# Learning for Robust Optimization

## Joint work with Irina Wang, Cole Becker, and Bart Van Parys

PRINCETON UNIVERSITY ORFE

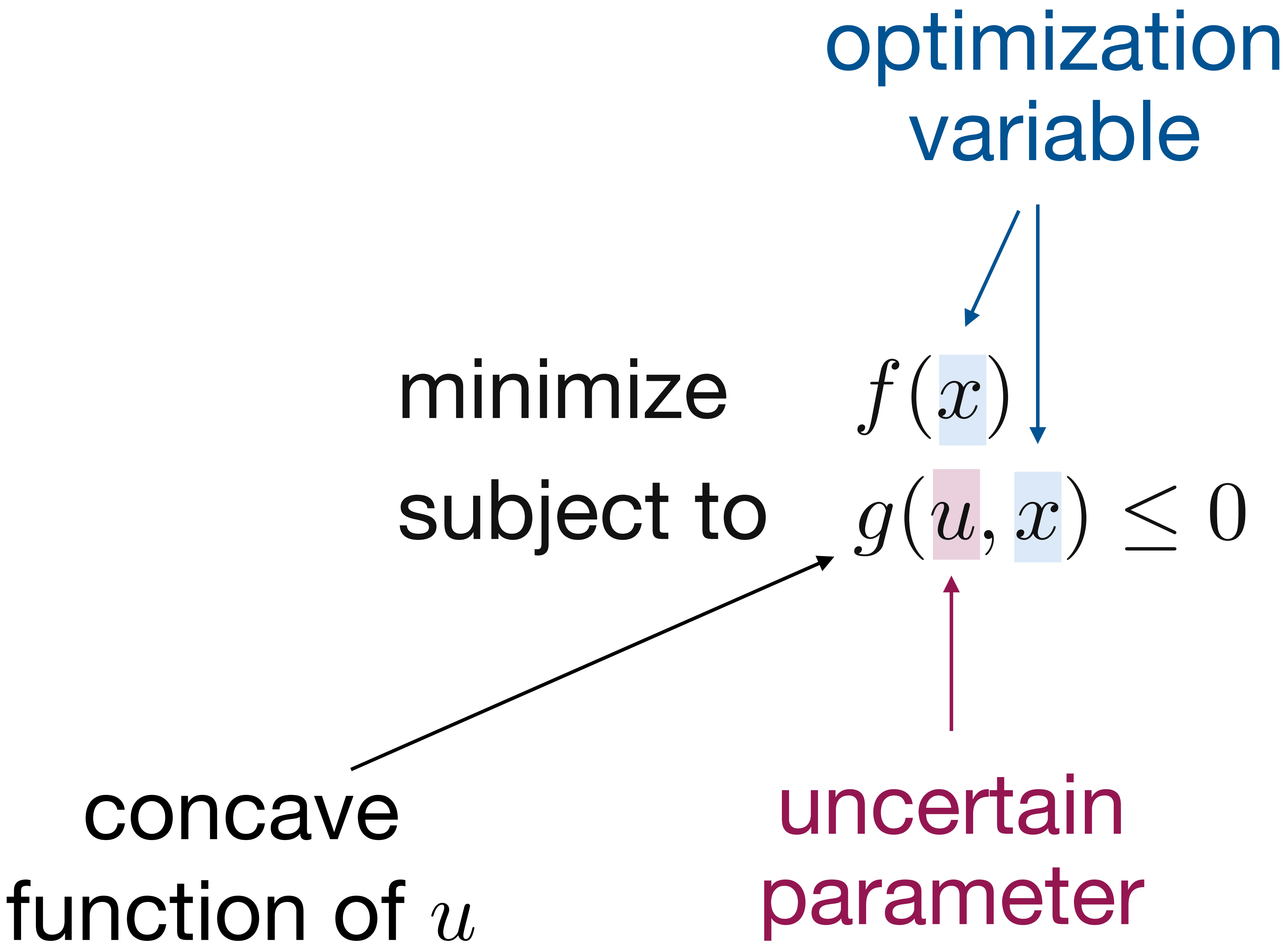
**Bartolomeo Stellato — Conference on Information Sciences and Systems, March 22 2023**

# Decision-making with uncertainty is hard


COVID-19


Finance


Energy


optimization variable

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

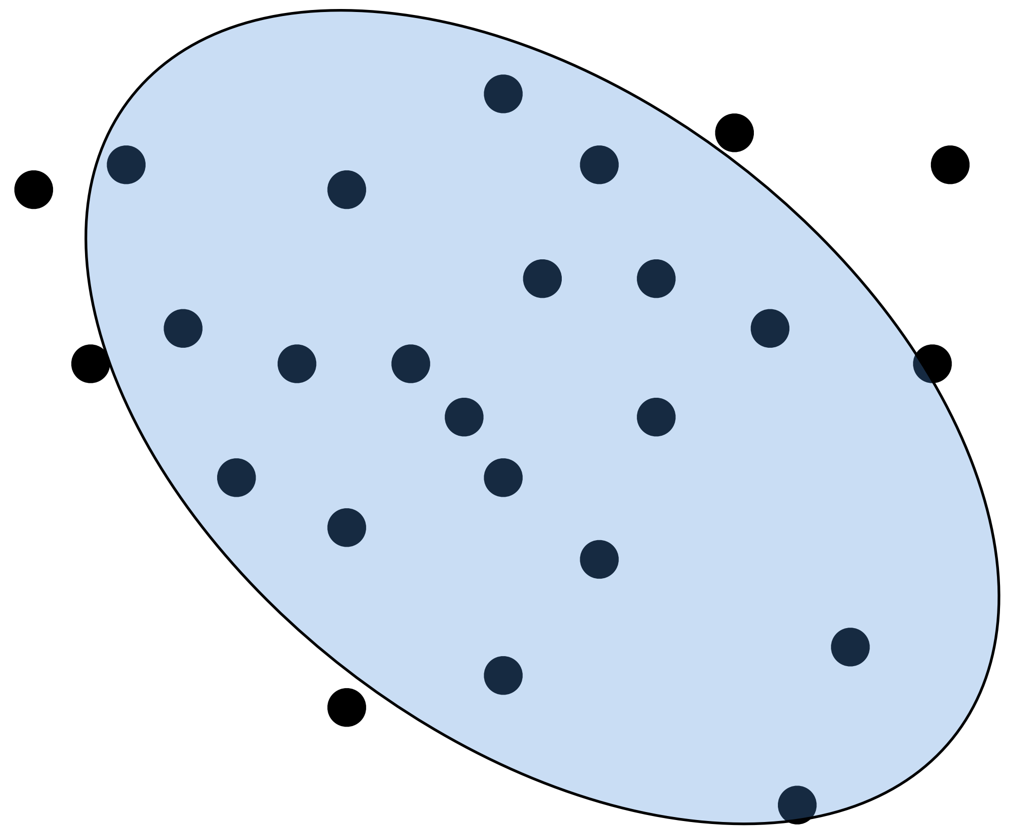
concave function of $u$

uncertain parameter

**We want to guarantee constraint satisfaction**

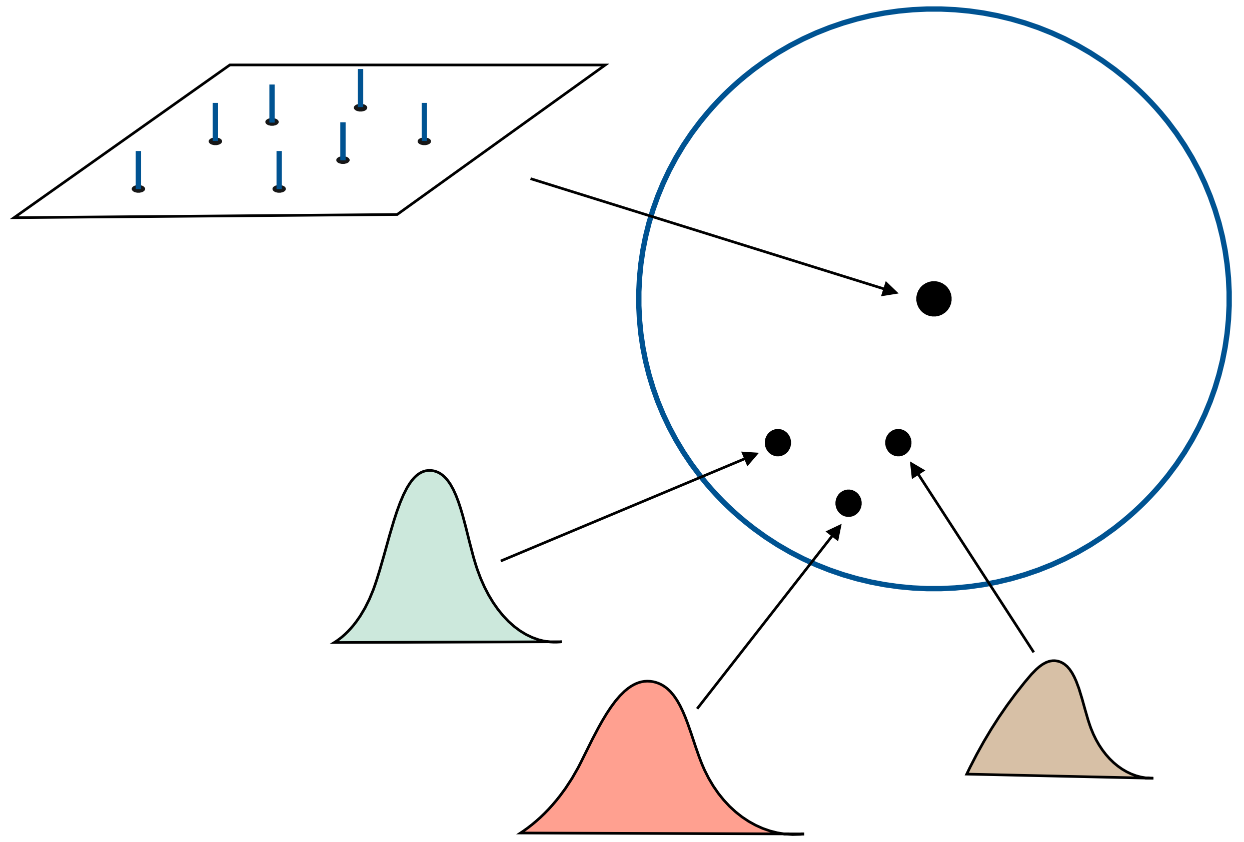
# Robust vs Distributionally Robust Optimization

## Robust optimization (RO)



- ✓ Tractable, expressive
- ✗ Can be conservative

## Distributionally Robust Optimization (DRO)



- ✓ Can be less conservative
- ✗ Computationally expensive

Can we get the best of both worlds?

# Probabilistic guarantees

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

$u \sim P$

(but we never know $P$!)

Data

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

**Data-driven probabilistic guarantees**

Product Distribution

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

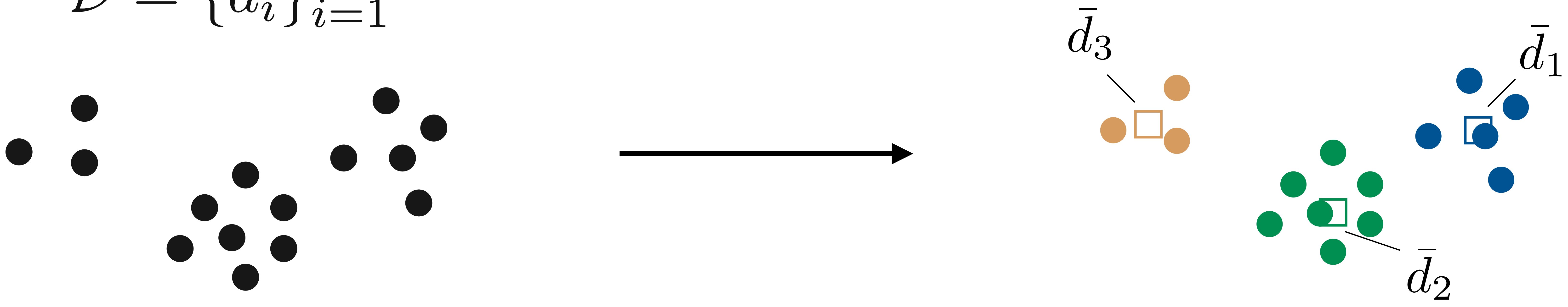
probability of constraint satisfaction

data-driven solution

# Machine Learning Clustering

Data

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

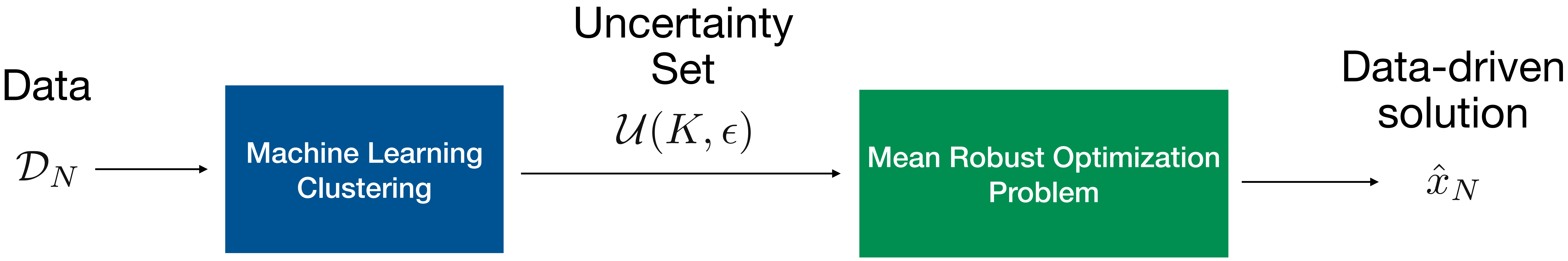


**Goal**

minimize $$\sum_{k=1}^{K} \sum_{i \in C_k} \|d_i - \bar{d}_k\|^2$$

cluster centers

# Mean Robust Optimization (MRO)



Data

$\mathcal{D}_N$

**Machine Learning Clustering**

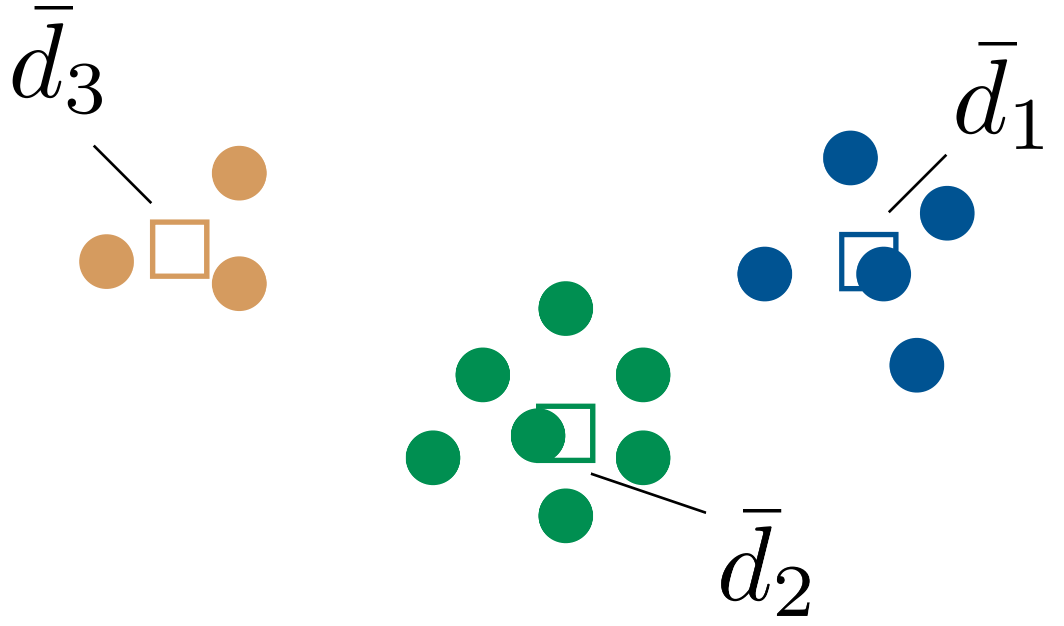
Uncertainty Set

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

**Mean Robust Optimization Problem**

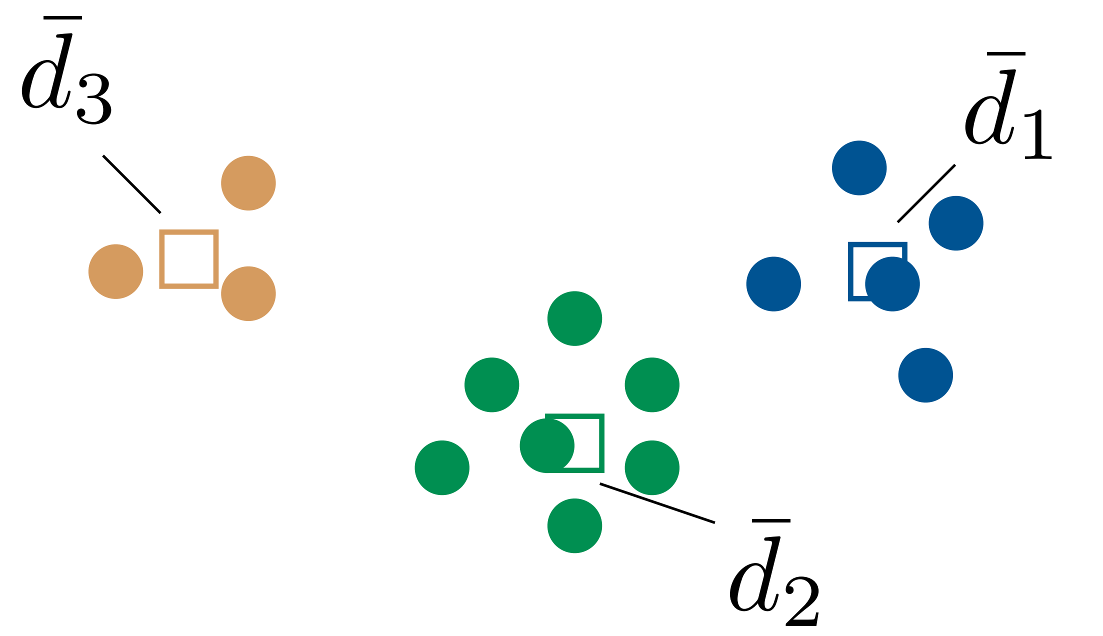
Data-driven solution

$\hat{x}_N$



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

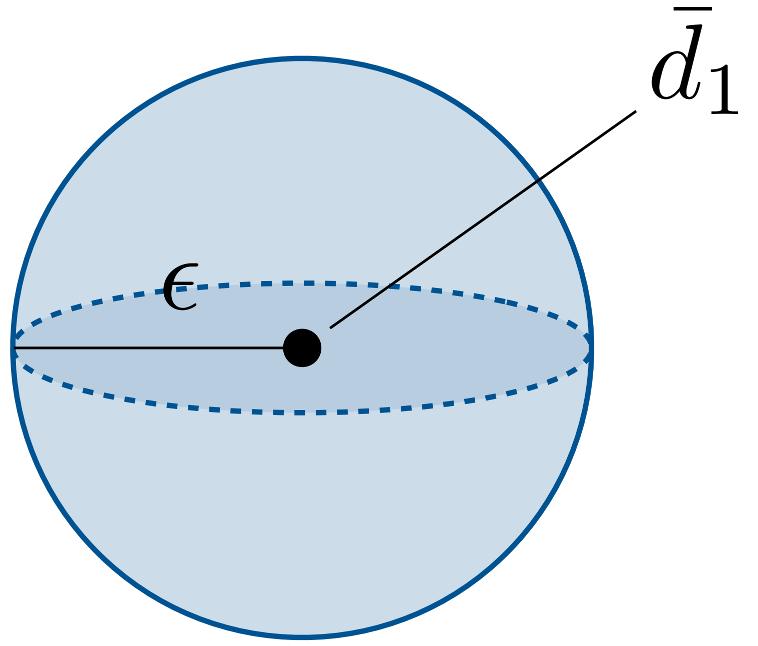
# Uncertainty set



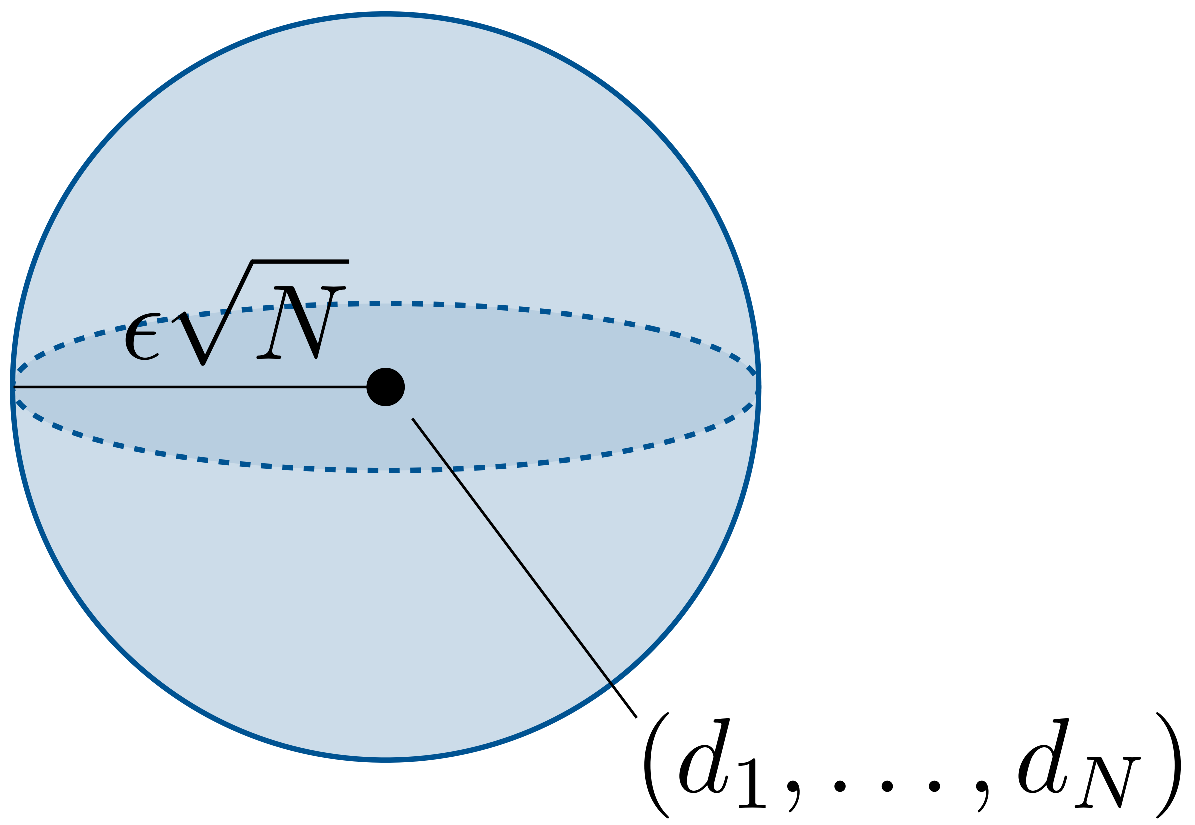
$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1,\dots,v_K) \;\middle|\; \sum_{k=1}^{K} \overset{\text{cluster weights}}{w_k} \|v_k - \underset{\text{cluster centers}}{\bar{d}_k}\|^{\overset{\text{order}}{p}} \le \epsilon^{\overset{\text{order}}{p}} \right\}$$

## Examples

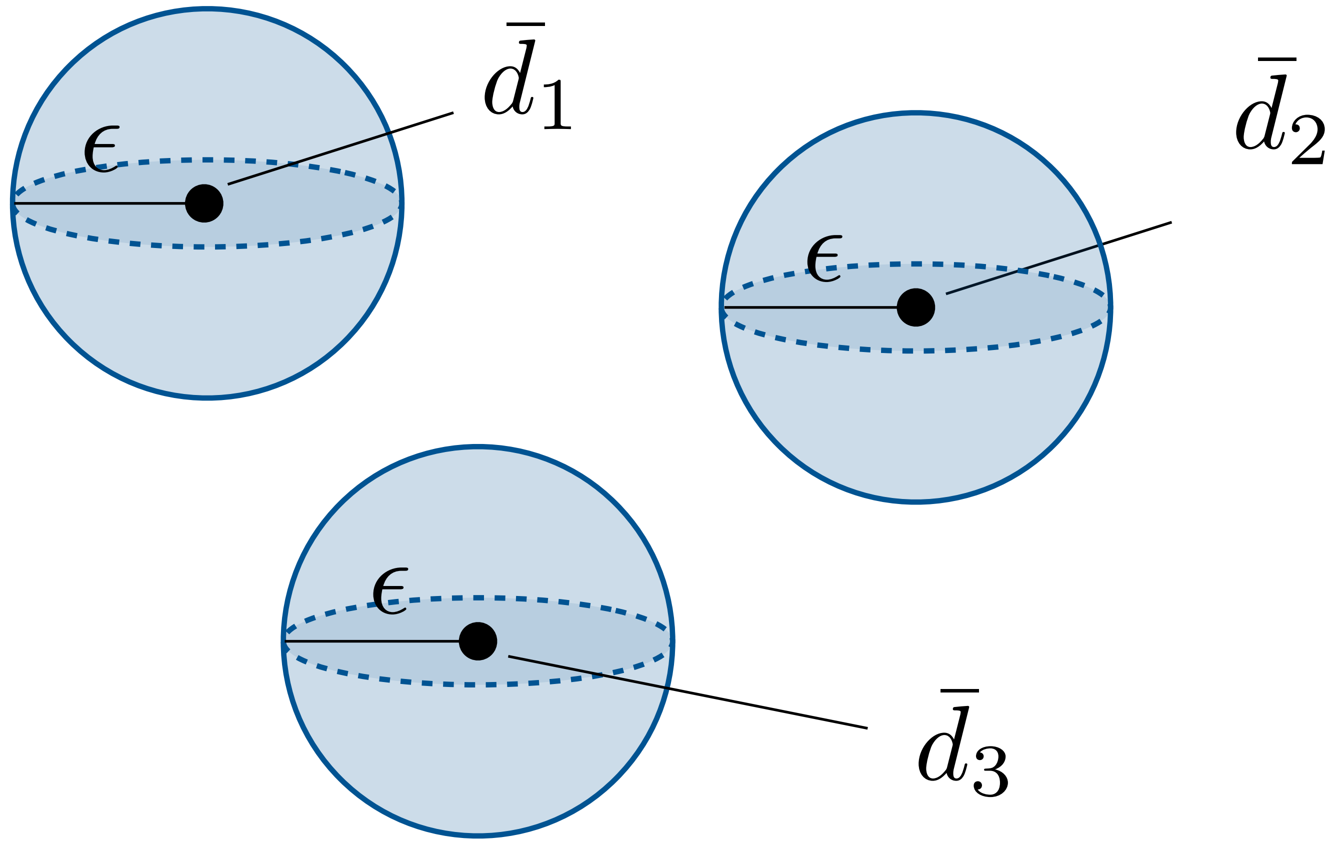
$K = 1$



$K = N, p = 2$



$K = 3, p = \infty$



# 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

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

$$\longrightarrow$$

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

constraint function

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

# Solving the MRO problem

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

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

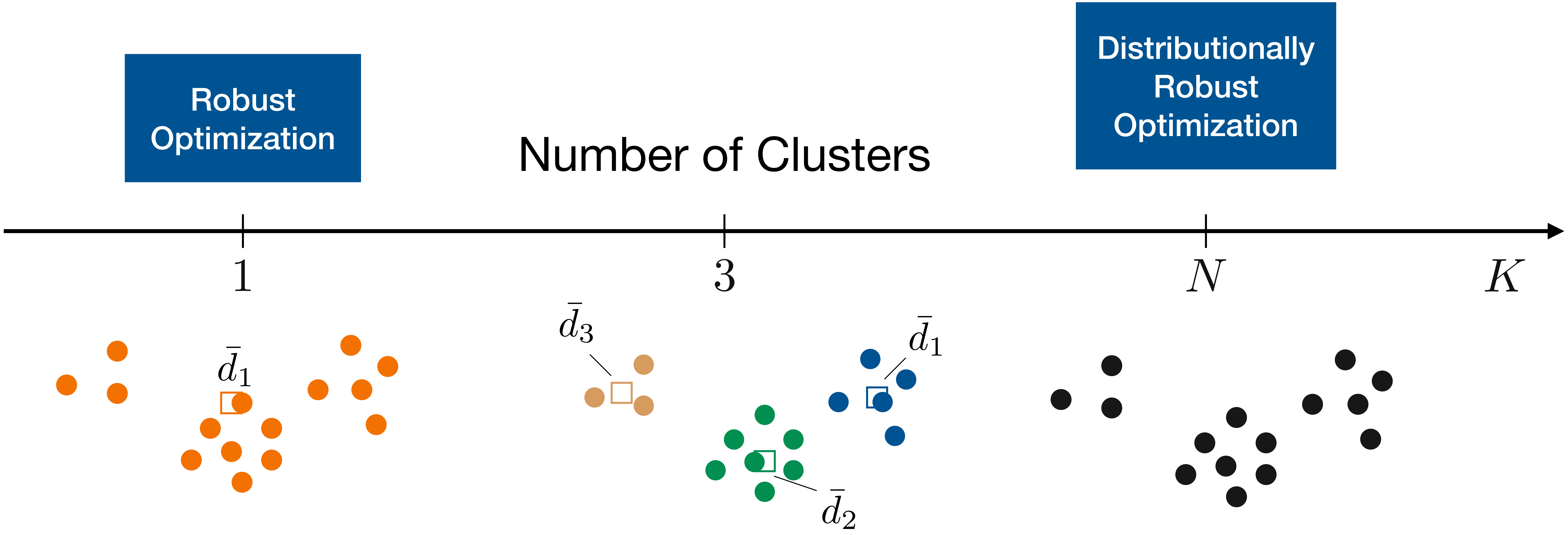
conjugate function

cluster centers

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

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

# MRO bridges between RO and DRO



**Robust Optimization**

Number of Clusters

**Distributionally Robust Optimization**

---

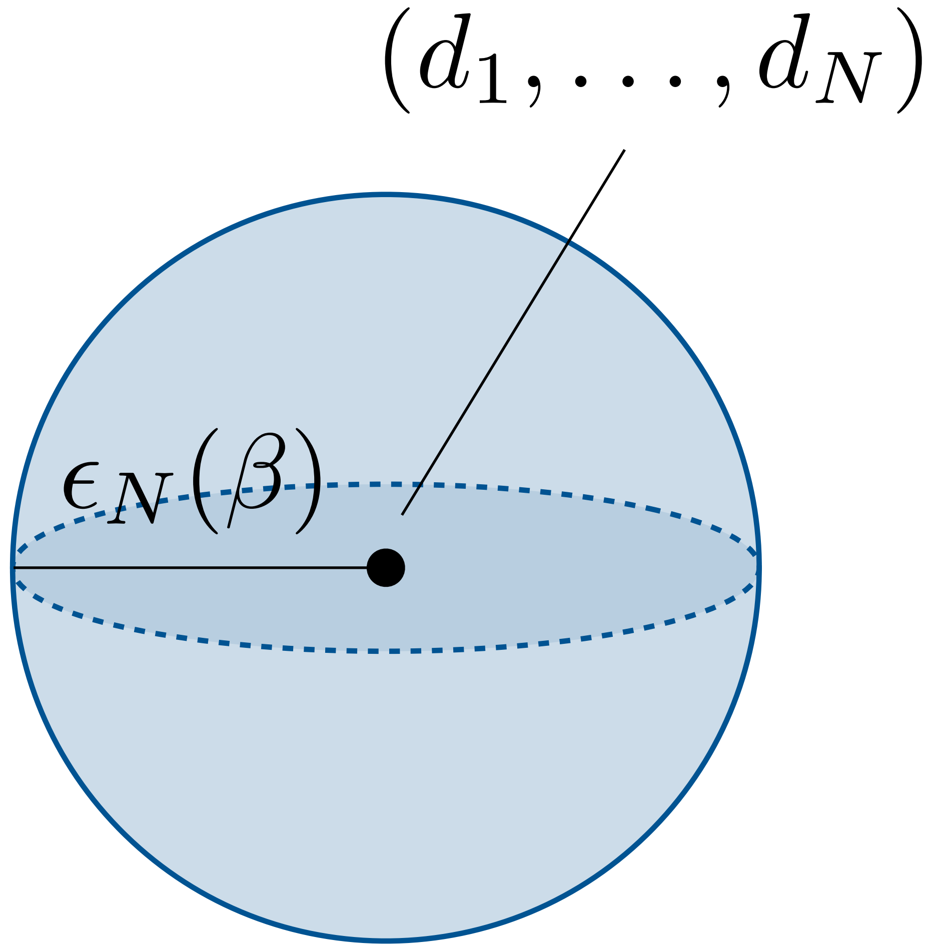
[KENSA19] D Kuhn, P M Esfahani, V A Nguyen, and S Shafieezadeh-Abadeh, “Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning

# Satisfying the probabilistic guarantees

probability of constraint satisfaction

uncertainty set radius

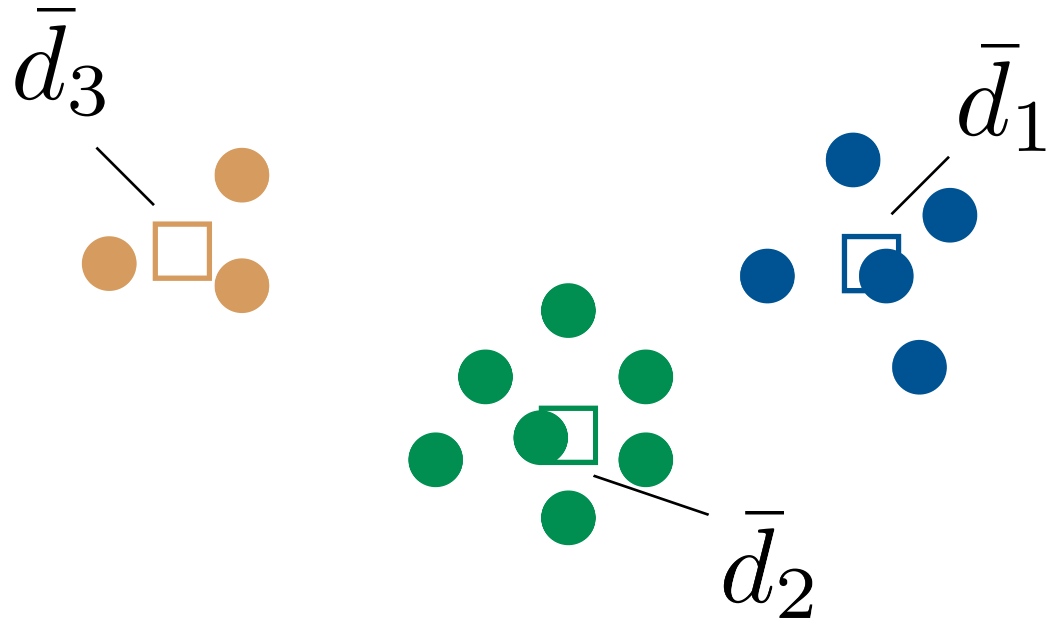
$$\mathbf{P}^N\left(\mathbf{E}(g(u,\hat{x}_N)) \le 0\right) \ge \boxed{1-\beta} \xrightarrow{\text{light-tailed}} \mathcal{U}(N, \boxed{\epsilon_N(\beta)})$$



### MRO clustering

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

$$\frac{1}{N}\sum_{k=1}^{K}\sum_{d_i\in C_k}\|d_i-\bar{d}_k\|^p$$



Quite conservative bounds… can we do better?

# Bounding the conservatism

Worst-case values

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

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

MRO constraint

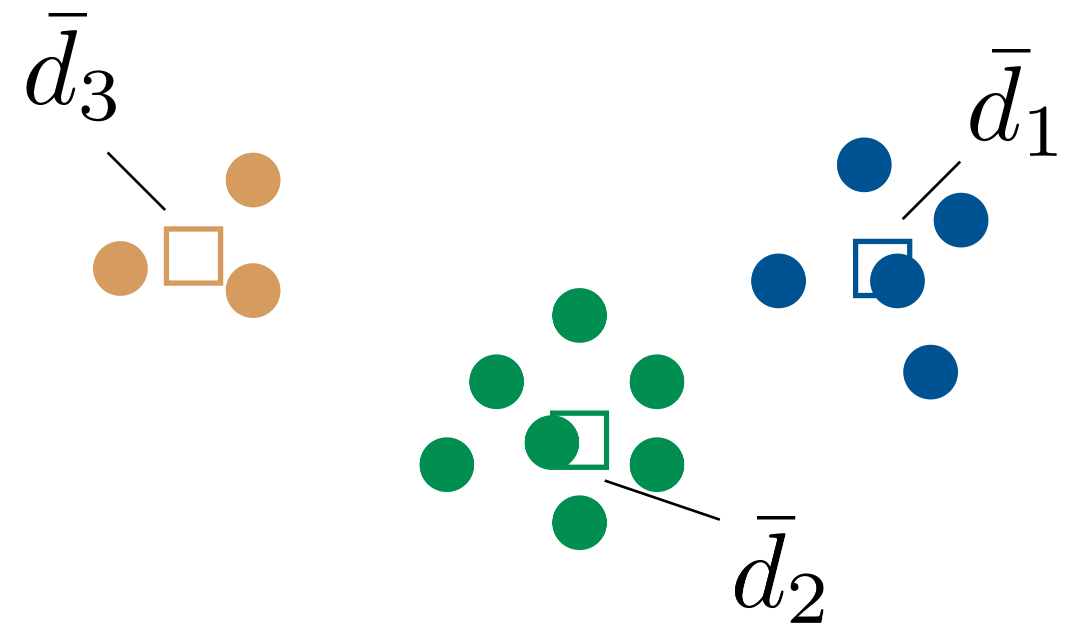
$$\bar{g}(u,x) \leq 0 \quad \forall u \in \mathcal{U}(K,\epsilon)$$

**Theorem**

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

clustering objective

$$\bar{g}^{N}(x) \leq \bar{g}^{K}(x) \leq \bar{g}^{N}(x) + \frac{L}{2} D(K) \longleftarrow \min \ \frac{1}{N} \sum_{k=1}^{K} \sum_{d_i \in C_k} \|d_i - \bar{d}_k\|^2$$



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

# Capital budgeting example

**Problem**

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

total net present value (NPV)

budget constraint

$$g(u, x, \tau) = -\eta(u)^T x - \tau$$

**MRO formulation**

$$\begin{array}{ll} \text{minimize} & \tau \\ \text{subject to} & \bar{g}(u, x, \tau) \leq 0, \quad \forall u \in \mathcal{U}(K, \epsilon) \\ & a^T x \leq b \\ & x \in \{0,1\}^n \end{array}$$

NPV of project $j$

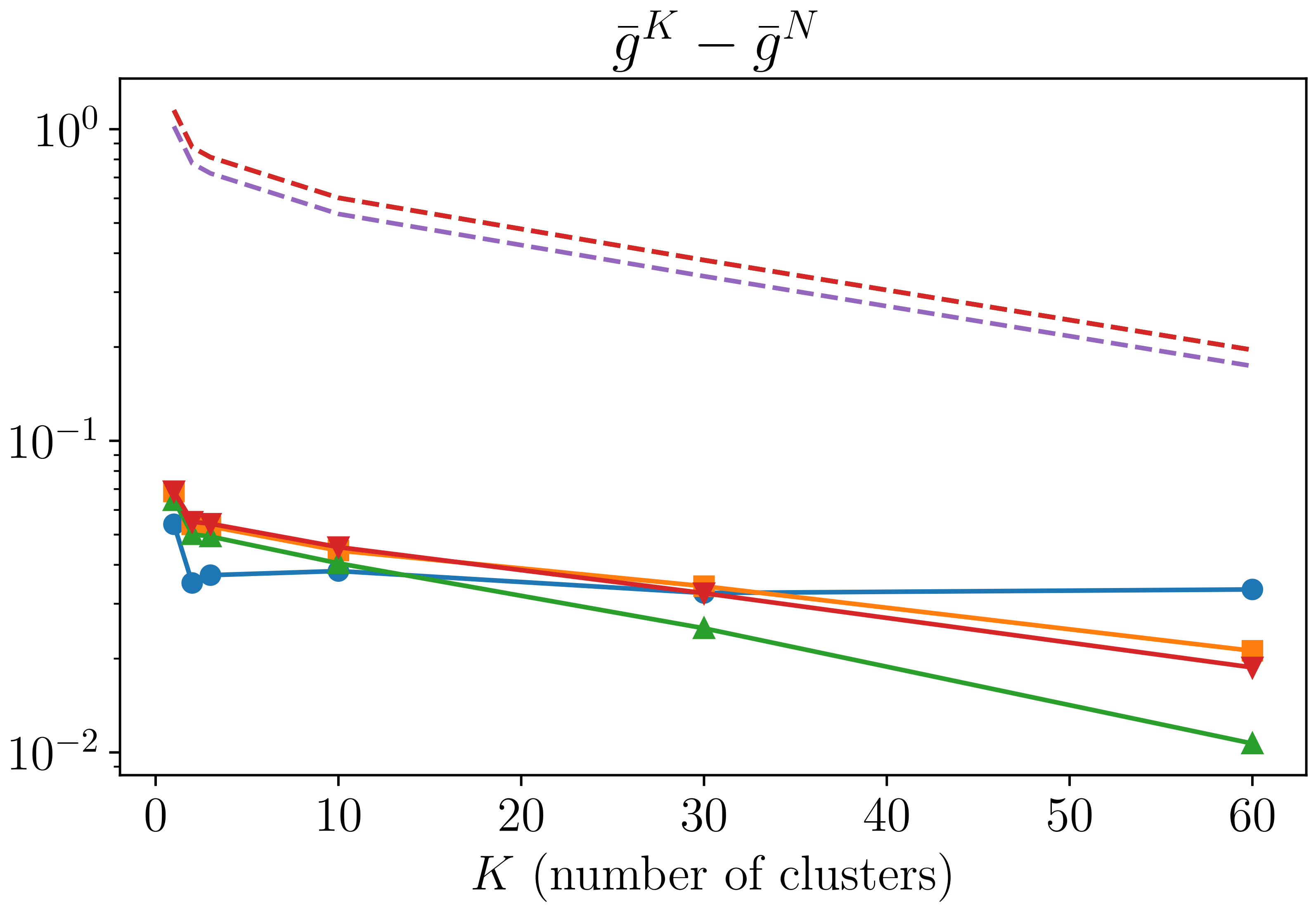
$$\eta_j(u) = \sum_{t=1}^{T} \frac{F_{jt}}{(1 + u_j)^t}$$

cash flow

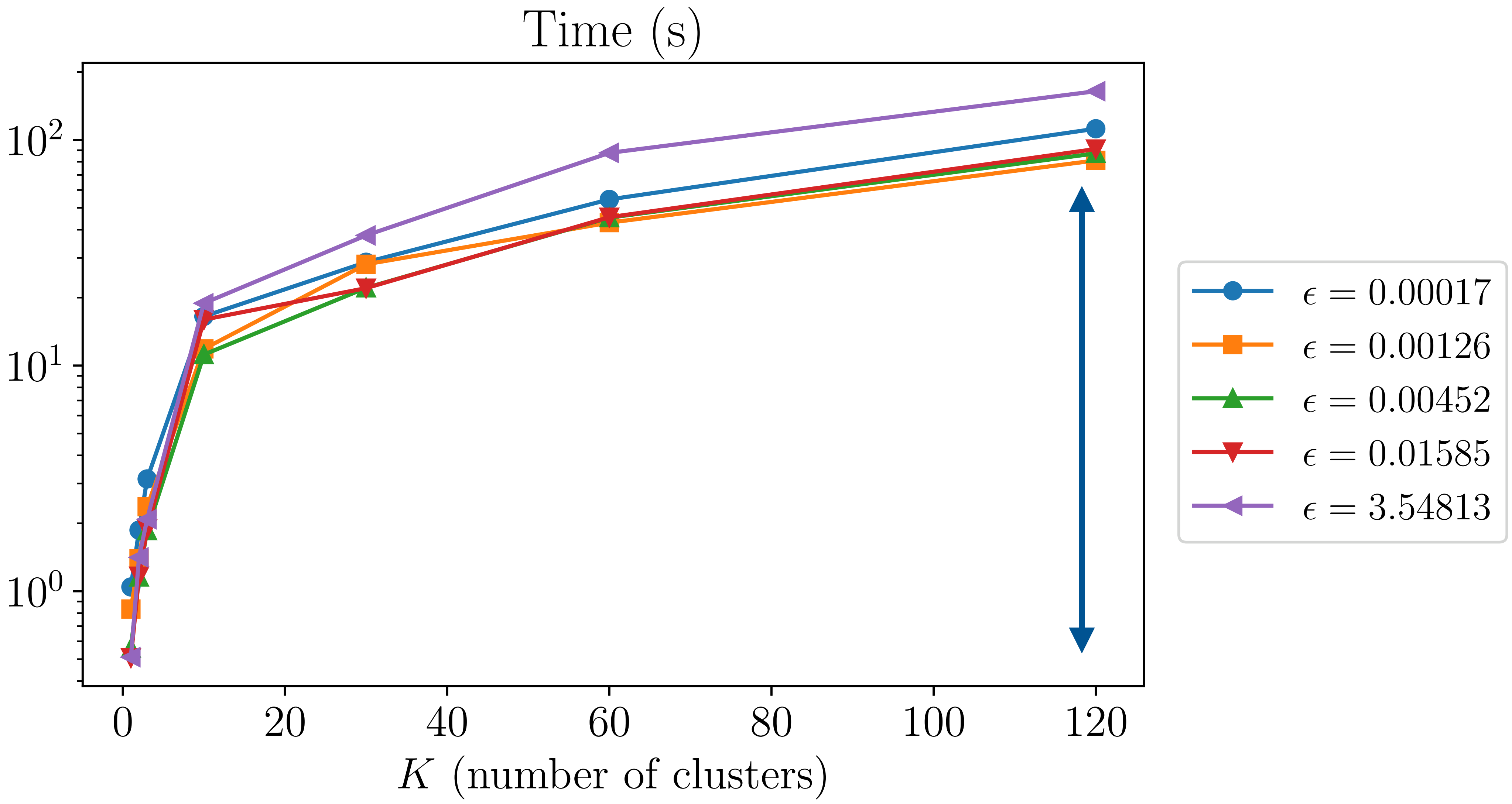
discount rate

# Capital budgeting results

$n = 20, N = 120, T = 5$

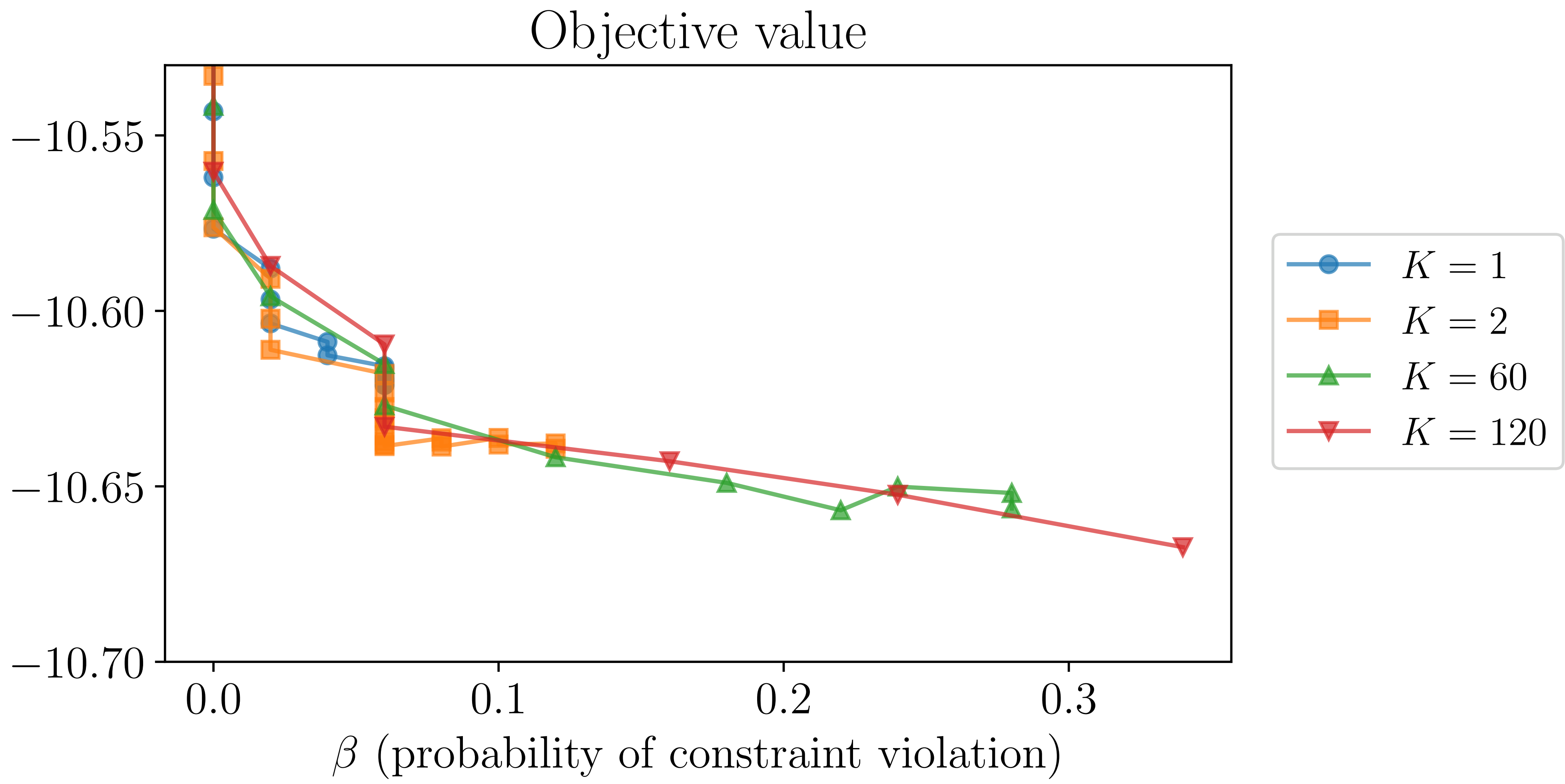


$\bar{g}^K - \bar{g}^N$



Time (s)

100x speedups!



Objective value

2 clusters give near-optimal performance

# Mean Robust Optimization

- **Bridge** between RO and DRO

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

- Multiple **orders of magnitude** speedups

https://github.com/stellatogrp/mro_experiments

arXiv https://arxiv.org/abs/2207.10820 ←

**INFORMS Computing Society Student Paper Award Winner**

stellato.io bstellato@princeton.edu @b_stellato