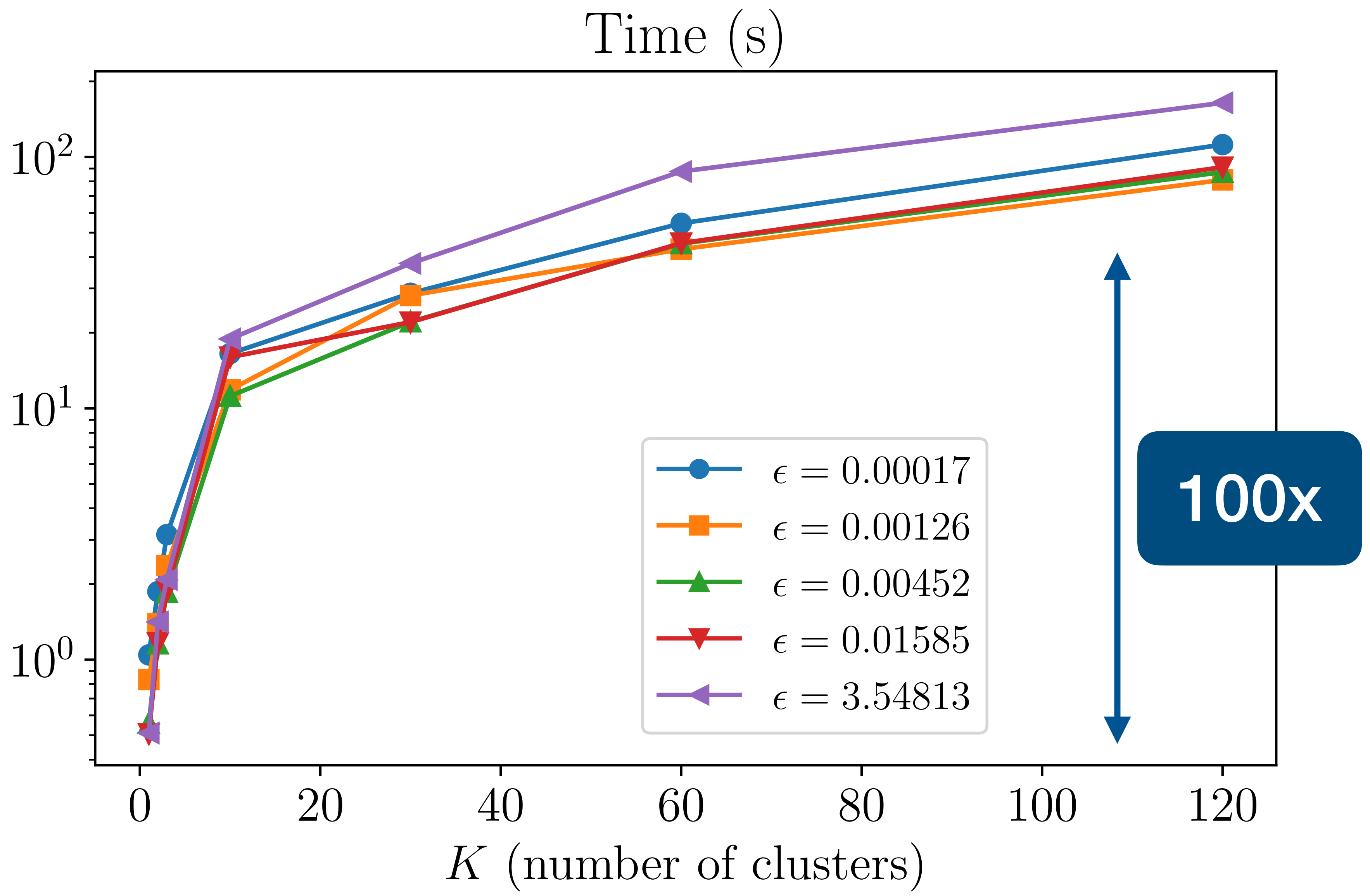
budget
constraint

NPV of project $j$

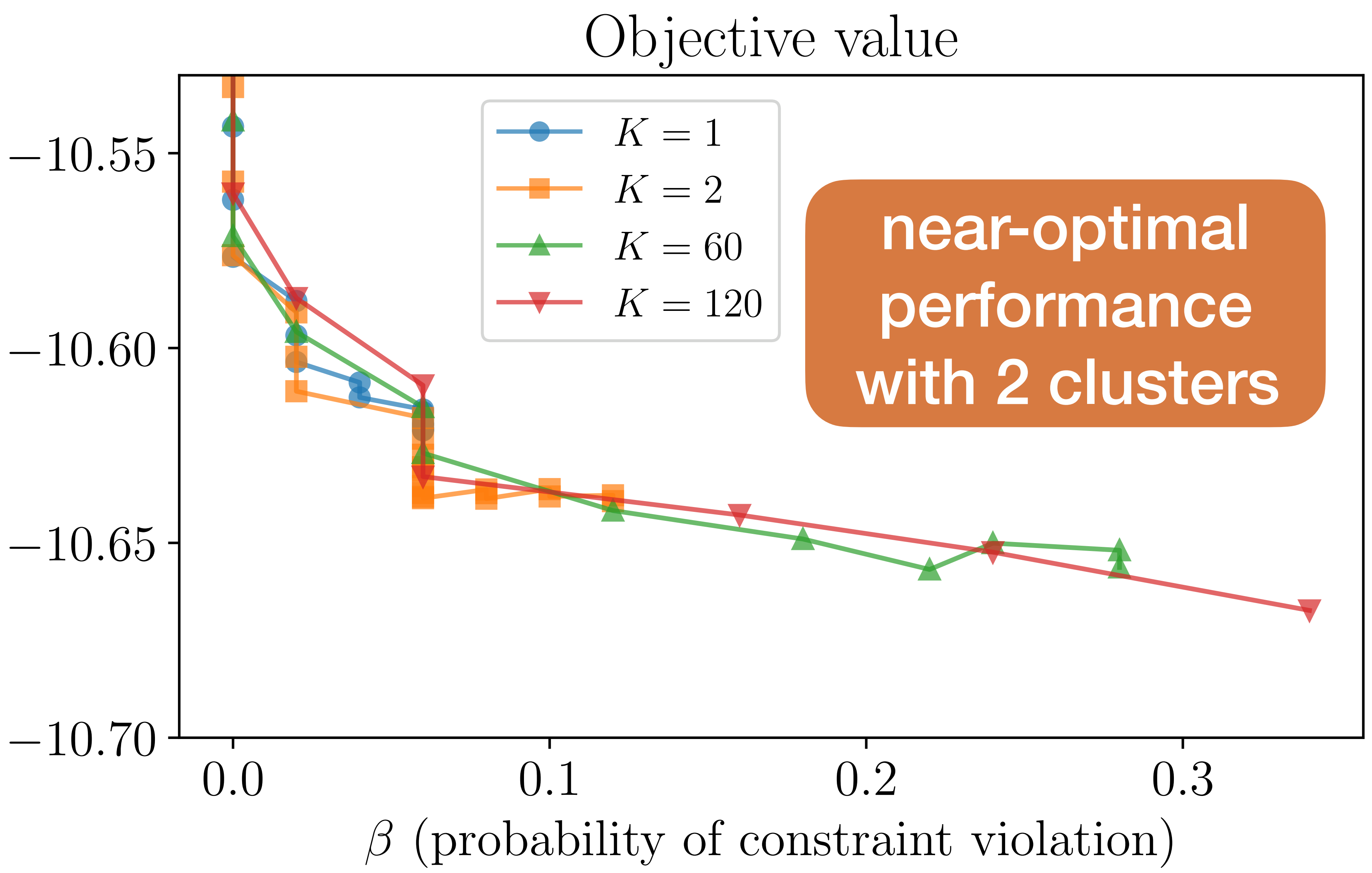
$$\eta_j(u) = \sum_{t=1}^{T} \frac{F_{jt}}{(1+u_j)^t}$$

cash
flow

discount
rate



100x



near-optimal
performance
with 2 clusters

# Mean Robust Optimization with streaming data

# Streaming data brings opportunities and challenges

streaming data

time step

$$\mathcal{D}_t = \{d_i\}_{i=1}^{t}$$

# Streaming data brings opportunities and challenges

streaming data

time step

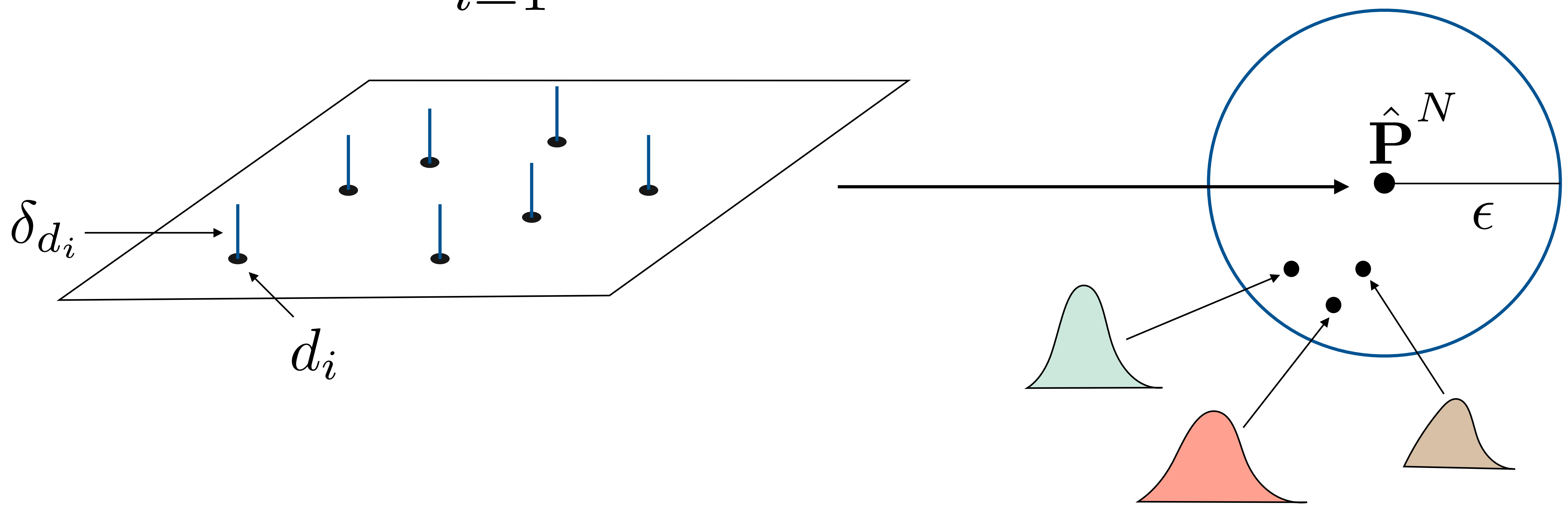
$$\mathcal{D}_t = \{d_i\}_{i=1}^{t}$$

empirical distribution

$$\hat{\mathbf{P}}_t = \frac{1}{t}\sum_{i=1}^{t}\delta_{d_i}$$

ambiguity set

$$\mathcal{P}_t = \left\{\mathbf{P} \mid W_p(\hat{\mathbf{P}}_t, \mathbf{P}) \leq \epsilon_t\right\}$$



# Streaming data brings opportunities and challenges

**streaming data**

$$\mathcal{D}_t = \{d_i\}_{i=1}^{t}$$

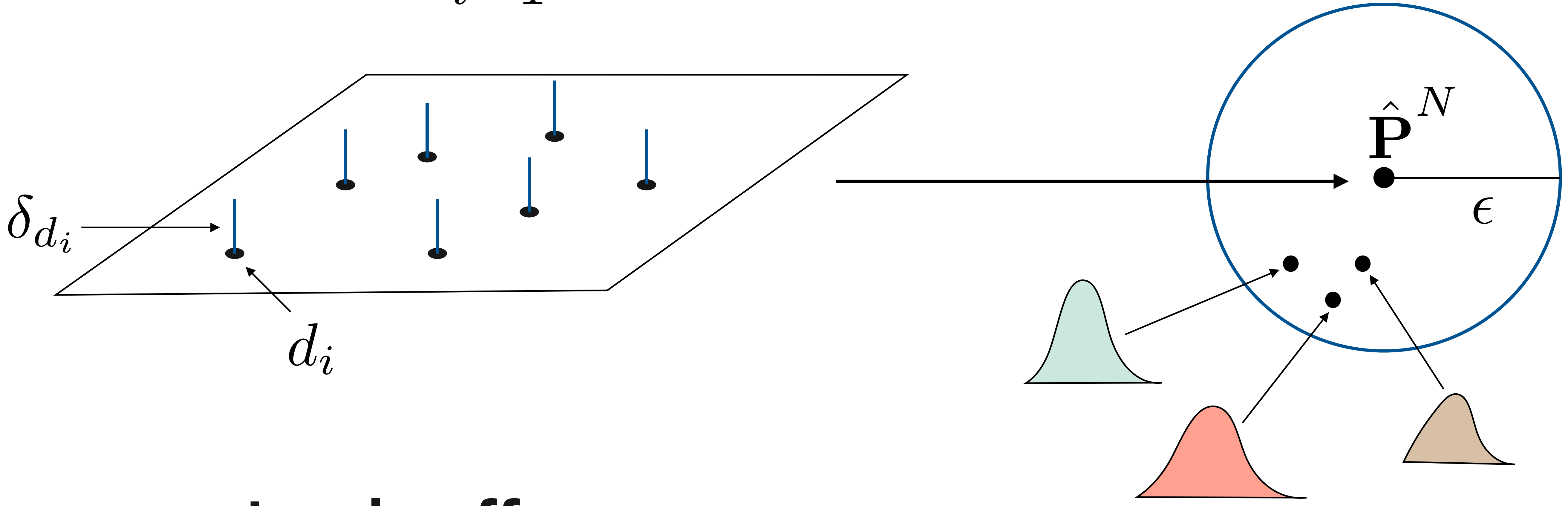
time step

**empirical distribution**

$$\hat{\mathbf{P}}_t = \frac{1}{t}\sum_{i=1}^{t}\delta_{d_i}$$

**ambiguity set**

$$\mathcal{P}_t = \left\{\mathbf{P} \mid W_p(\hat{\mathbf{P}}_t, \mathbf{P}) \leq \epsilon_t\right\}$$



**tradeoff**

**statistics**

*shrinking radius*

$$\epsilon_t \sim t^{-1/m}$$

Fournier and Guillin (2013)

**computation**

*increasing computational cost*

$1$ datapoint $\leftrightarrow$ $1$ constraint

# Many problems with streaming data are sequential

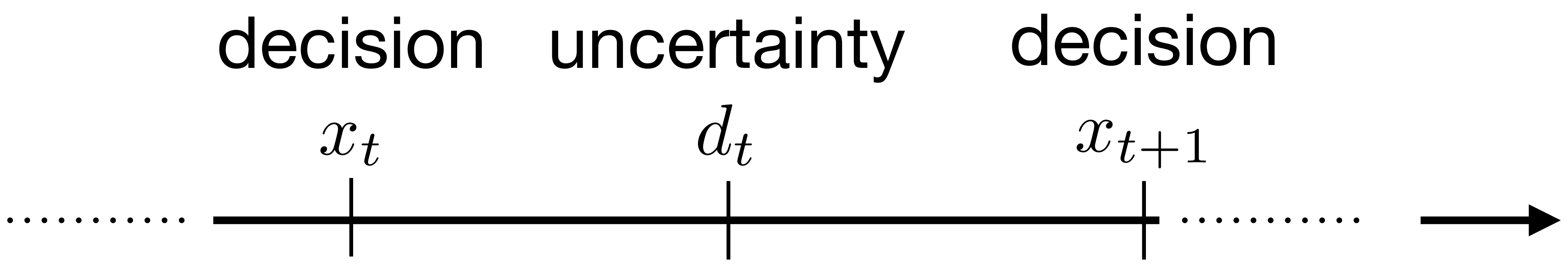
goal

$$H_\star = \inf_{x \in X} \mathbf{E}(f(x,u))$$

# Many problems with streaming data are sequential

**goal**

$$H_\star = \inf_{x \in X} \mathbf{E}(f(x,u))$$



decision $x_t$

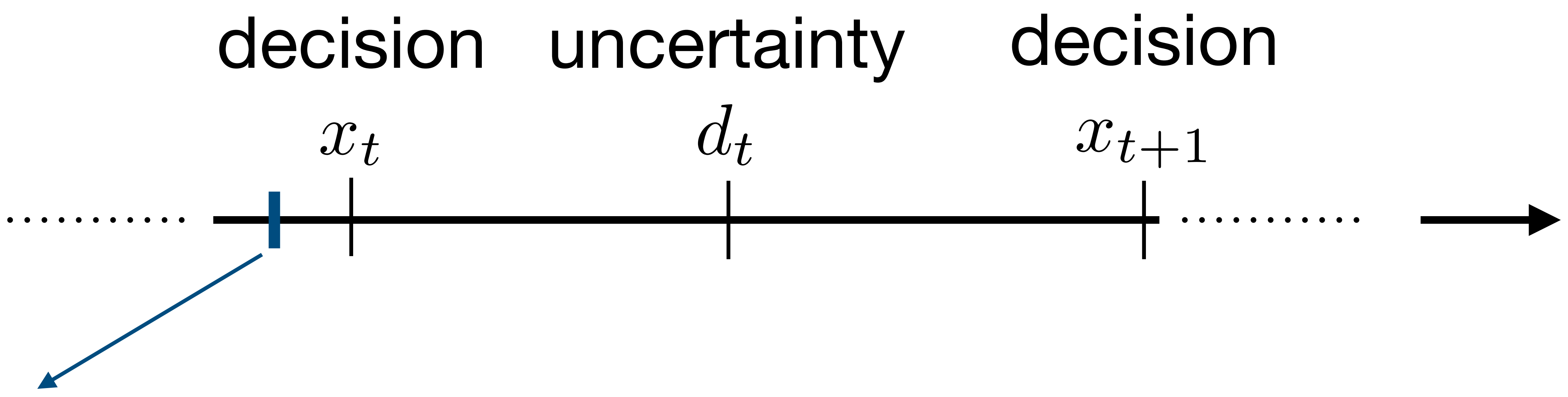
uncertainty $d_t$

decision $x_{t+1}$

# Many problems with streaming data are sequential

**goal**

$$H_\star = \inf_{x\in X} \mathbf{E}(f(x,u))$$



decision $x_t$

uncertainty $d_t$

decision $x_{t+1}$

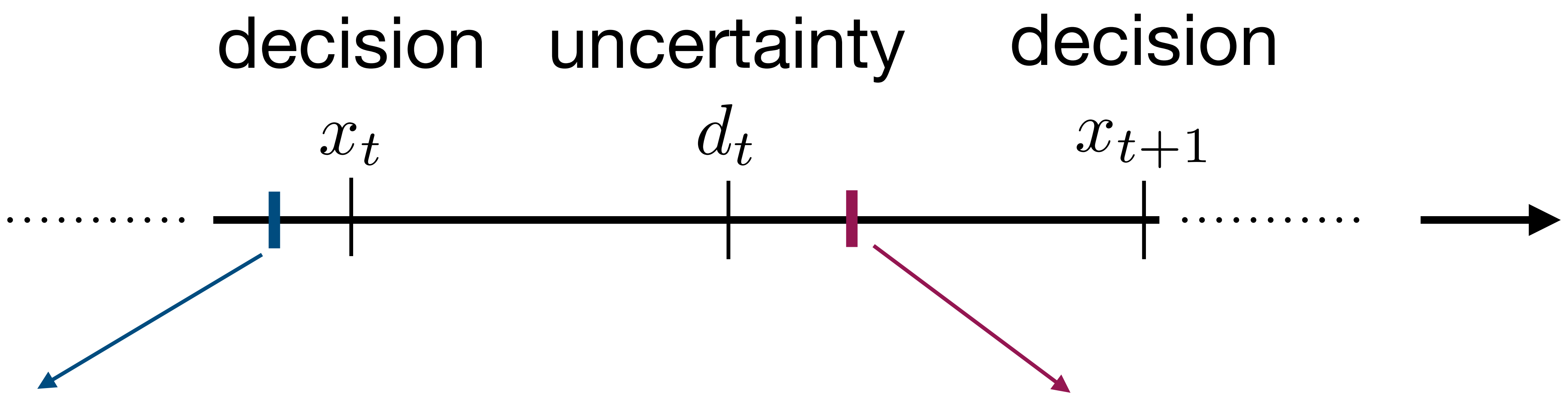
compute decision

$$x_t = \operatorname*{argmin}_{x\in X} \sup_{\mathbf{P}\in\mathcal{P}_t} \mathbf{E}(f(x,u))$$

# Many problems with streaming data are sequential

**goal**

$$H_\star = \inf_{x \in X} \mathbf{E}(f(x,u))$$



compute decision

$$x_t = \text{argmin}_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t} \mathbf{E}(f(x,u))$$

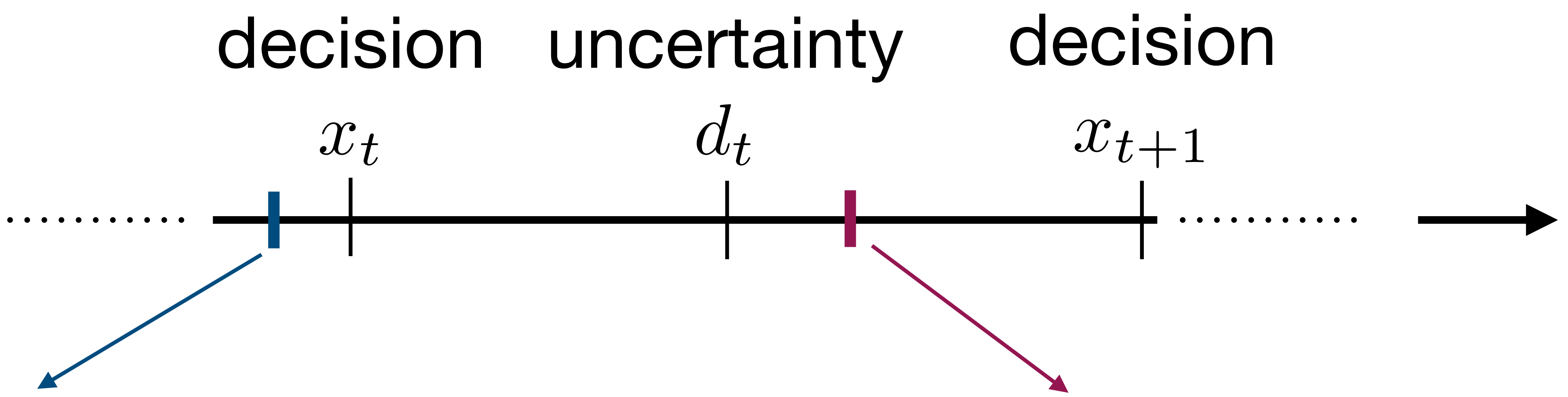
update ambiguity set

$$\mathcal{P}_t = \left\{ \mathbf{Q} \mid W_p(\hat{\mathbf{P}}_t, \mathbf{Q}) \leq \epsilon_t \right\}$$

# Many problems with streaming data are sequential

**goal**

$$H_\star = \inf_{x \in X} \mathbf{E}(f(x,u))$$



compute decision

$$x_t = \text{argmin}_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t} \mathbf{E}(f(x,u))$$

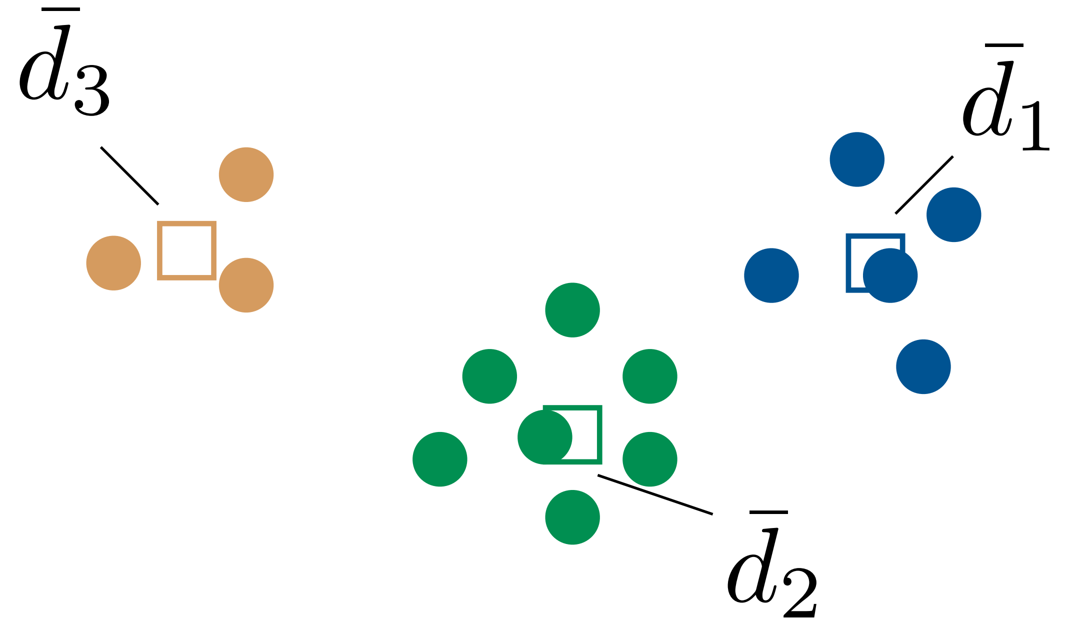
update ambiguity set

$$\mathcal{P}_t = \left\{ \mathbf{Q} \mid W_p(\hat{\mathbf{P}}_t, \mathbf{Q}) \leq \epsilon_t \right\}$$

**Can Mean Robust Optimization help?**

# Idea: compress data via online clustering

# Idea: compress data via online clustering



**clustered empirical distribution**

$$\hat{\mathbf{P}}_t^K = \sum_{k=1}^{K} w_t^k \delta_{\bar{d}_t^k}$$

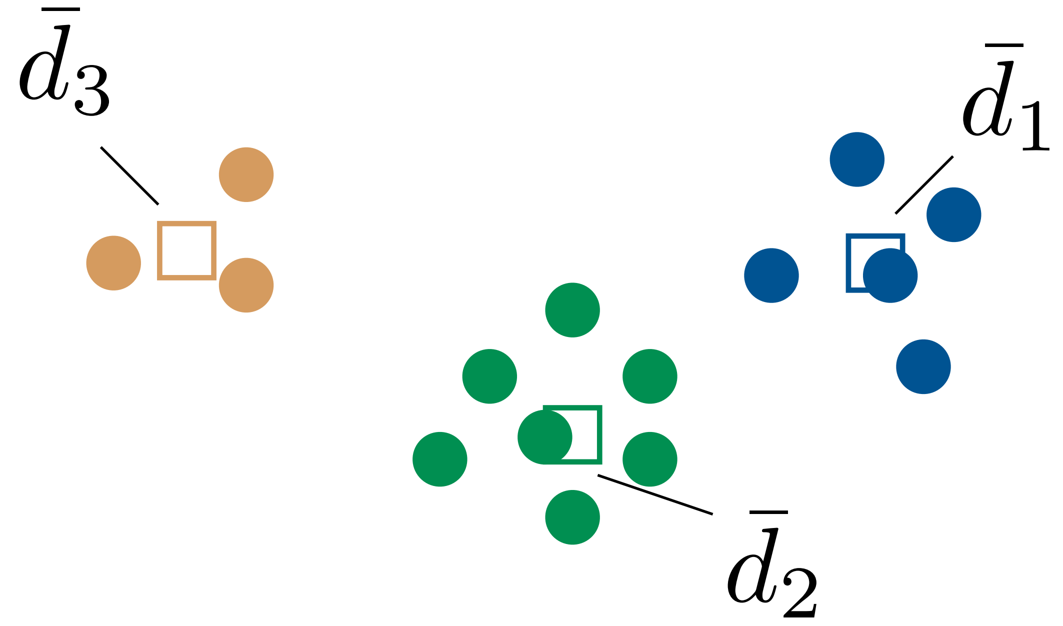
cluster centers

cluster weights

dynamic clustering-based ambiguity set (via MRO)

$$\mathcal{P}_t^K = \left\{ \mathbf{Q} \mid W_p(\hat{\mathbf{P}}_t^K, \mathbf{Q}) \leq \epsilon_t \right\}$$

# Idea: compress data via online clustering



clustered empirical distribution

$$\hat{\mathbf{P}}_t^K = \sum_{k=1}^{K} w_t^k \delta_{\bar{d}_t^k} \longleftarrow \text{cluster centers}$$

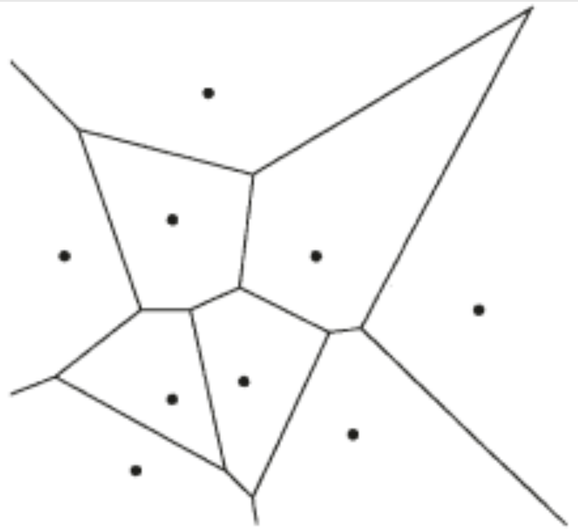
cluster weights

dynamic clustering-based ambiguity set (via MRO)

$$\longrightarrow \mathcal{P}_t^K = \left\{\mathbf{Q} \mid W_p(\hat{\mathbf{P}}_t^K, \mathbf{Q}) \le \epsilon_t\right\}$$

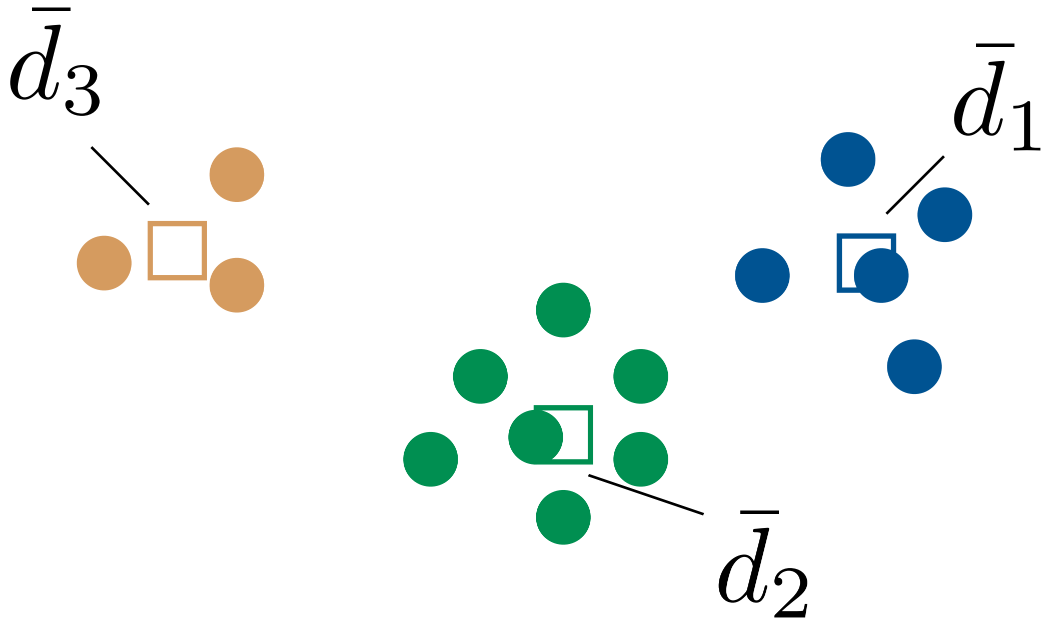
it works with many clustering algorithms

Fixed partitioning



Zubaroğlu and Atalay (2021)

# Idea: compress data via online clustering



**clustered empirical distribution**

$$\hat{\mathbf{P}}_t^K = \sum_{k=1}^{K} w_t^k \delta_{\bar{d}_t^k}$$

cluster centers

cluster weights

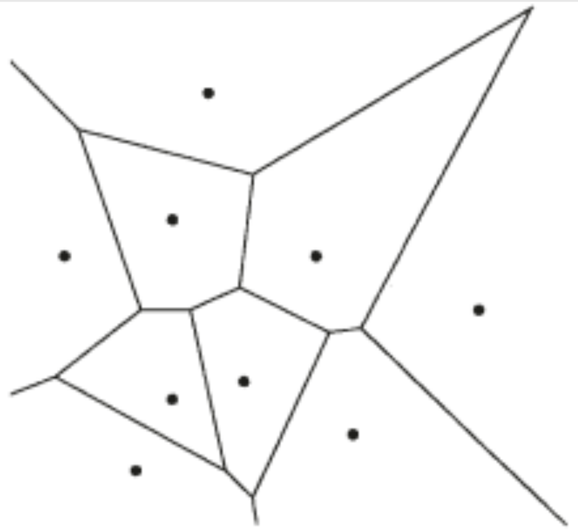
→

**dynamic clustering-based ambiguity set (via MRO)**

$$\mathcal{P}_t^K = \left\{ \mathbf{Q} \mid W_p(\hat{\mathbf{P}}_t^K, \mathbf{Q}) \leq \epsilon_t \right\}$$

**it works with many clustering algorithms**

Fixed partitioning

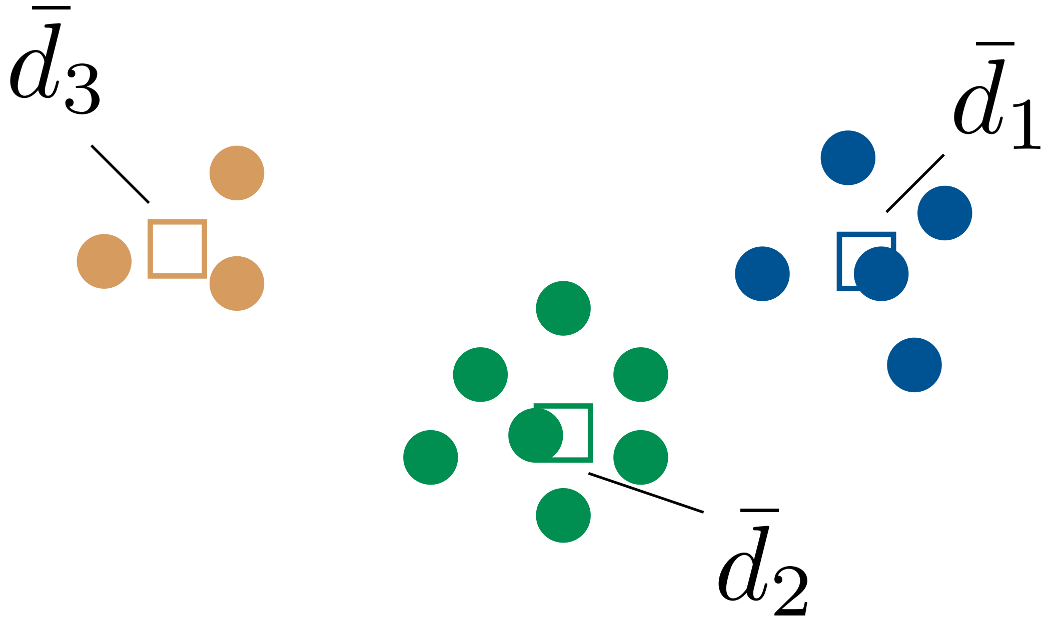


*k*-means *reclustering* at each *t*

$$\min \sum_{i=1}^{t} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^2$$

Zubaroğlu and Atalay (2021)

# Idea: compress data via online clustering



clustered empirical distribution

$$\hat{\mathbf{P}}_t^K = \sum_{k=1}^{K} w_t^k \delta_{\bar{d}_t^k}$$

cluster centers

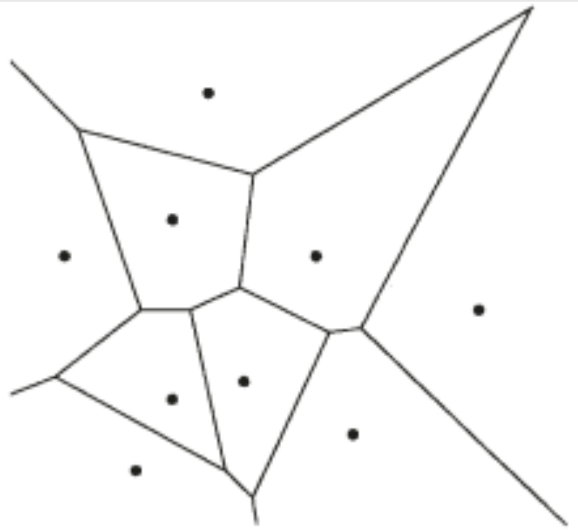
cluster weights

dynamic clustering-based ambiguity set (via MRO)

$$\mathcal{P}_t^K = \left\{ \mathbf{Q} \mid W_p(\hat{\mathbf{P}}_t^K, \mathbf{Q}) \leq \epsilon_t \right\}$$

it works with many clustering algorithms

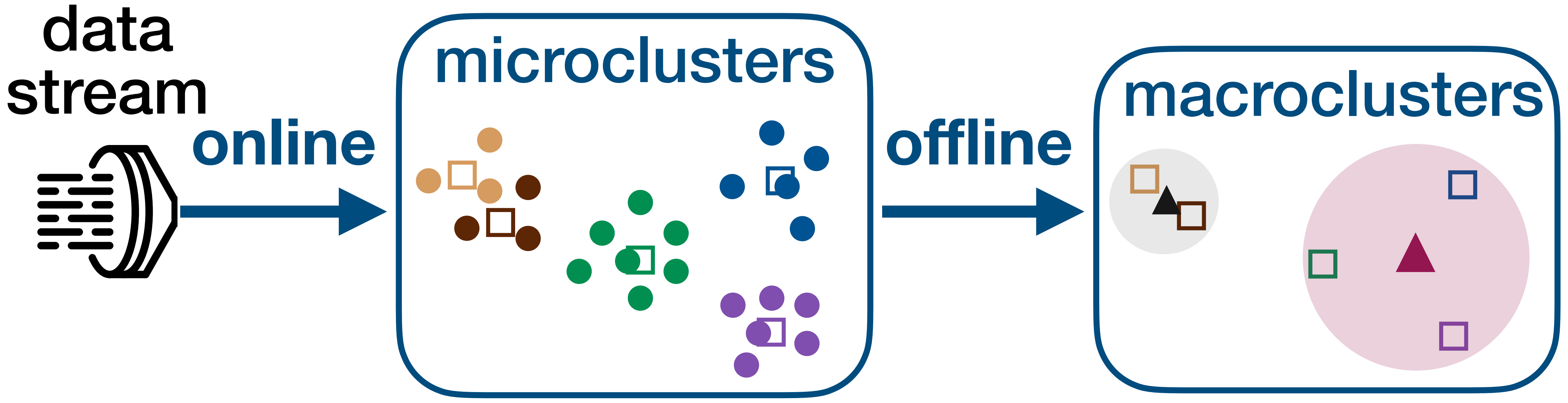
Fixed partitioning



*k*-means *reclustering* at each *t*

$$\min \quad \sum_{i=1}^{t} \sum_{k=1}^{K} a_{ik} \| d_i - \bar{d}_k \|^2$$

two-phase clustering



Zubaroğlu and Atalay (2021)

# Bounding the performance lost via clustering

optimal DRO value
at time $t$

$$H_t = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t} \mathbf{E}(f(x,u))$$

optimal MRO value
at time $t$ with $K$ clusters

$$H_t^K = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t^K} \mathbf{E}(f(x,u))$$

# Bounding the performance lost via clustering

optimal DRO value
at time $t$

$$H_t = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t} \mathbf{E}(f(x,u))$$

optimal MRO value
at time $t$ with $K$ clusters

$$H_t^K = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t^K} \mathbf{E}(f(x,u))$$

### Theorem (concave function)

If $f$ is $L$-smooth and $M$-Lipschitz in $u$, we have

$$H_t \leq H_t^K \leq H_t + \min\left\{\frac{L}{2} D_t(K), M(2\epsilon_t + W_p(\hat{P}_t, \hat{P}_t^K))\right\}$$

# Bounding the performance lost via clustering

optimal DRO value at time $t$

$$H_t = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t} \mathbf{E}(f(x,u))$$

optimal MRO value at time $t$ with $K$ clusters

$$H_t^K = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t^K} \mathbf{E}(f(x,u))$$

### Theorem (concave function)

If $f$ is $L$-smooth and $M$-Lipschitz in $u$, we have

$$H_t \leq H_t^K \leq H_t + \min\left\{\frac{L}{2} D_t(K), M(2\epsilon_t + W_p(\hat{P}_t, \hat{P}_t^K))\right\}$$

clustering value at time $t$

$$\sum_{i=1}^{t} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^2$$

# Bounding the performance lost via clustering

optimal DRO value
at time $t$

$$H_t = \inf_{x\in X} \sup_{\mathbf{P}\in\mathcal{P}_t} \mathbf{E}(f(x,u))$$

optimal MRO value
at time $t$ with $K$ clusters

$$H_t^K = \inf_{x\in X} \sup_{\mathbf{P}\in\mathcal{P}_t^K} \mathbf{E}(f(x,u))$$

## Theorem (concave function)

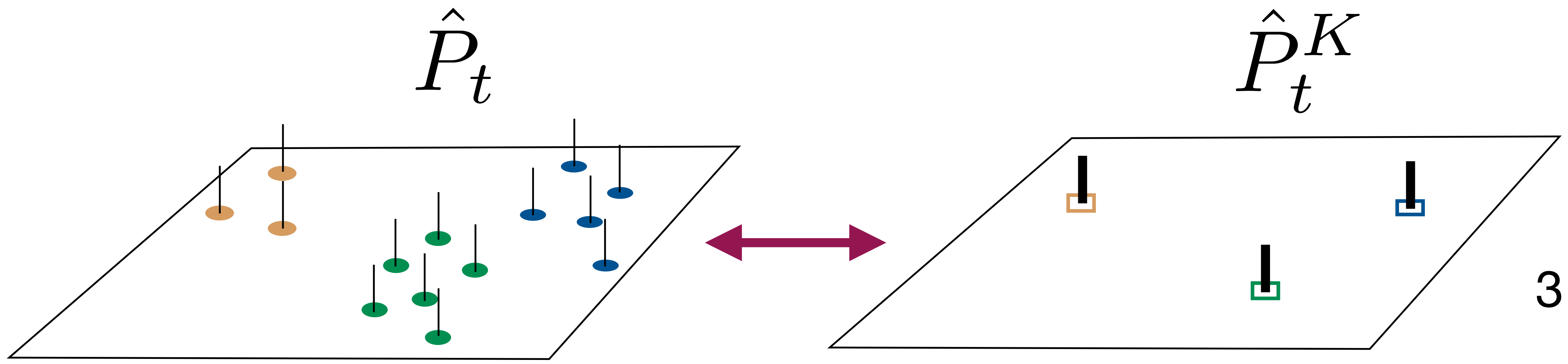
If $f$ is $L$-smooth and $M$-Lipschitz in $u$, we have

$$H_t \le H_t^K \le H_t + \min\left\{\frac{L}{2}D_t(K), M(2\epsilon_t + W_p(\hat{P}_t, \hat{P}_t^K))\right\}$$

clustering
value at time $t$

$$\sum_{i=1}^{t}\sum_{k=1}^{K} a_{ik}\|d_i - \bar{d}_k\|^2$$

Wasserstein distance



# Bounding the performance lost via clustering

optimal DRO value
at time $t$

$$H_t = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t} \mathbf{E}(f(x,u))$$

optimal MRO value
at time $t$ with $K$ clusters

$$H_t^K = \inf_{x \in X} \sup_{\mathbf{P} \in \mathcal{P}_t^K} \mathbf{E}(f(x,u))$$

### Theorem (concave function)

If $f$ is $L$-smooth and $M$-Lipschitz in $u$, we have

$$H_t \leq H_t^K \leq H_t + \min\left\{ \frac{L}{2} D_t(K), M(2\epsilon_t + W_p(\hat{P}_t, \hat{P}_t^K)) \right\}$$

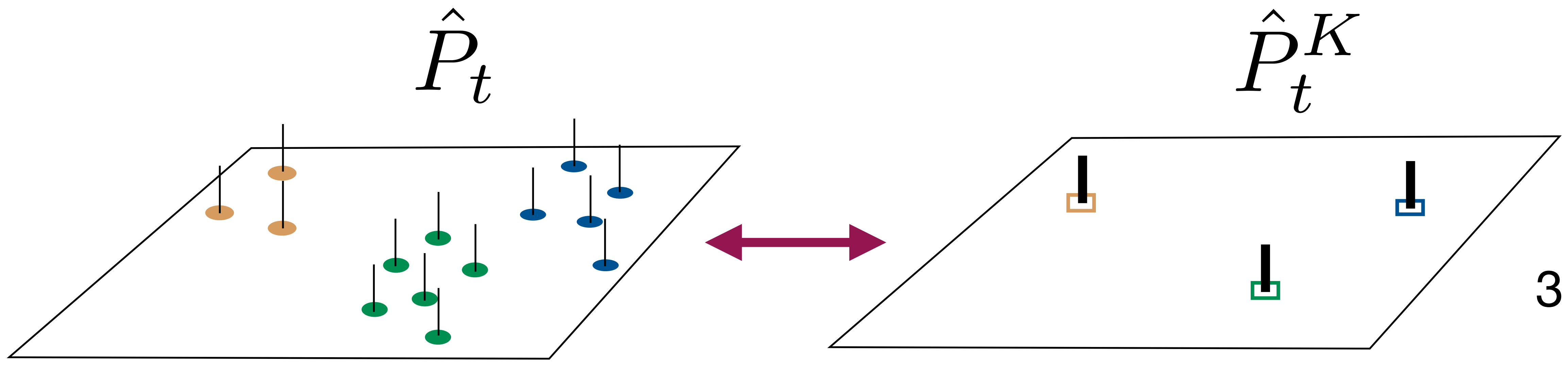
clustering
value at time $t$

$$\sum_{i=1}^{t} \sum_{k=1}^{K} a_{ik} \|d_i - \bar{d}_k\|^2$$

radius
*same as DRO*
Gao (2022)

$$\tilde{O}(1/\sqrt{t})$$

Wasserstein distance



# Computational speedups on online sparse portfolio with high confidence

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

# Computational speedups on online sparse portfolio with high confidence

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

← conditional value-at-risk

# Computational speedups on online sparse portfolio with high confidence

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

conditional value-at-risk

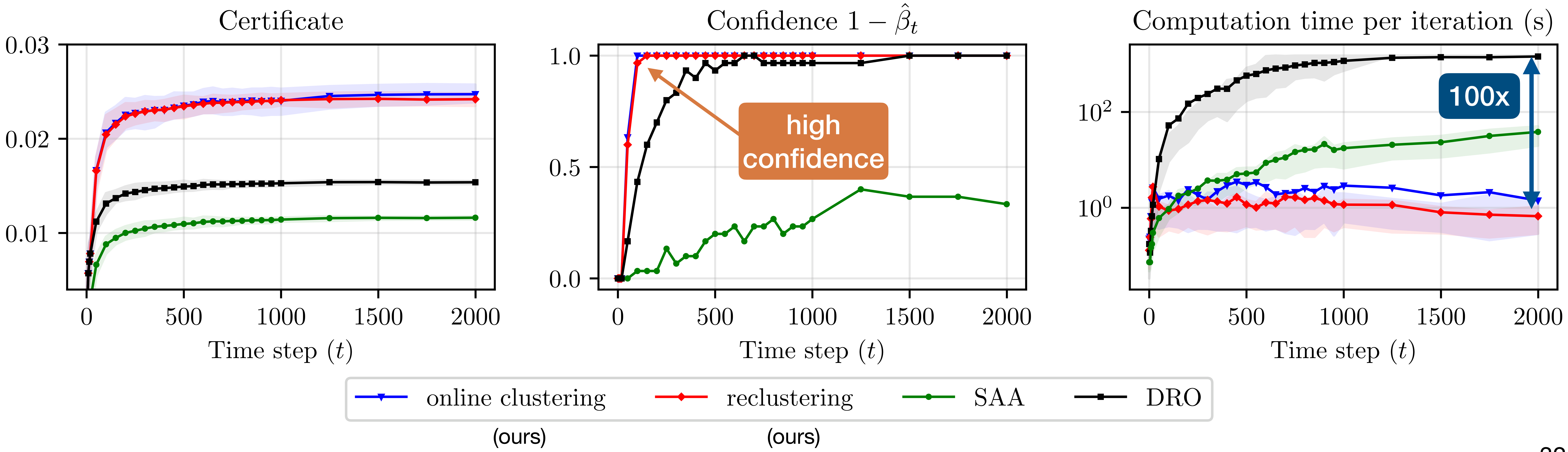
cardinality constraint

# Computational speedups on online sparse portfolio with high confidence

$$\begin{array}{ll} \text{minimize} & \mathbf{CVaR}(-u^T x, \eta) \\ \text{subject to} & \mathbf{1}^T x = 1, \quad x \geq 0 \\ & \mathbf{card}(x) \leq C \end{array}$$

conditional
value-at-risk

cardinality
constraint

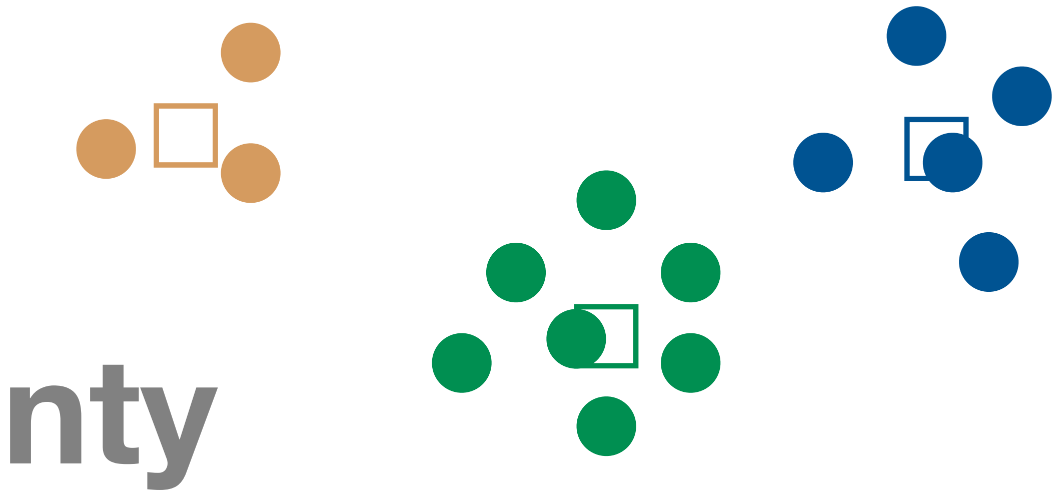


# Mean Robust Optimization

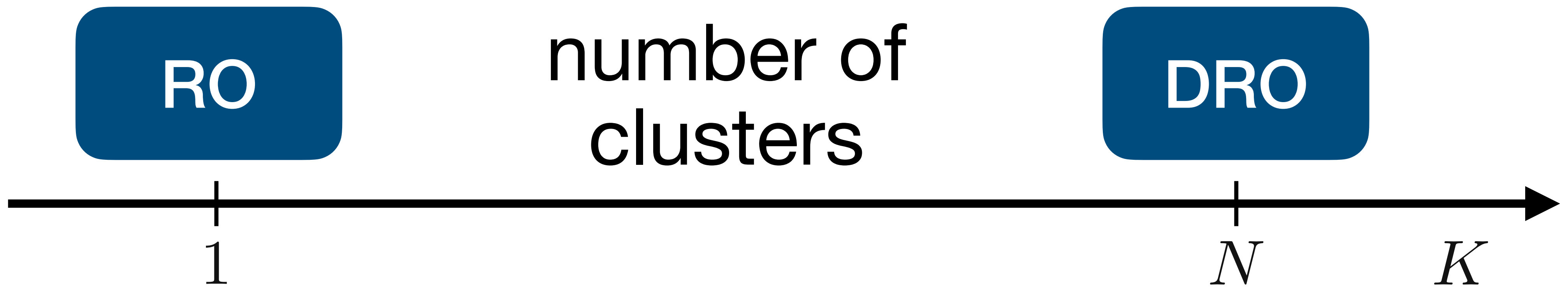
## Using data wisely in decision-making under uncertainty

# Mean Robust Optimization

## Using data wisely in decision-making under uncertainty

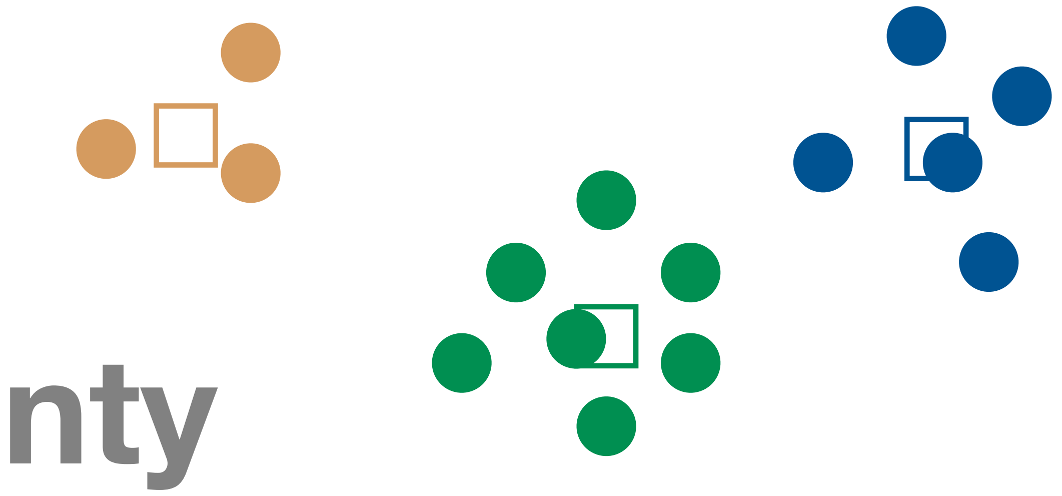


- **Bridges** RO and DRO

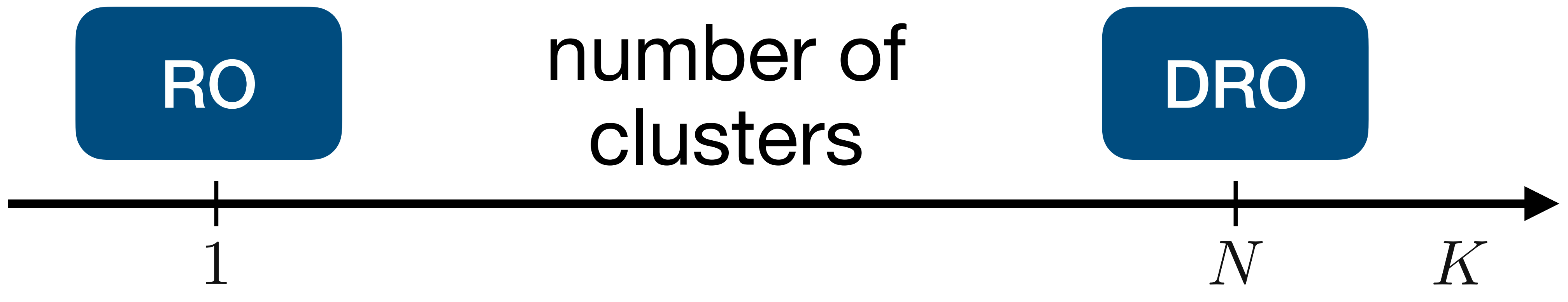


# Mean Robust Optimization

## Using data wisely in decision-making under uncertainty



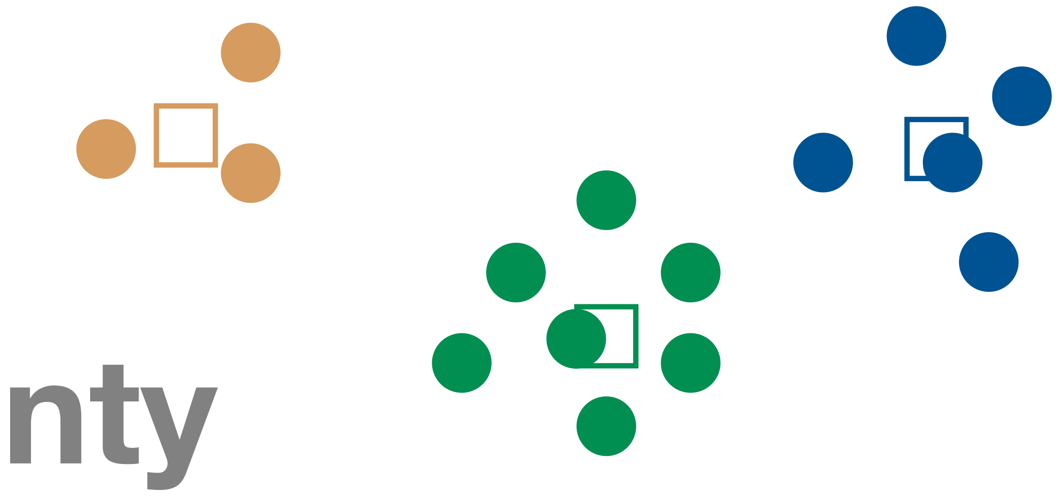
- **Bridges** RO and DRO



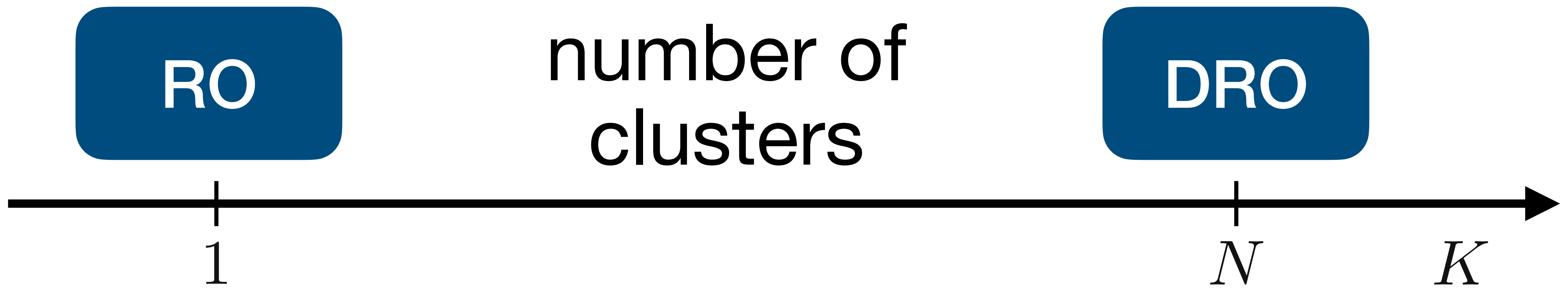
- Clustering effect
  - $g$ affine in $u$ → **zero clustering effect!**
  - $g$ (max of) concave in $u$ → **performance bound**

# Mean Robust Optimization

## Using data wisely in decision-making under uncertainty



- Bridges RO and DRO



RO

number of clusters

DRO

$1$ $N$ $K$

- Clustering effect

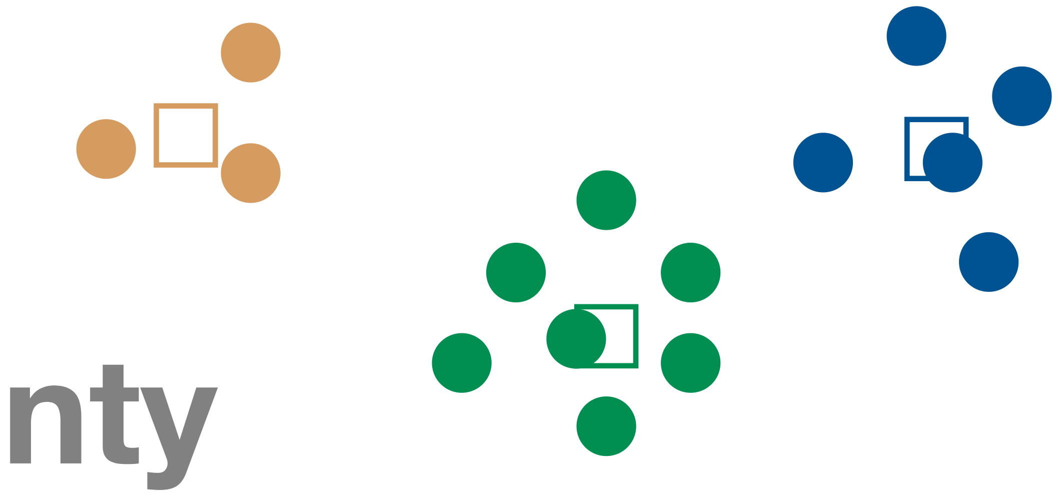
$g$ affine in $u$ → zero clustering effect!

$g$ (max of) concave in $u$ → performance bound

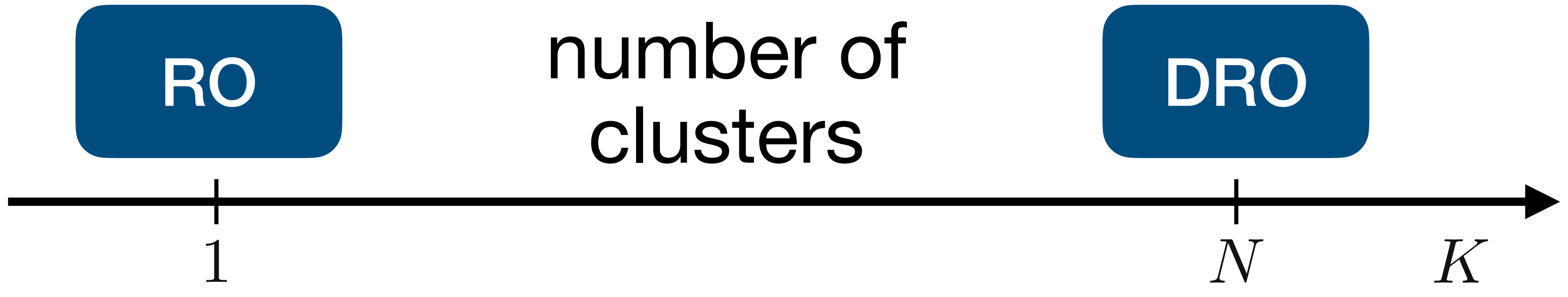
- Multiple orders of magnitude speedups

# Mean Robust Optimization

## Using data wisely in decision-making under uncertainty



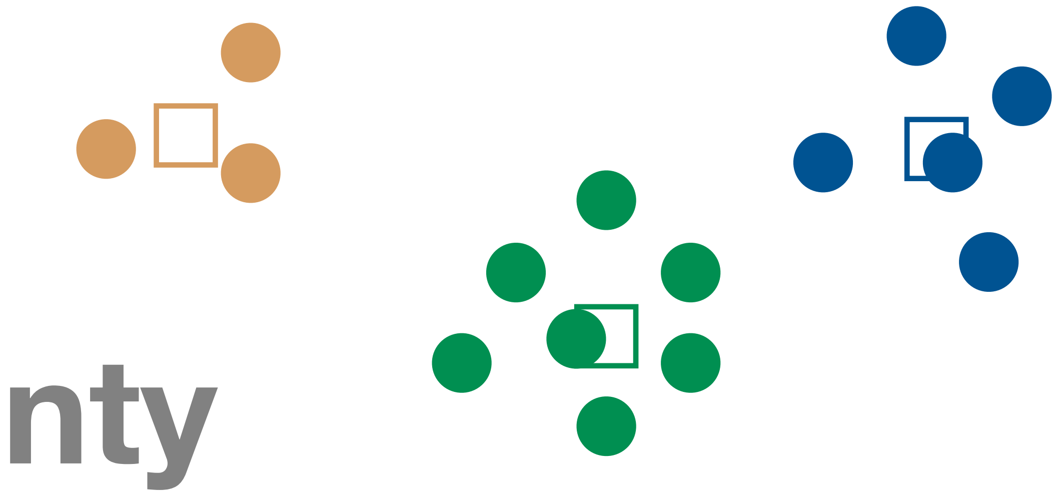
- **Bridges** RO and DRO



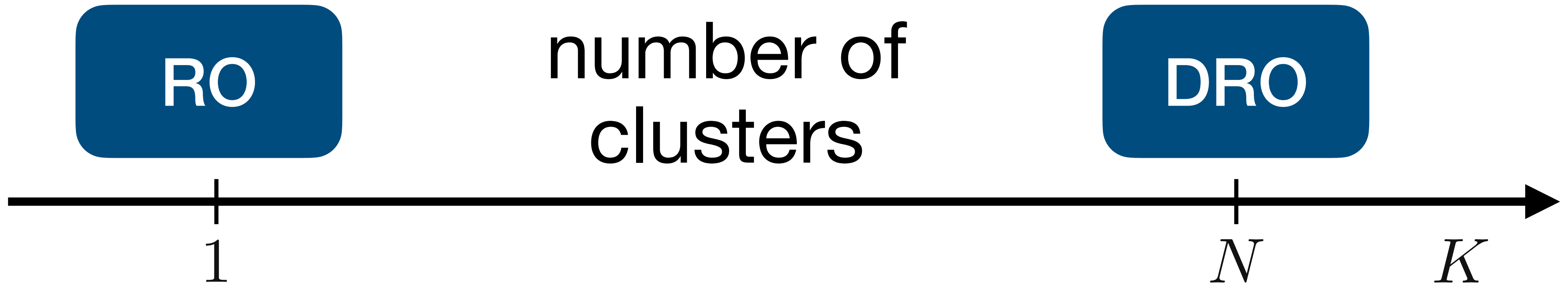
- Clustering effect
  - $g$ affine in $u$ → zero clustering effect!
  - $g$ (max of) concave in $u$ → performance bound
- Multiple **orders of magnitude** speedups
- Works well with **streaming data**

# Mean Robust Optimization

## Using data wisely in decision-making under uncertainty



- Bridges RO and DRO



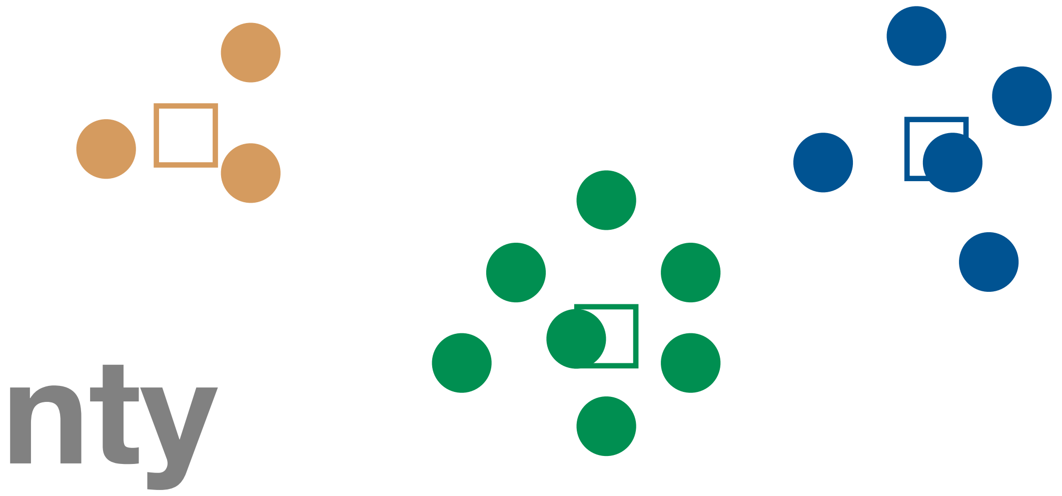
- Clustering effect
  - $g$ affine in $u$ → zero clustering effect!
  - $g$ (max of) concave in $u$ → performance bound
- Multiple orders of magnitude speedups
- Works well with streaming data

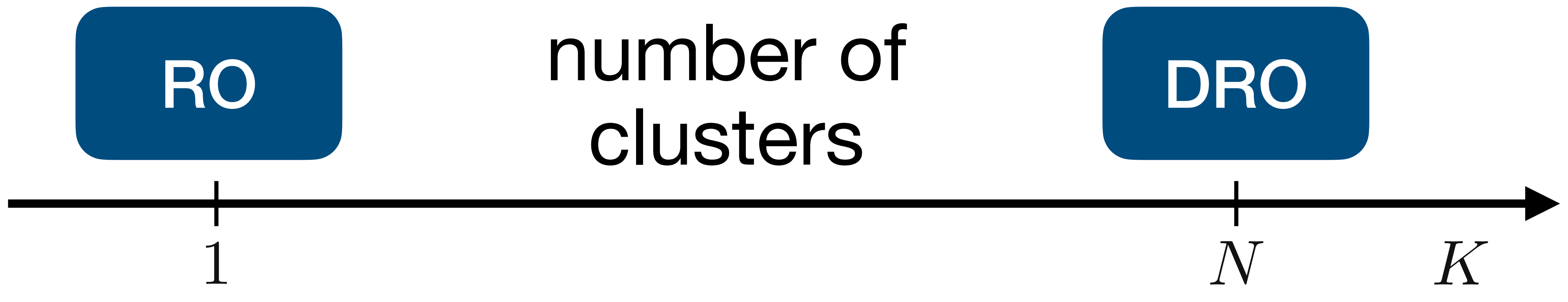
**Mean Robust Optimization**
I. Wang, C. Becker, B. Van Parys, and B. Stellato
*Mathematical Programming*, 2024
github.com/stellatogrp/mro_experiments

**Data Compression for Fast Online Stochastic Optimization**
I. Wang, M. Fochesato, and B. Stellato
*arXiv: 2504.08097*, 2025
github.com/stellatogrp/online_mro

# Mean Robust Optimization

## Using data wisely in decision-making under uncertainty





- **Bridges** RO and DRO
- Clustering effect
  - $g$ affine in $u$ → **zero clustering effect!**
  - $g$ (max of) concave in $u$ → **performance bound**
- Multiple **orders of magnitude** speedups
- Works well with **streaming data**

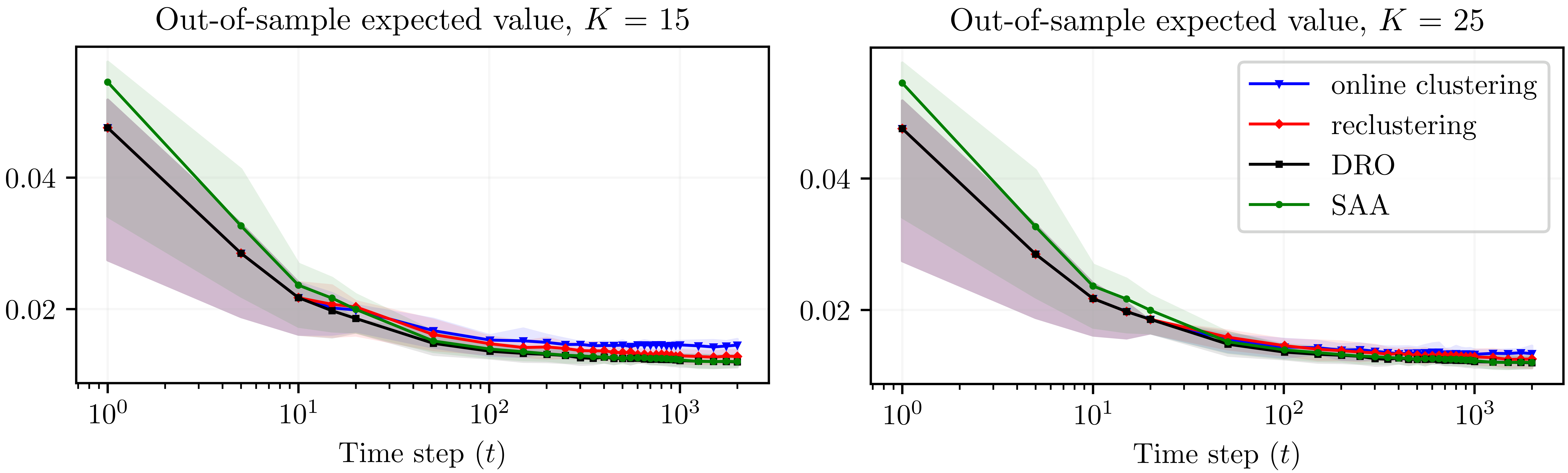
**Mean Robust Optimization**
I. Wang, C. Becker, B. Van Parys, and B. Stellato
*Mathematical Programming*, 2024
`github.com/stellatogrp/mro_experiments`

**Data Compression for Fast Online Stochastic Optimization**
I. Wang, M. Fochesato, and B. Stellato
*arXiv: 2504.08097*, 2025
`github.com/stellatogrp/online_mro`

@stellato.io bstellato@princeton.edu stellato.io

# Backup

# Online out-of-sample performance


Out-of-sample expected value, $K = 15$

Out-of-sample expected value, $K = 25$

**Obtain near-optimal out-of-sample values**