Learning for Decision-Making under Uncertainty

Bartolomeo Stellato

Department of Operations Research and Financial Engineering
Department of Electrical and Computer Engineering
Center for Statistics and Machine Learning

PRINCETON
UNIVERSITY

Bartolomeo Stellato — MIP Workshop, May 25 2023

🌐 `stellato.io`
✉ `bstellato@princeton.edu`

It is hard to make decisions under uncertainty

Transportation

Finance

Energy

It is hard to make decisions under uncertainty

Transportation

Finance

Energy

But we have data!

Problem setup with uncertain constraints

Problem setup with uncertain constraints

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

Problem setup with uncertain constraints

optimization
variable

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

Problem setup with uncertain constraints

optimization
variable

minimize $f(x)$

subject to $g(u, x) \leq 0$

uncertain
parameter

Problem setup with uncertain constraints

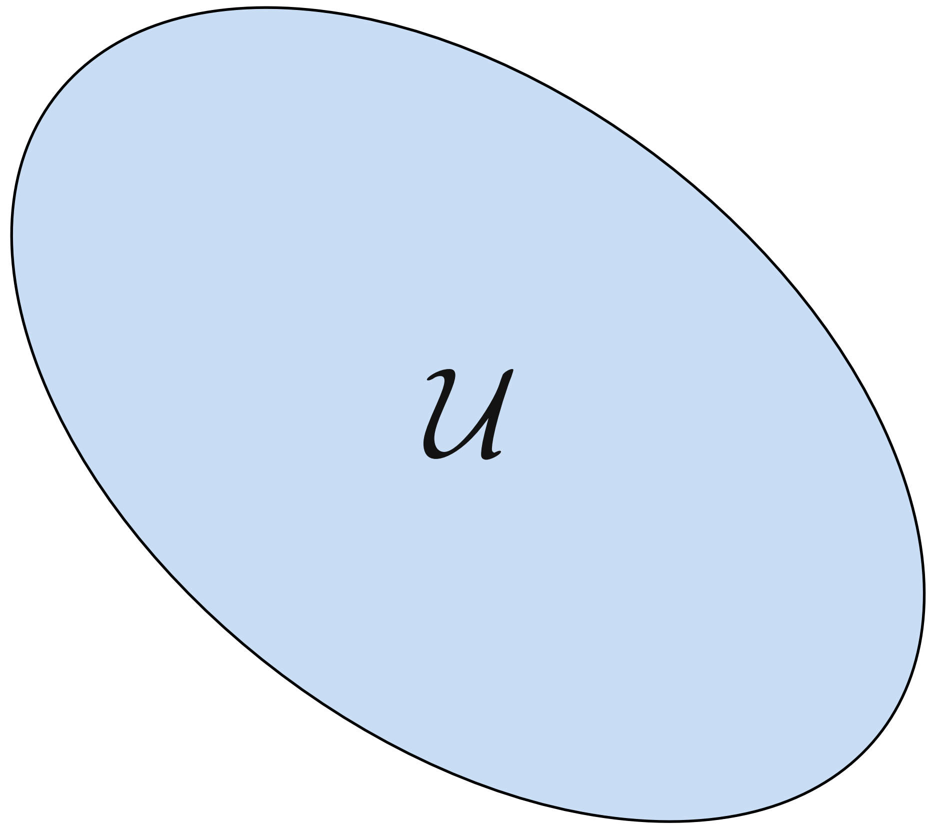
optimization
variable

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

uncertain
parameter

We want to guarantee constraint satisfaction

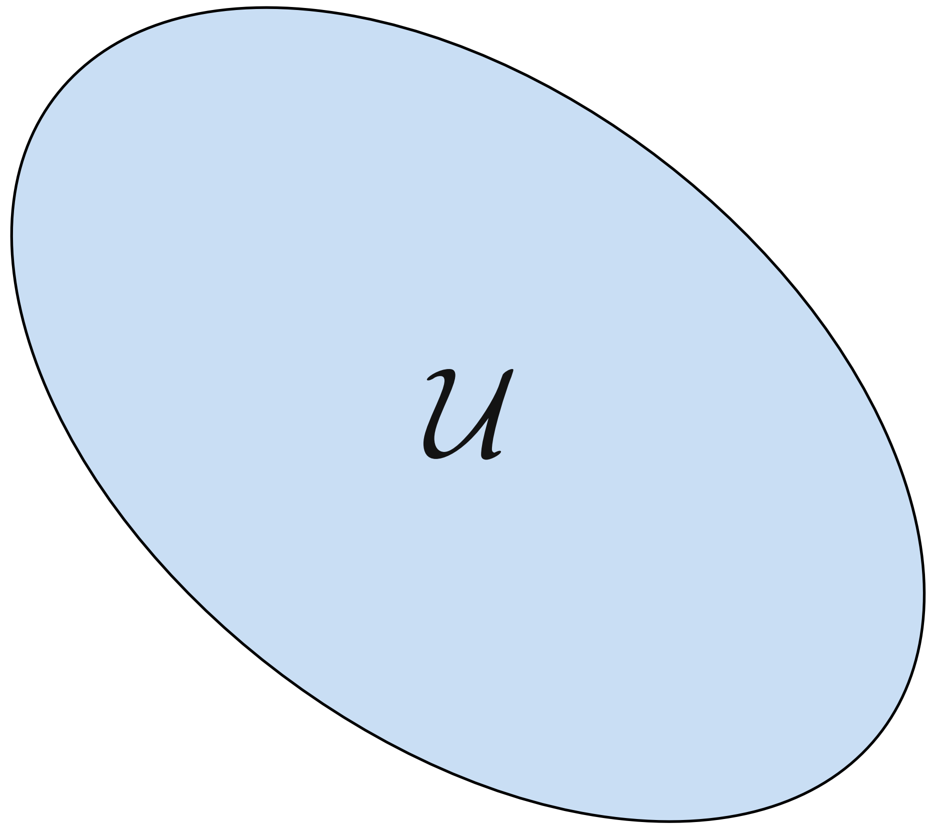
Robust optimization recipe



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

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

Robust optimization recipe



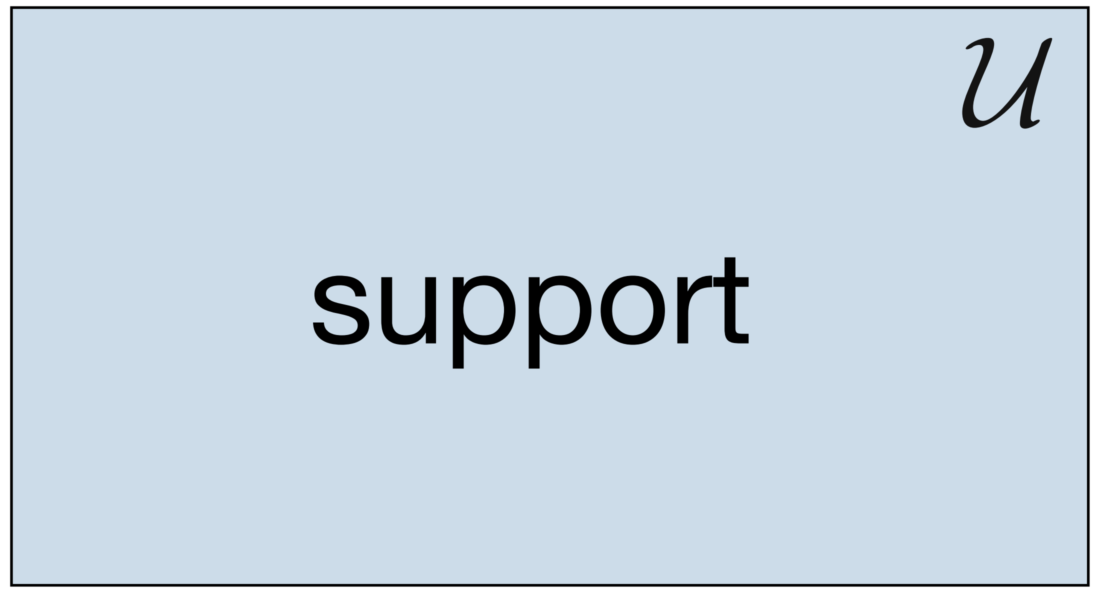
1. Pick uncertainty set $\mathcal{U}$
2. Ensure constraint satisfaction $\forall u \in \mathcal{U}$

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

How do we pick the uncertainty set?

Picking the uncertainty set is difficult

Worst-case approach



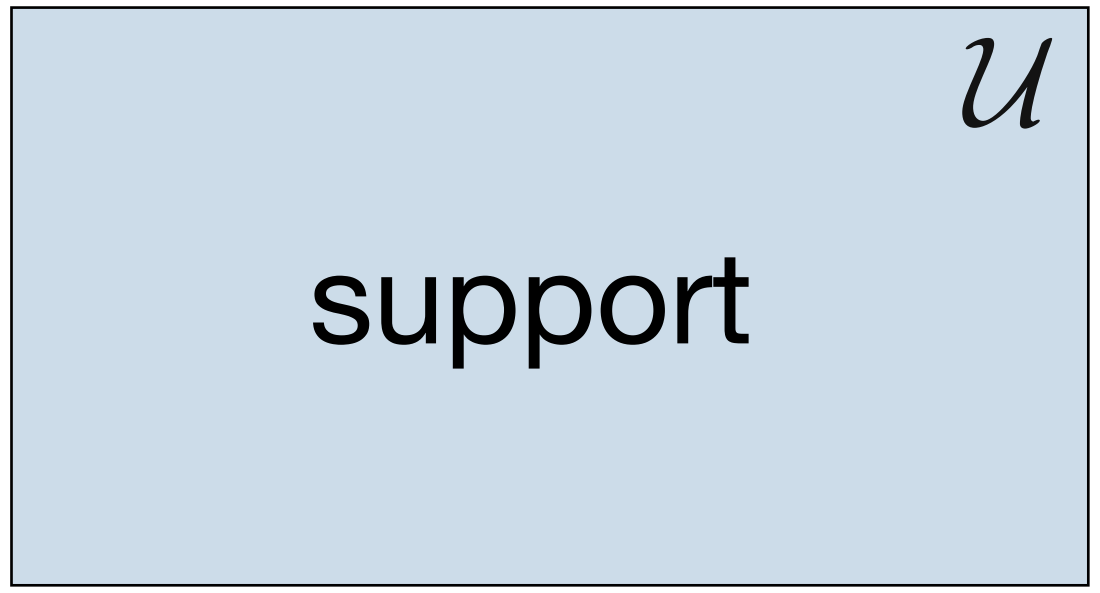
$\mathcal{U}$

support

❌ Very conservative

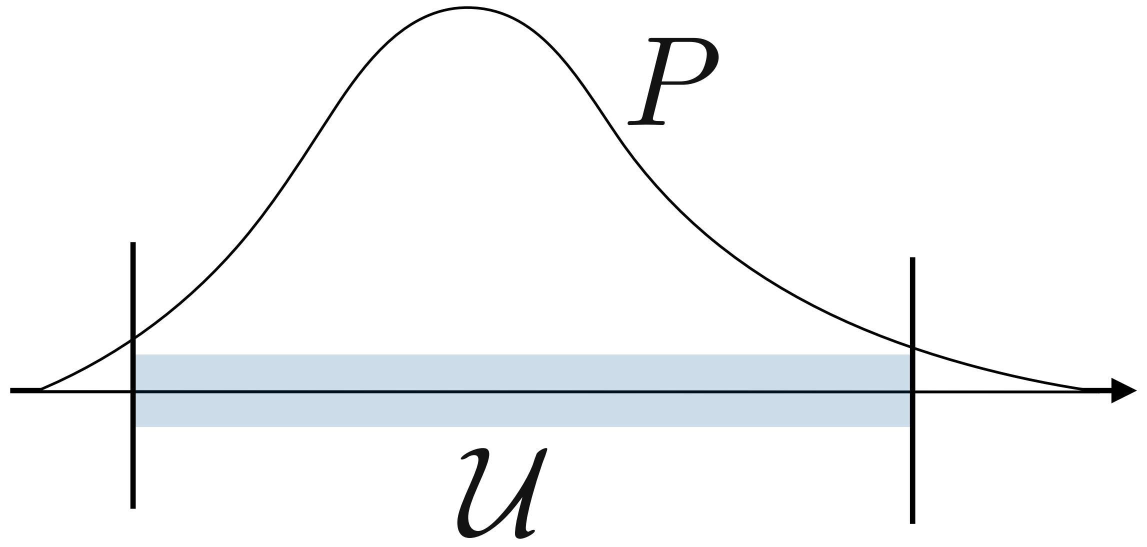
Picking the uncertainty set is difficult

Worst-case approach



❌ Very conservative

Probabilistic approach

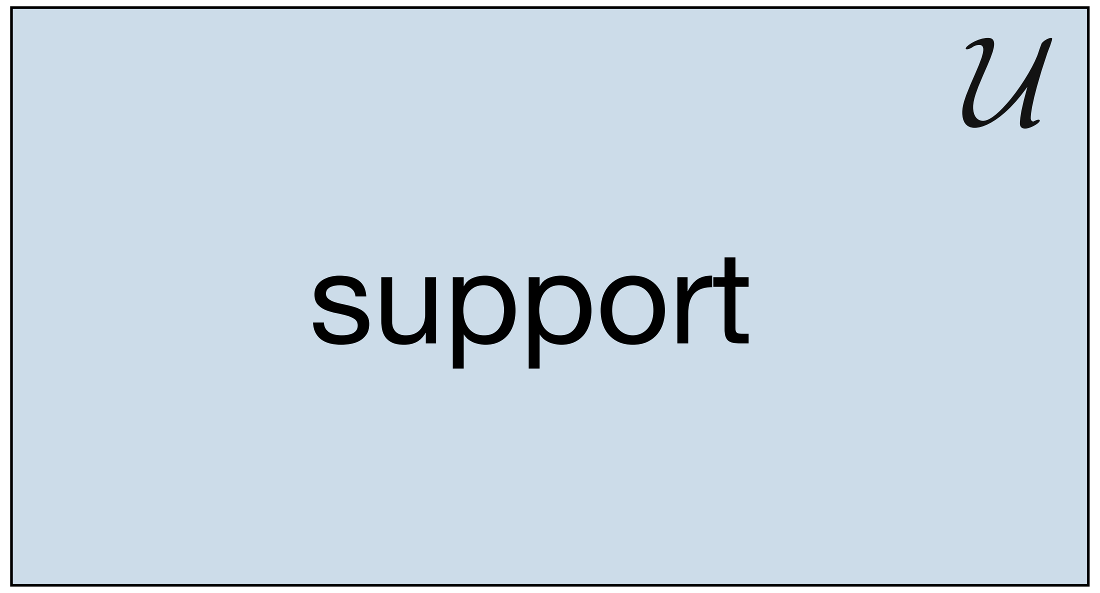


❌ nobody knows P

“Probabilistic Guarantees in Robust Optimization”, D. Bertsimas, D. den Hertog, and J. Pauphilet (2019)

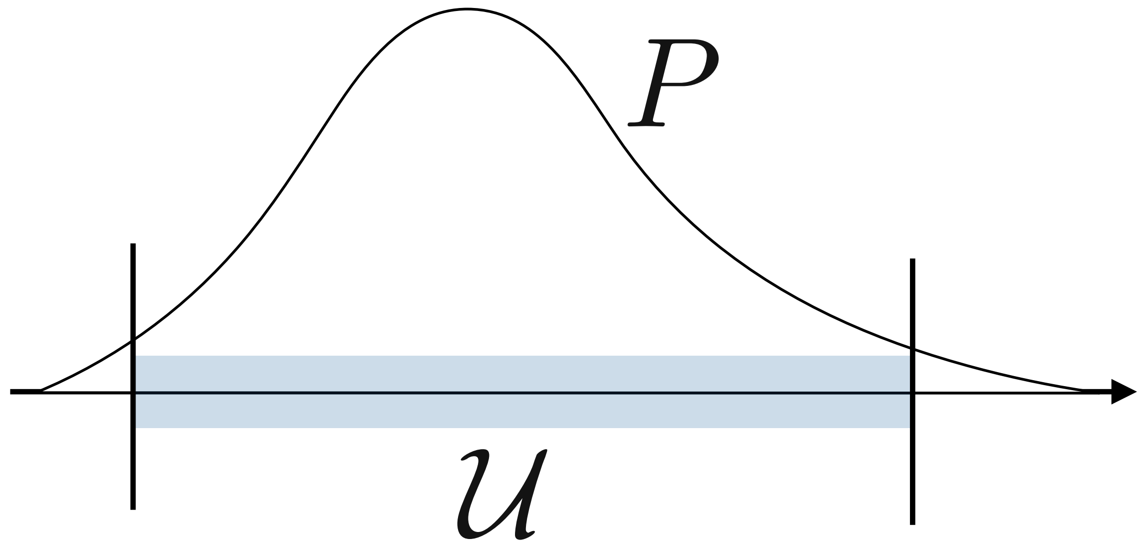
Picking the uncertainty set is difficult

Worst-case approach



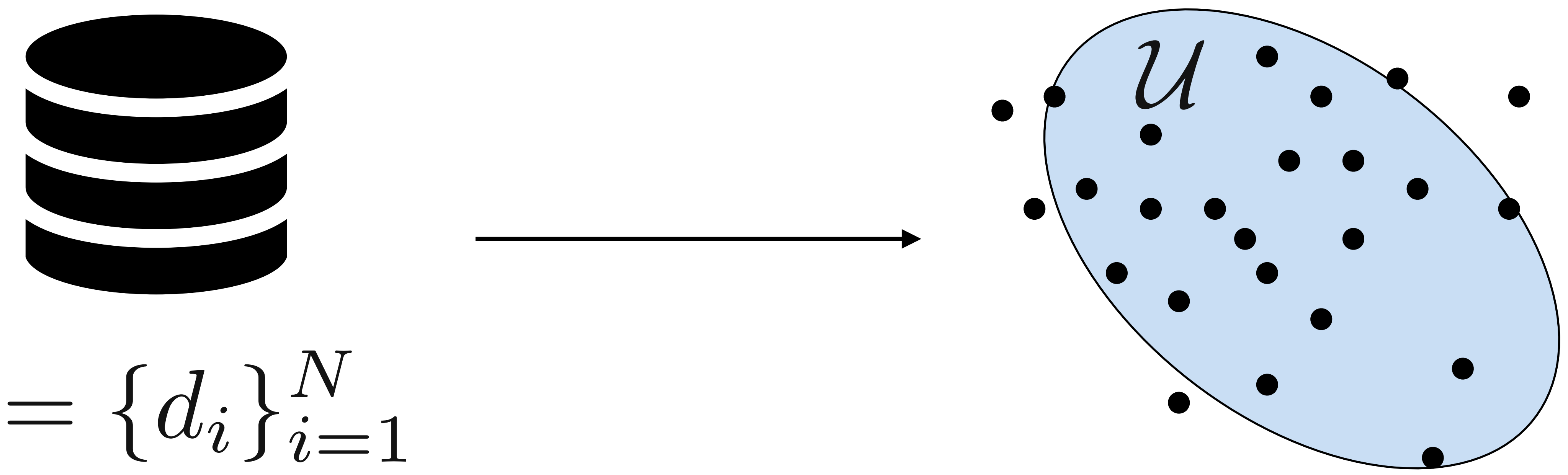
❌ Very conservative

Probabilistic approach



❌ nobody knows P

Can we use data?



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

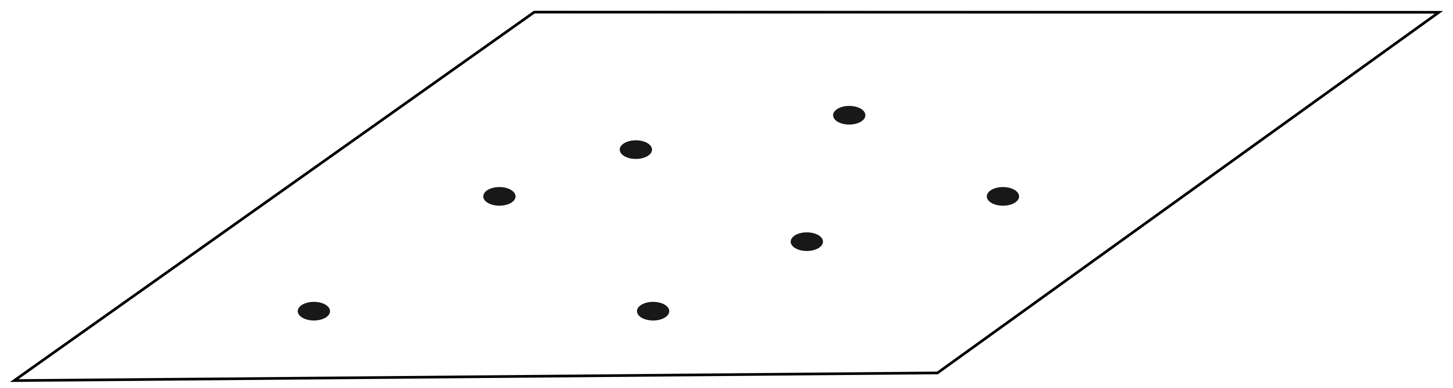
"Probabilistic Guarantees in Robust Optimization", D. Bertsimas, D. den Hertog, and J. Pauphilet (2019)
"Data-driven robust optimization", D. Bertsimas, V. Gupta, and N. Kallus (2014)
"Learning-Based Robust Optimization: Procedures and Statistical Guarantees", L. Jeff Hong, Z. Huang, and H. Lam (2021)

Estimating true distribution from data

Data-Driven Distributionally Robust Optimization

Data

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



Estimating true distribution from data

Data-Driven Distributionally Robust Optimization

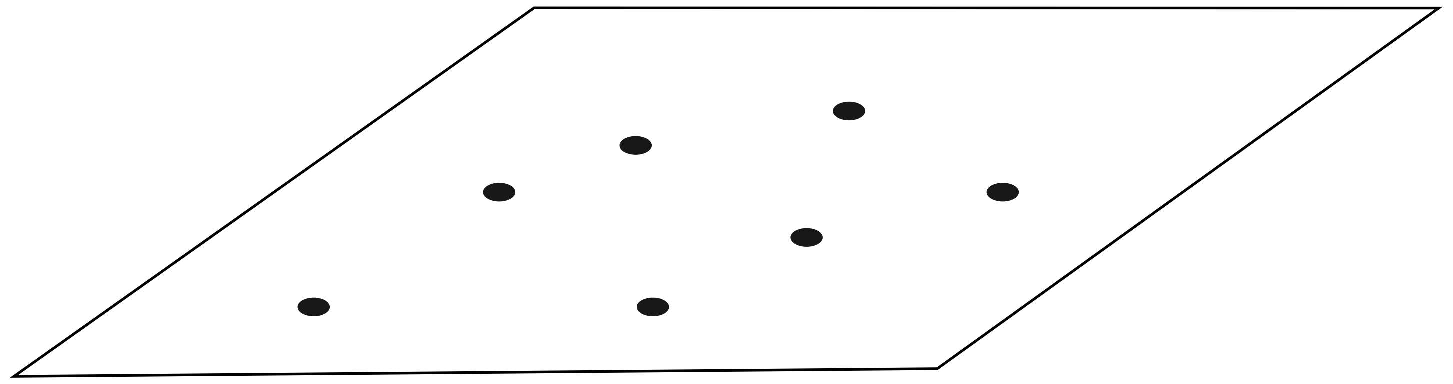
Data

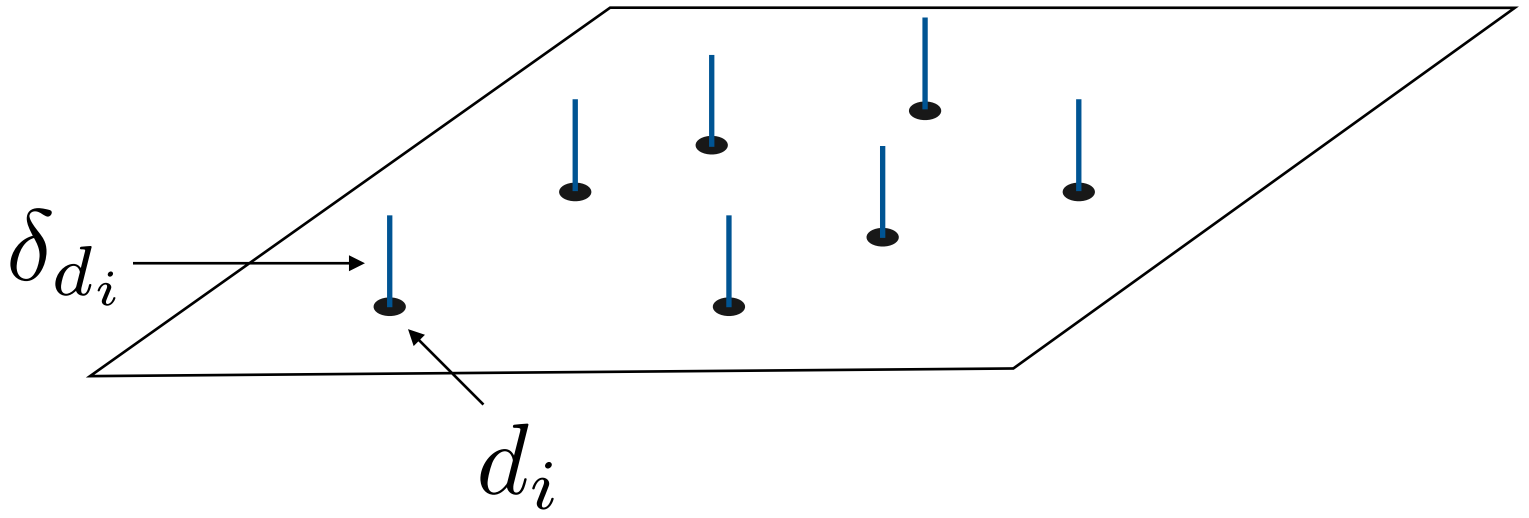
$\mathcal{D} = \{d_i\}_{i=1}^{N}$

⟶

Empirical Distribution

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





Estimating true distribution from data

Data-Driven Distributionally Robust Optimization

Data

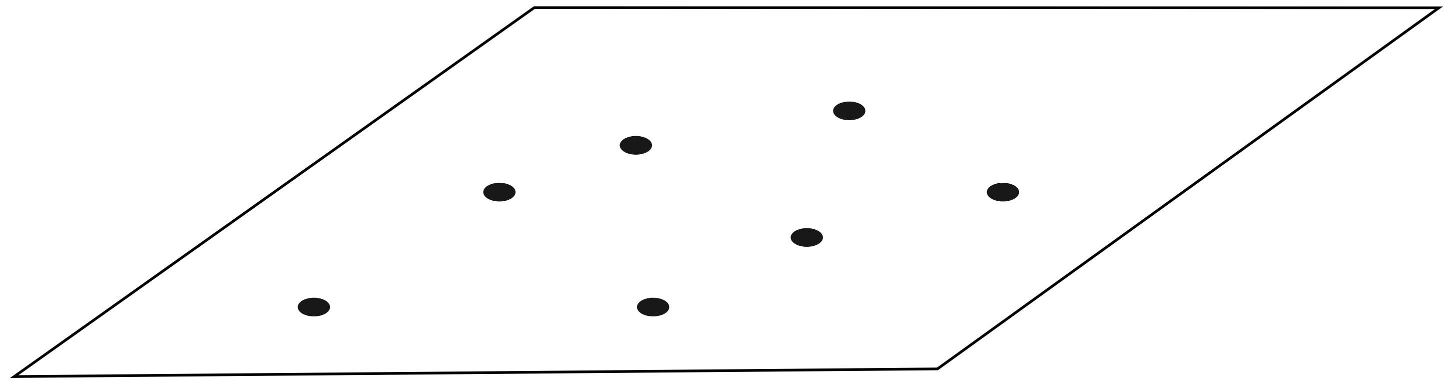
$$\mathcal{D} = \{d_i\}_{i=1}^{N}$$

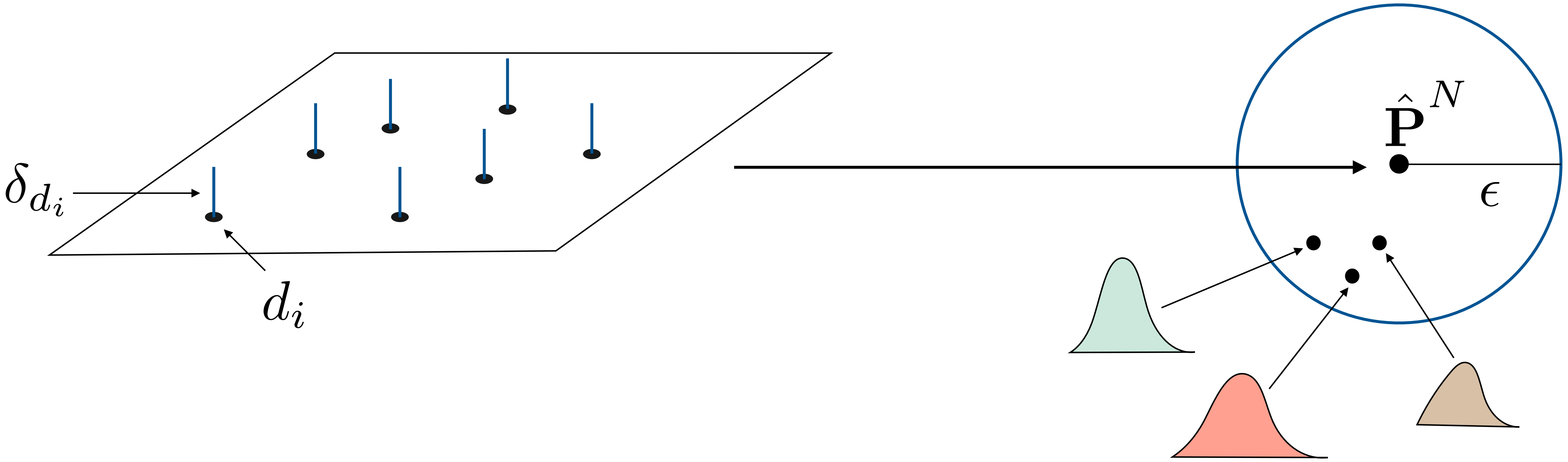
Empirical Distribution

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

Ambiguity Set

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





Estimating true distribution from data

Data-Driven Distributionally Robust Optimization

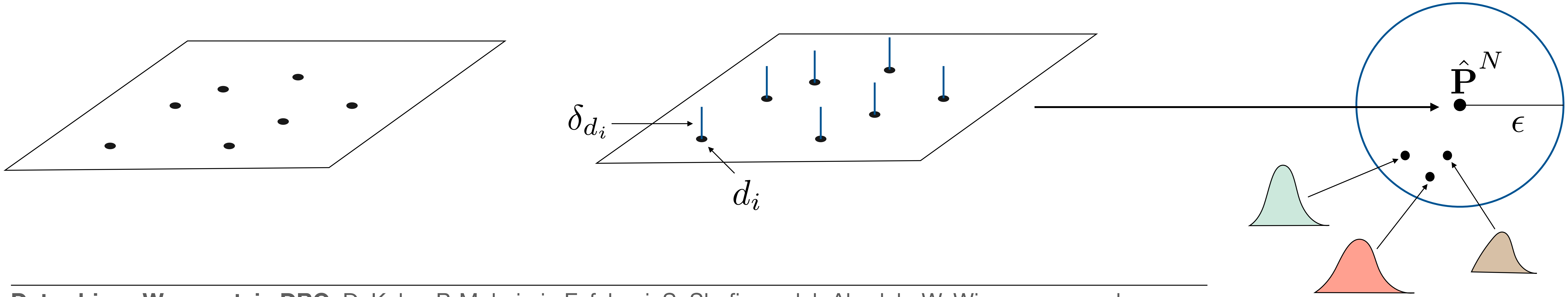
Empirical Distribution

Ambiguity Set

Data

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

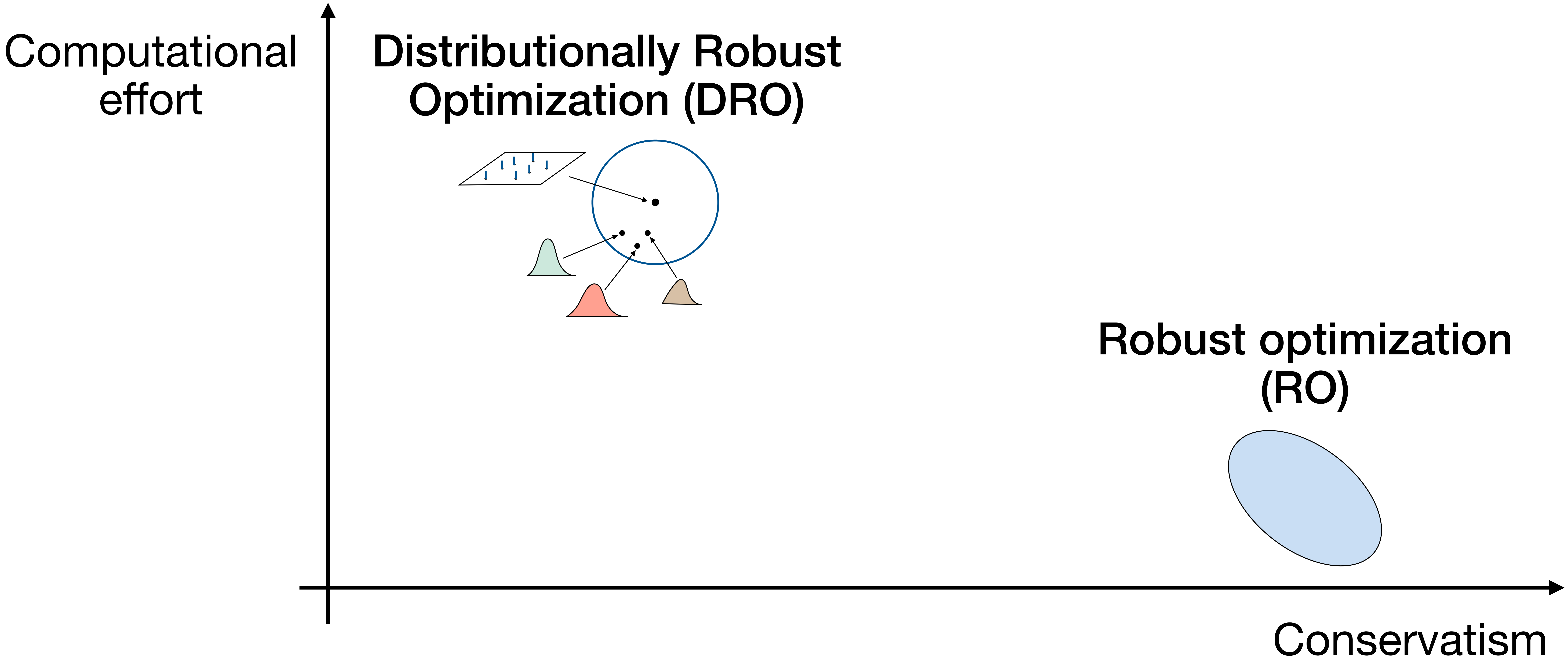
Wasserstein distance



Data-driven Wasserstein DRO. D. Kuhn, P. Mohajerin Esfahani, S. Shafieezadeh Abadeh, W. Wiesemann, and many more

Optimal Transport. G. Monge, L. Kantorovic, L. Wasserstein, C. Villani, …and many more

Robust vs Distributionally Robust Optimization



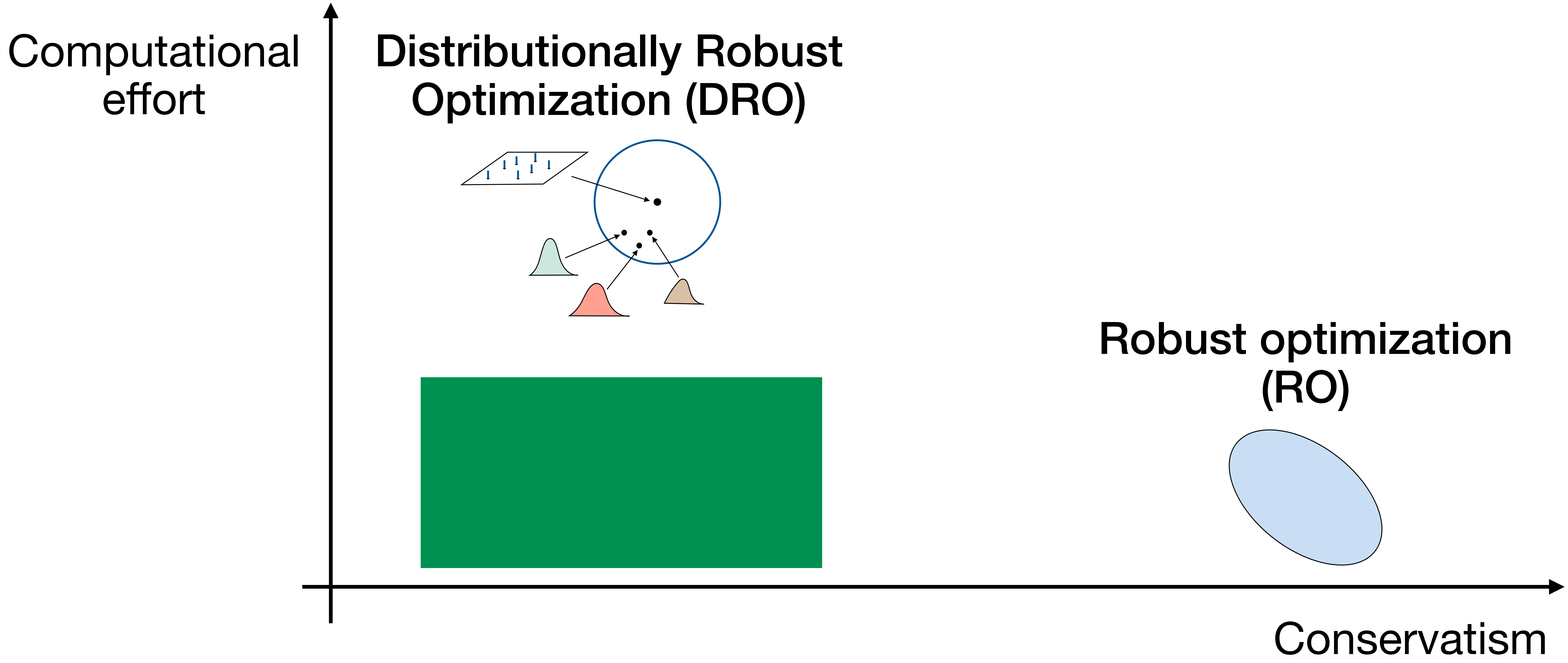
Computational effort

Distributionally Robust Optimization (DRO)

Robust optimization (RO)

Conservatism

Robust vs Distributionally Robust Optimization



Computational effort

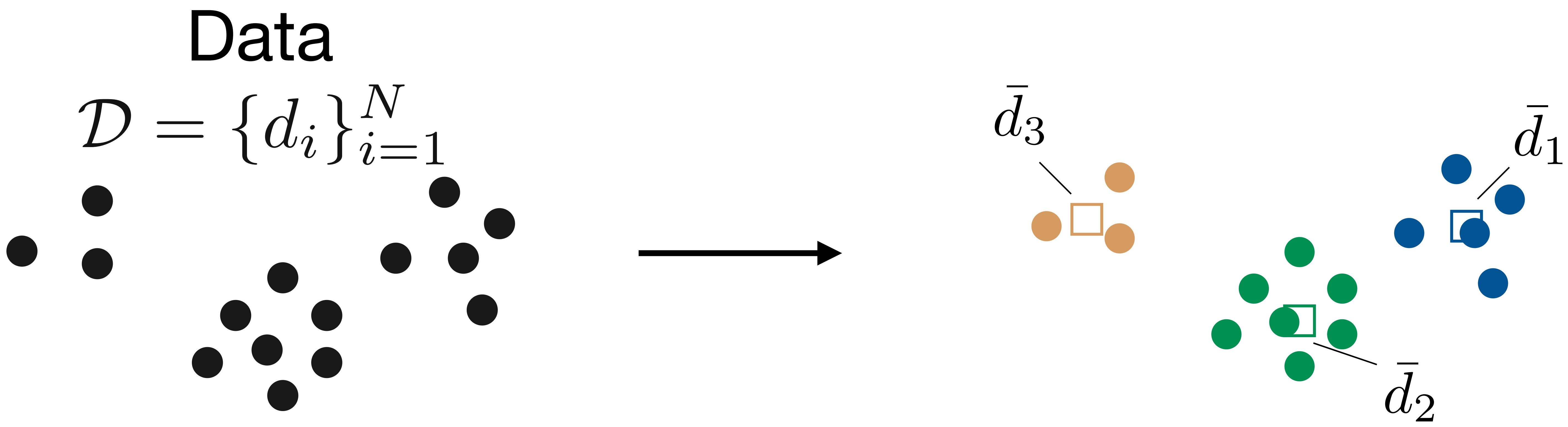
Distributionally Robust Optimization (DRO)

Robust optimization (RO)

Conservatism

Can we get the best of both worlds?

Clustering reduces dimensionality



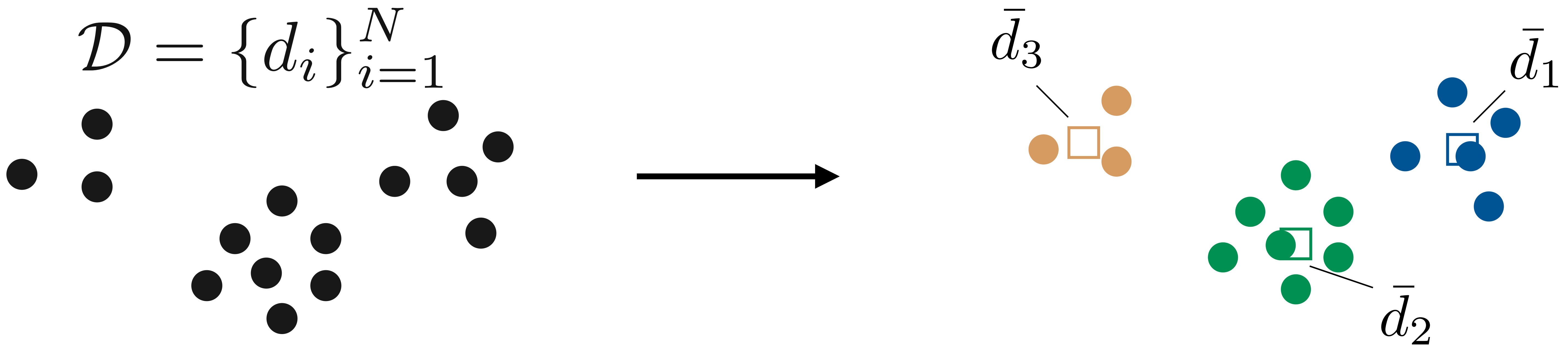
Data

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

Clustering reduces dimensionality

Data

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

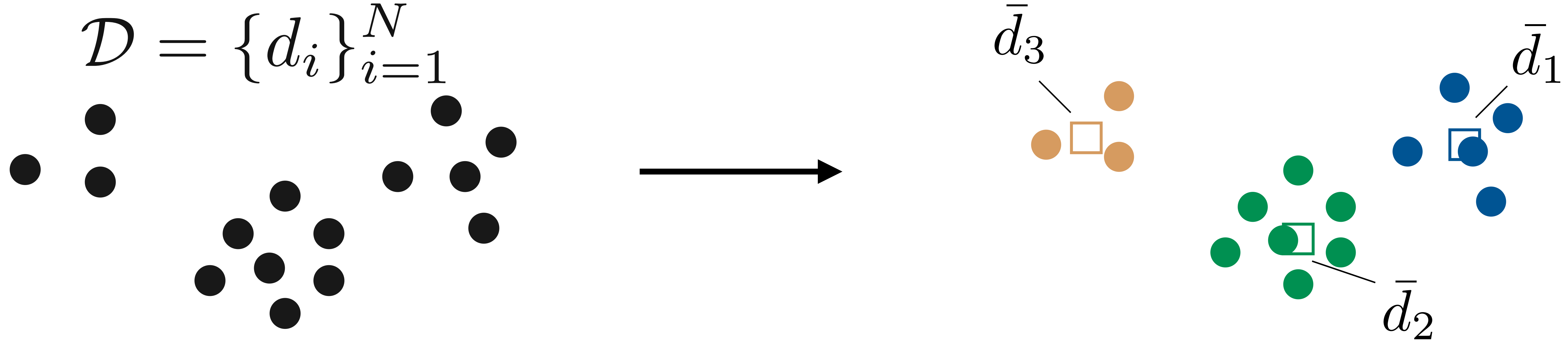


minimize $$\sum_{k=1}^{K} \sum_{i \in C_k} \|d_i - \bar{d}_k\|^2$$

Clustering reduces dimensionality

Data

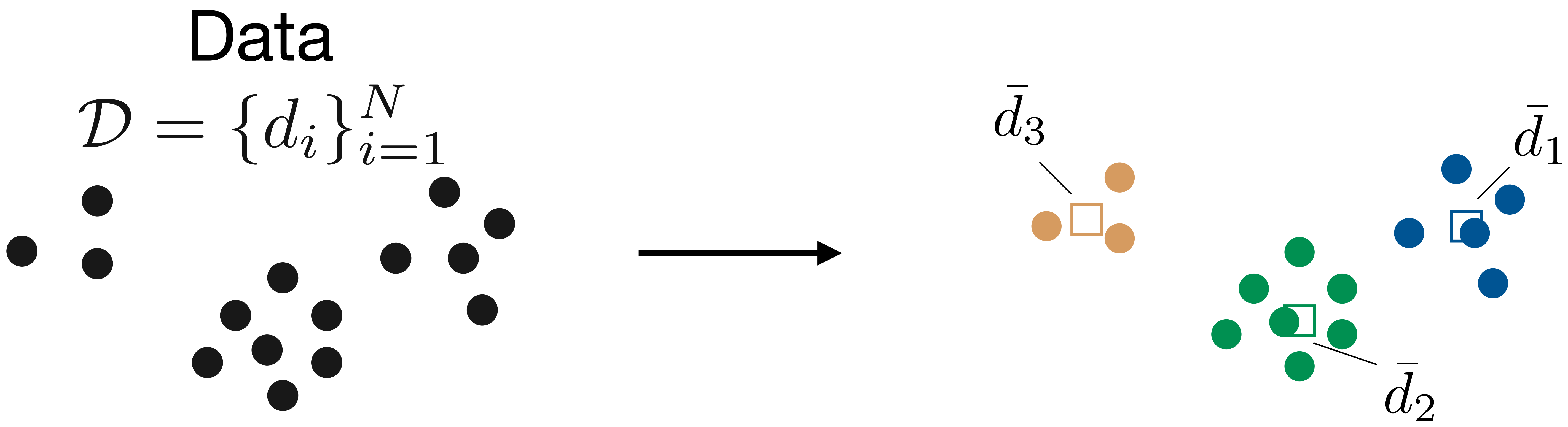
$\mathcal{D} = \{d_i\}_{i=1}^{N}$



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

cluster centers

Clustering reduces dimensionality



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

cluster centers

Main idea

Use cluster centers
instead of original data

Probabilistic guarantees

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

concave
function

Probabilistic guarantees

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

concave
function

$$u \sim P$$

(but we never know P!)

Probabilistic guarantees

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

concave function

$$u \sim P$$

(but we never know P!)

Data

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

Probabilistic guarantees

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

concave function

$$u \sim P$$

(but we never know P!)

Data

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

Data-driven probabilistic guarantees

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

Probabilistic guarantees

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

concave function

$u \sim P$

(but we never know P!)

Data

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

Data-driven probabilistic guarantees

Product Distribution

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

Probabilistic guarantees

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

concave function

$u \sim P$

(but we never know P!)

Data

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

Data-driven probabilistic guarantees

Product Distribution

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

data-driven solution

Probabilistic guarantees

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

concave function

$u \sim P$

(but we never know P!)

Data

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

Data-driven probabilistic guarantees

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

Product Distribution

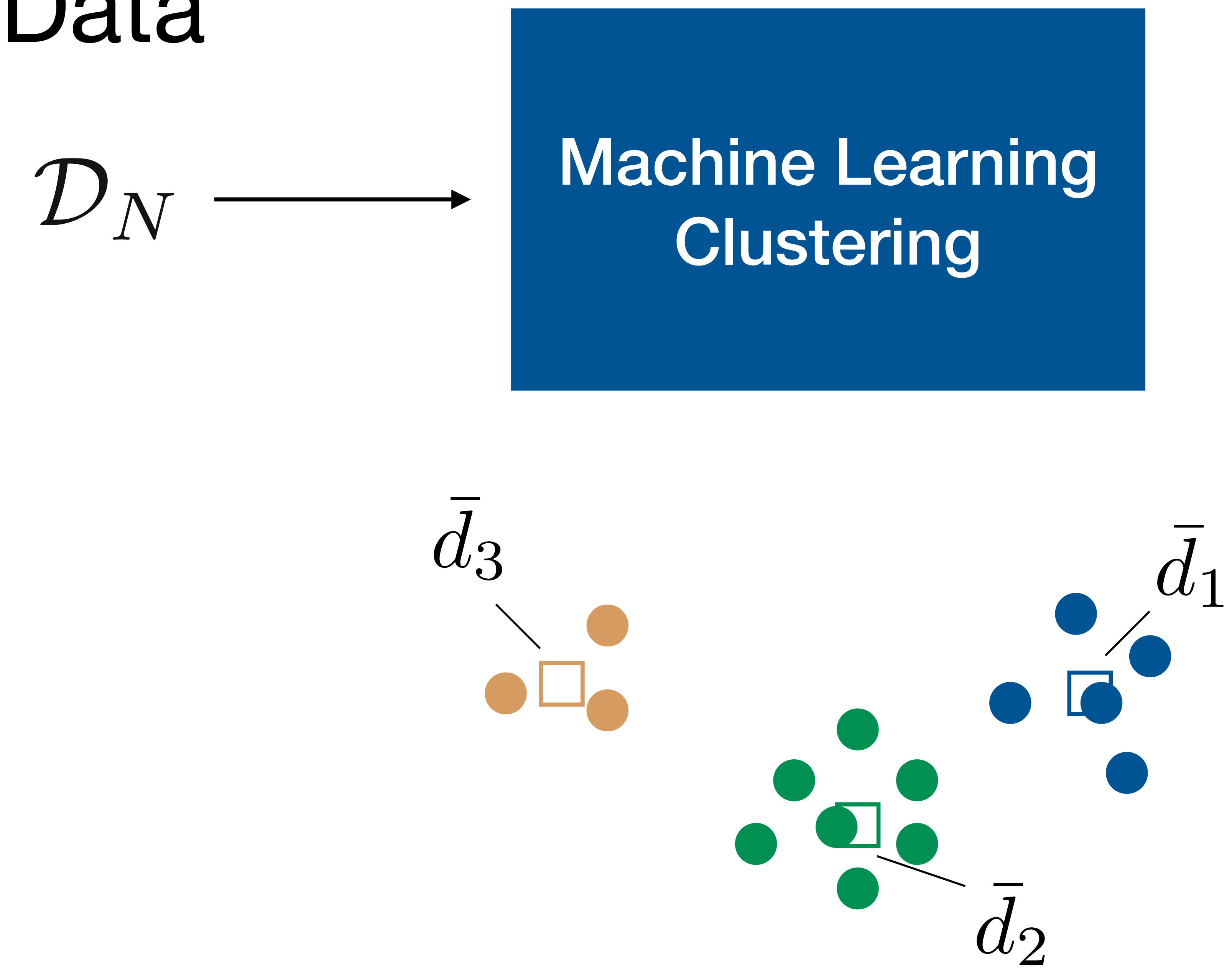
data-driven solution

probability of constraint satisfaction

Mean Robust Optimization

Mean Robust Optimization (MRO)

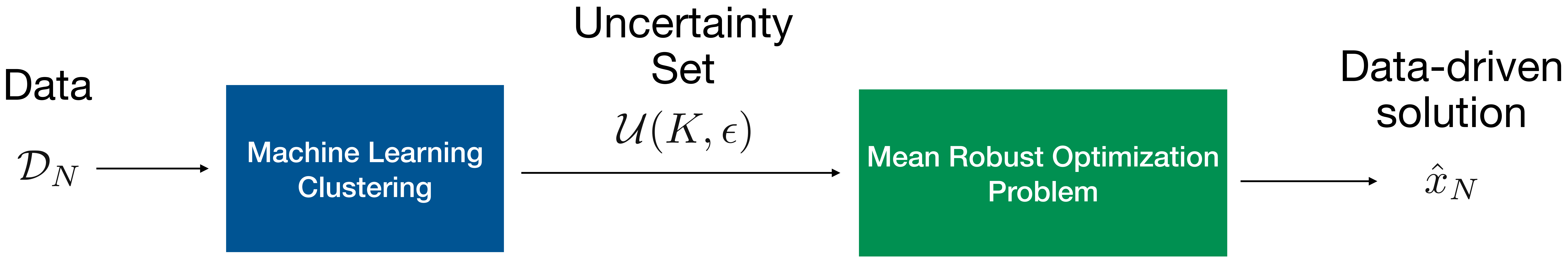
Data



$\mathcal{D}_N$

Machine Learning
Clustering

Mean Robust Optimization (MRO)



Data

$\mathcal{D}_N$

Machine Learning Clustering

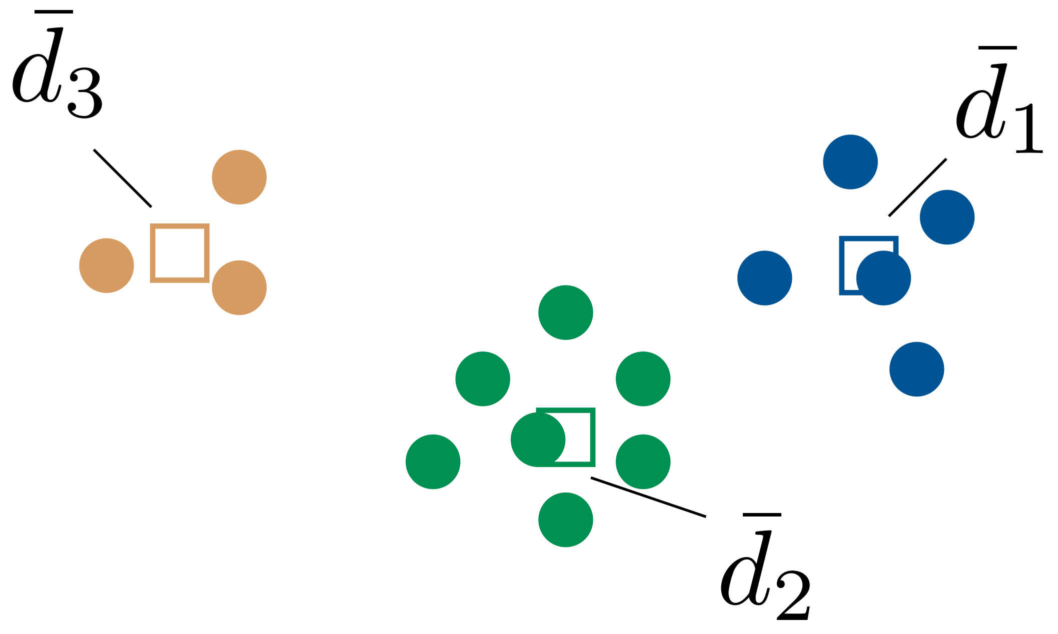
Uncertainty Set

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

Mean Robust Optimization Problem

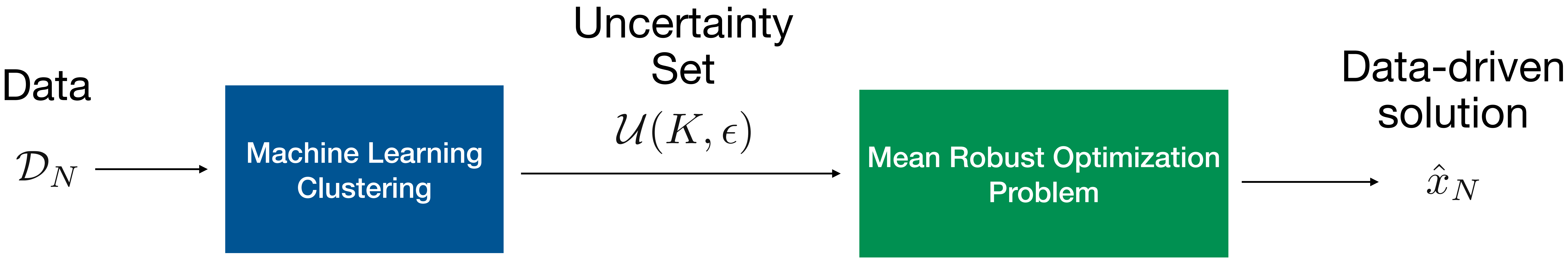
Data-driven solution

$\hat{x}_N$



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

Mean Robust Optimization (MRO)



Data

$\mathcal{D}_N$

Machine Learning Clustering

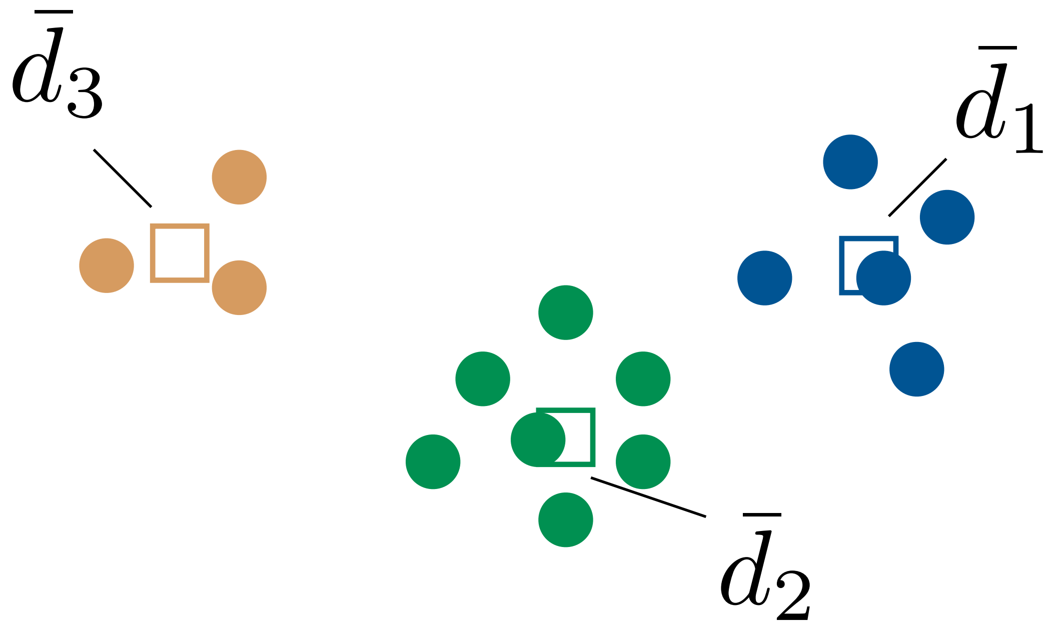
Uncertainty Set

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

Mean Robust Optimization Problem

Data-driven solution

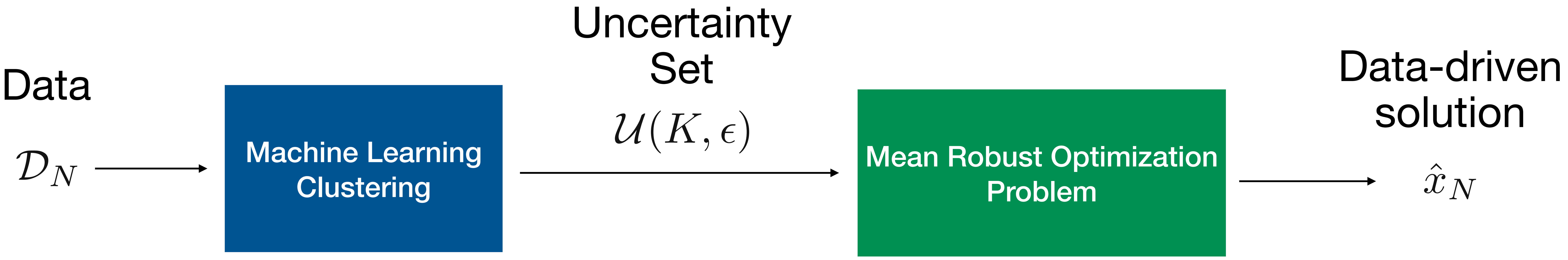
$\hat{x}_N$



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

constraint function

Mean Robust Optimization (MRO)



Data

$\mathcal{D}_N$

Machine Learning Clustering

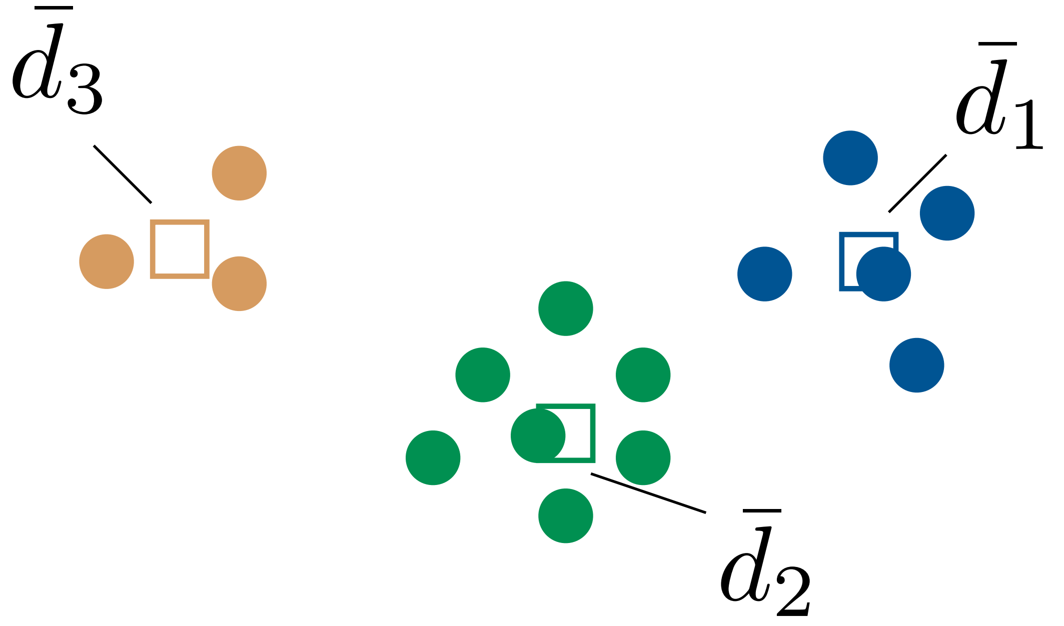
Uncertainty Set

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

Mean Robust Optimization Problem

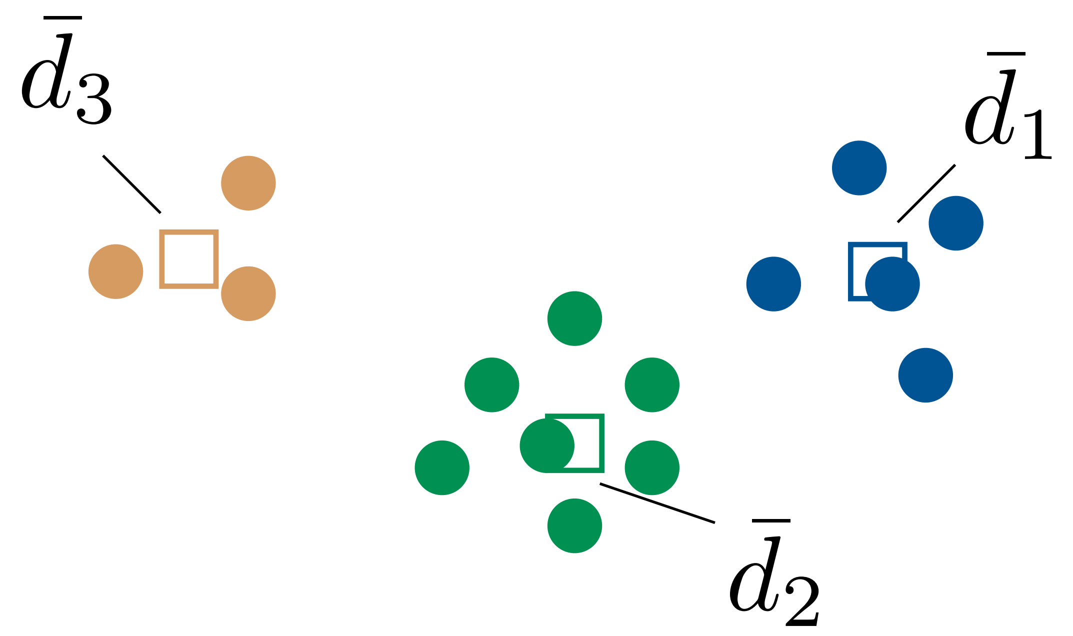
Data-driven solution

$\hat{x}_N$

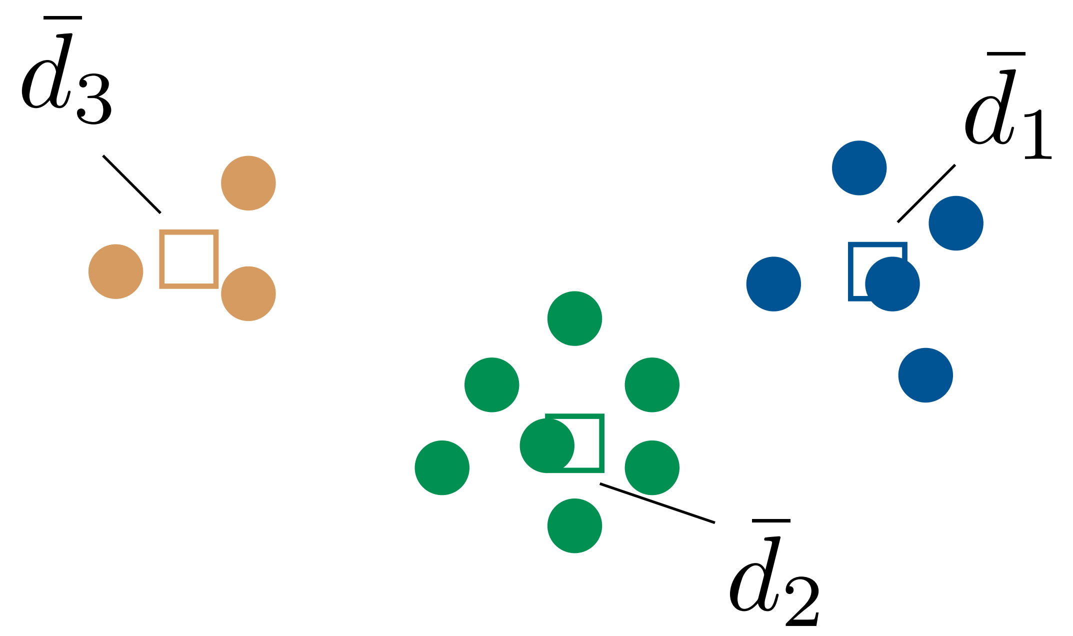


$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \underbrace{\bar{g}(u,x)}_{\text{constraint function}} \leq 0 \quad \forall u \in \underbrace{\mathcal{U}(K,\epsilon)}_{\text{uncertainty set}} \end{array}$$

Uncertainty set



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

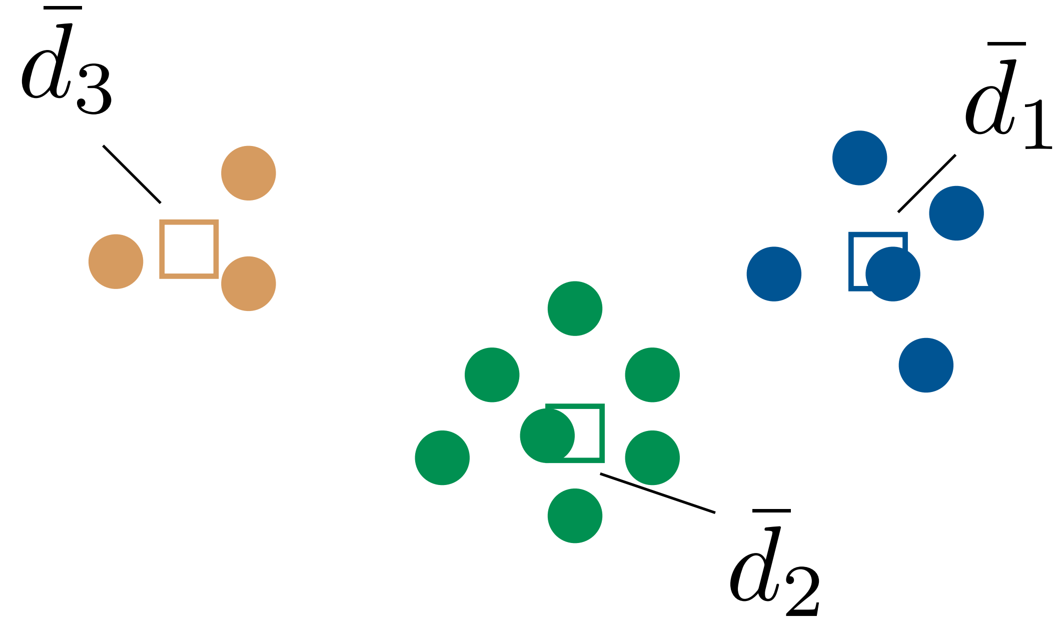


Uncertainty set

cluster weights

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

Uncertainty set

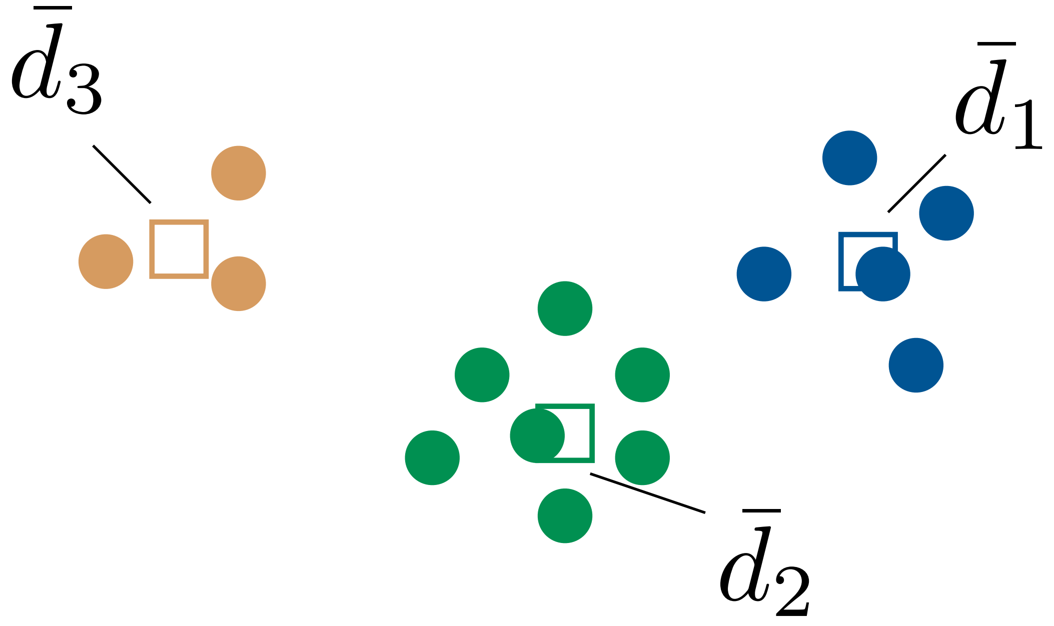


cluster weights

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

cluster centers

Uncertainty set



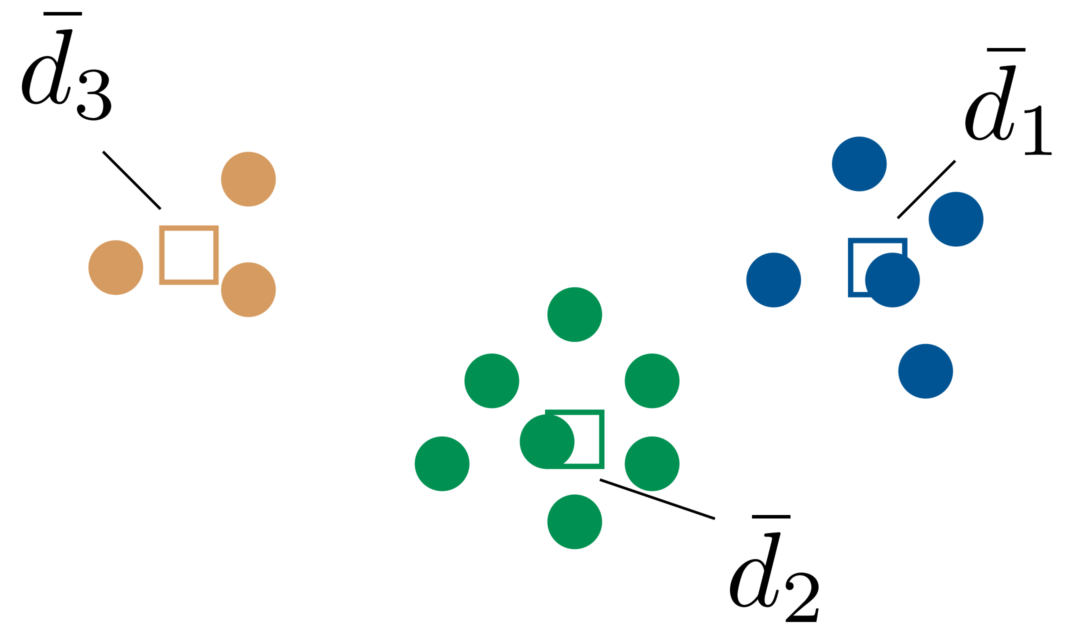
cluster weights

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

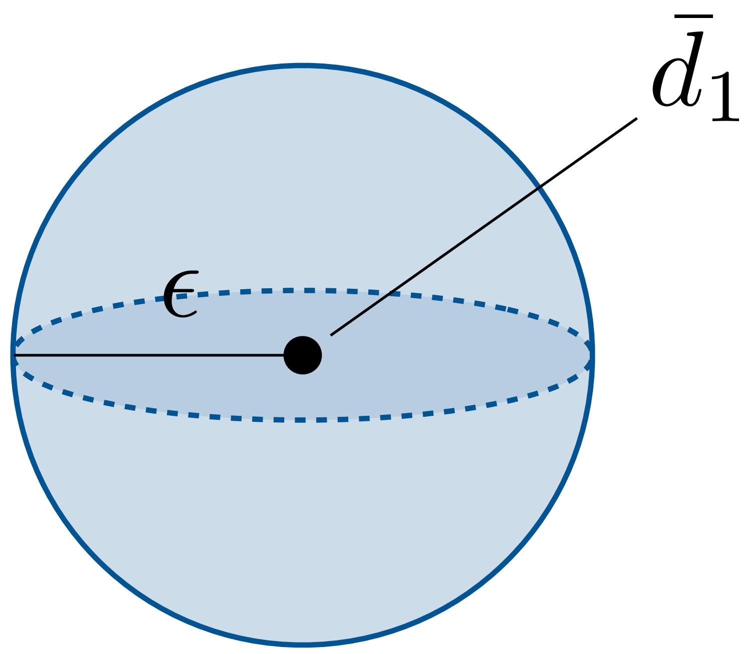
$$\mathcal{U}(K, \epsilon) = \left\{ u = (v_1, \dots, v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \le \epsilon^p \right\}$$

cluster centers

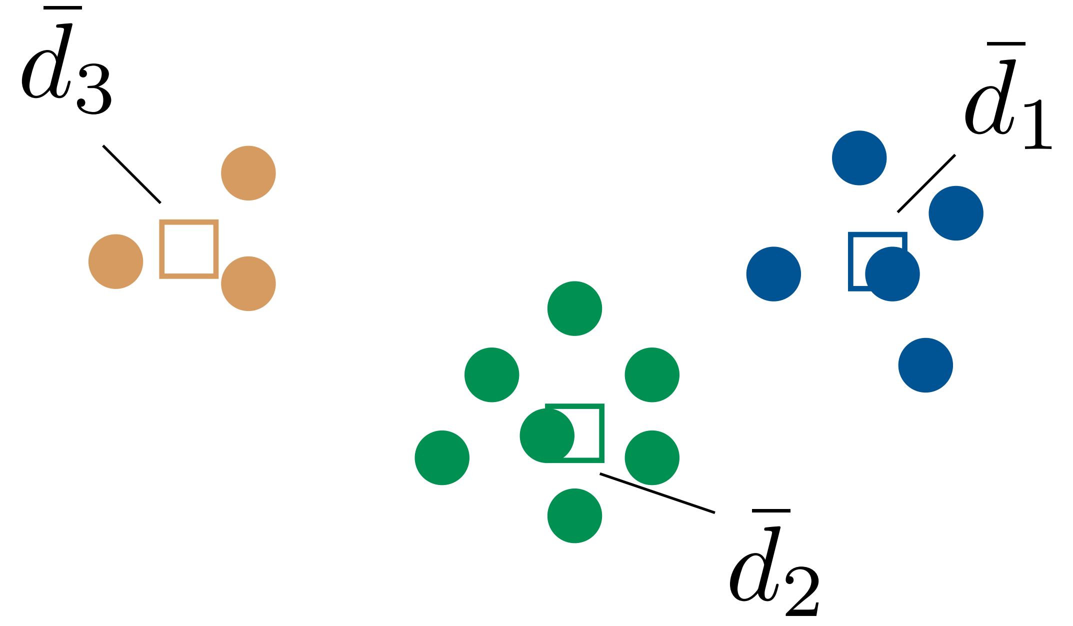
Examples

K

$K = 1$



Uncertainty set



cluster weights

order

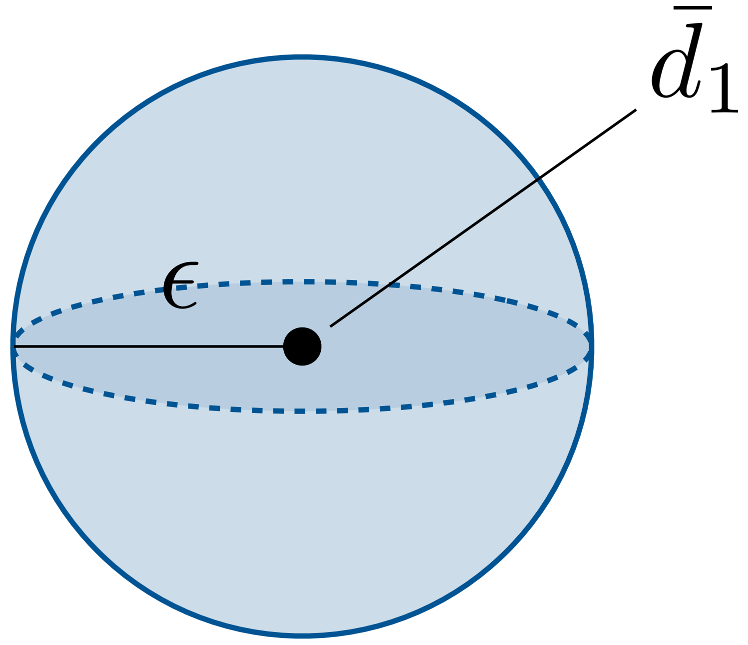
$$\mathcal{U}(K,\epsilon) = \left\{ u = (v_1,\ldots,v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \leq \epsilon^p \right\}$$

cluster centers

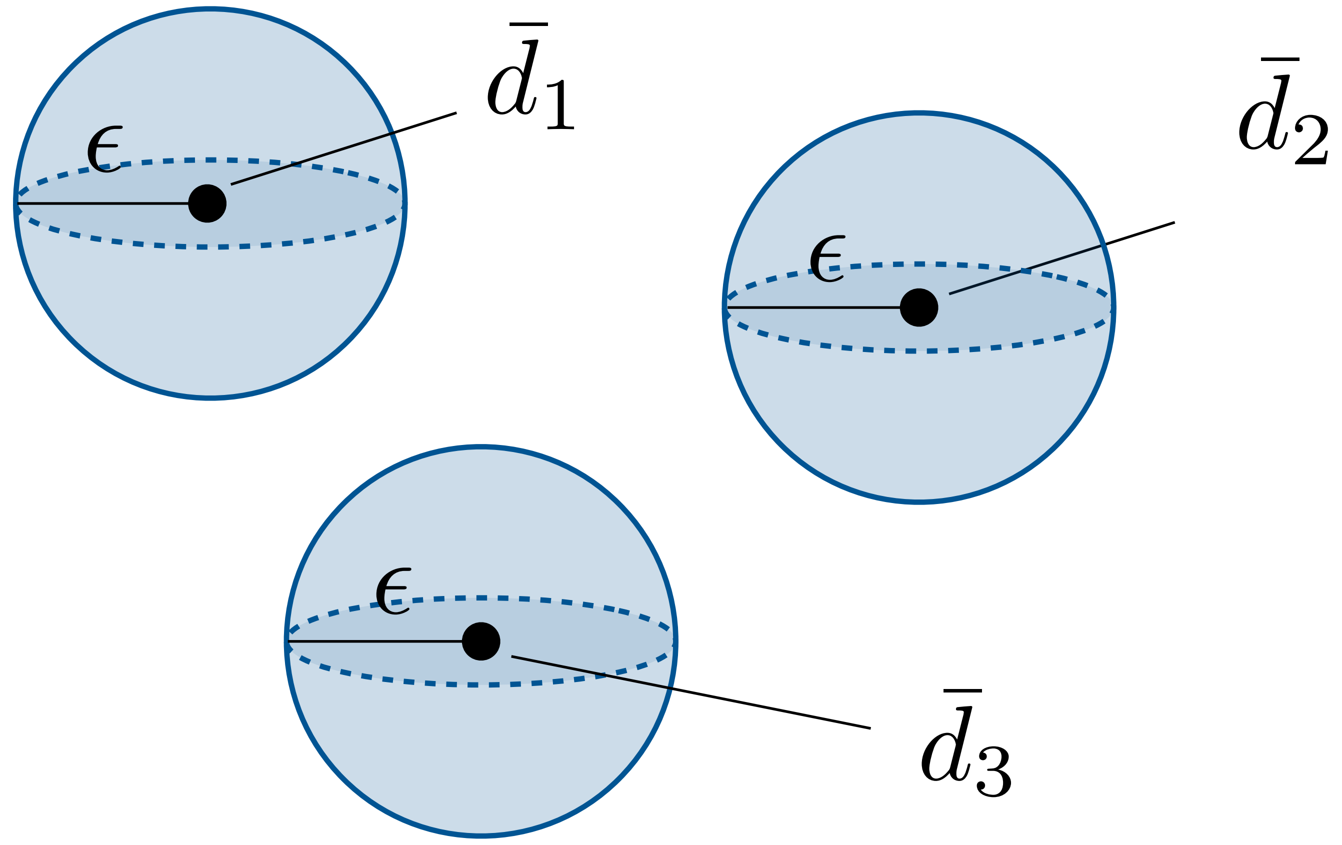
Examples

K

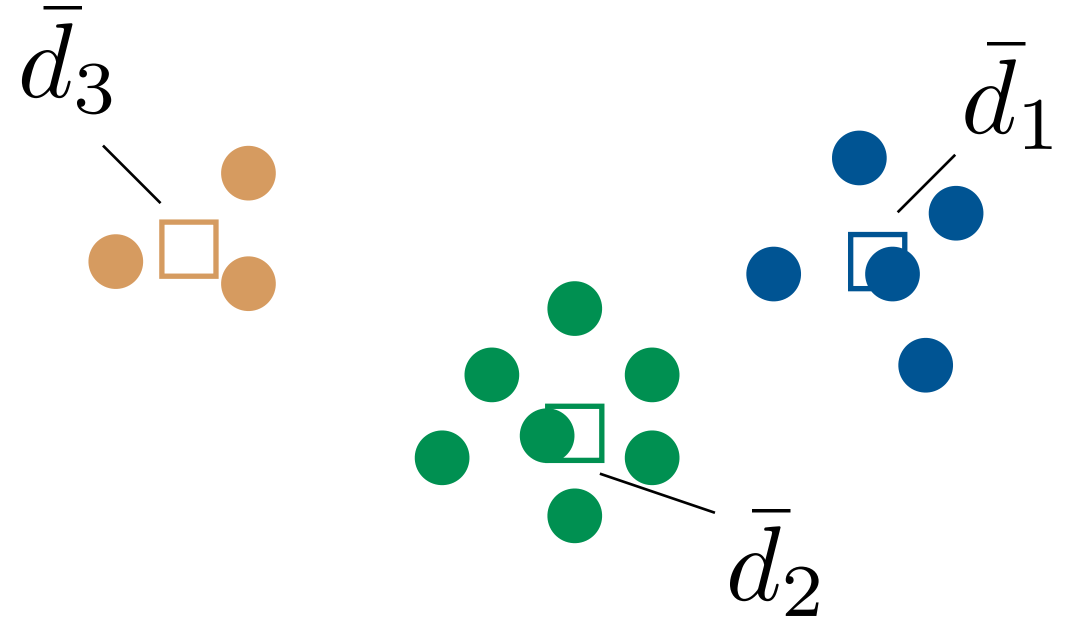
$K = 1$



$K = 3, p = \infty$



Uncertainty set



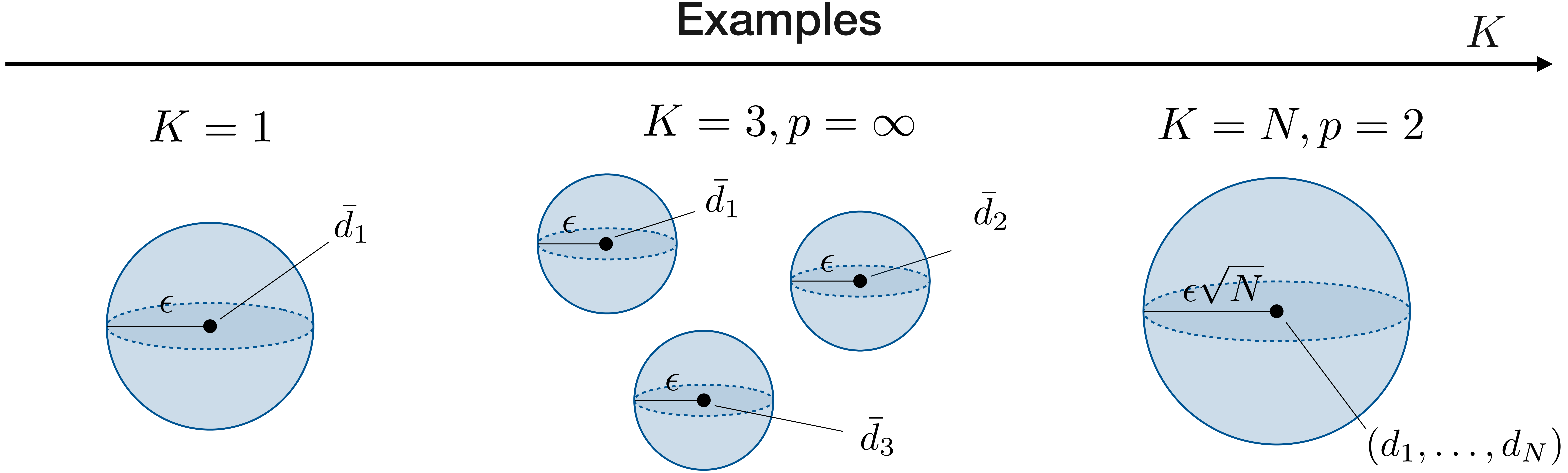
$$\mathcal{U}(K, \epsilon) = \left\{ u = (v_1, \ldots, v_K) \;\middle|\; \sum_{k=1}^{K} w_k \|v_k - \bar{d}_k\|^p \leq \epsilon^p \right\}$$

cluster weights

order

cluster centers

Examples



$K = 1$

$K = 3, p = \infty$

$K = N, p = 2$

Mean Robust Optimization Problem

Uncertain variable lifting

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

$\longrightarrow$

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

Mean Robust Optimization Problem

Uncertain variable lifting

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

$$\longrightarrow$$

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

uncertainty set

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

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

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

→

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

constraint function

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

Solving the MRO problem

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

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

Solving the MRO problem

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

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

conjugate function

Solving the MRO problem

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

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

conjugate function

cluster centers

Solving the MRO problem

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

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

conjugate function

cluster centers

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

Solving the MRO problem

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

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

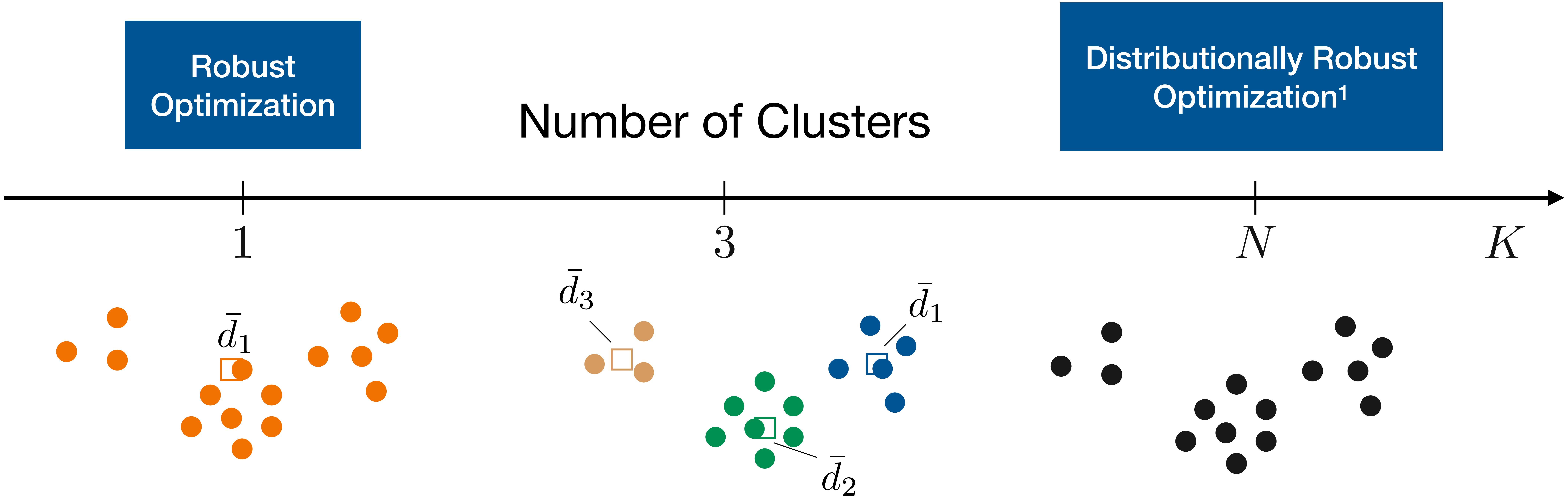
conjugate function

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 RO and DRO



Robust Optimization

Number of Clusters

Distributionally Robust Optimization[1]

---

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

Guarantees

Satisfying the probabilistic guarantees

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

Satisfying the probabilistic guarantees

probability of constraint satisfaction

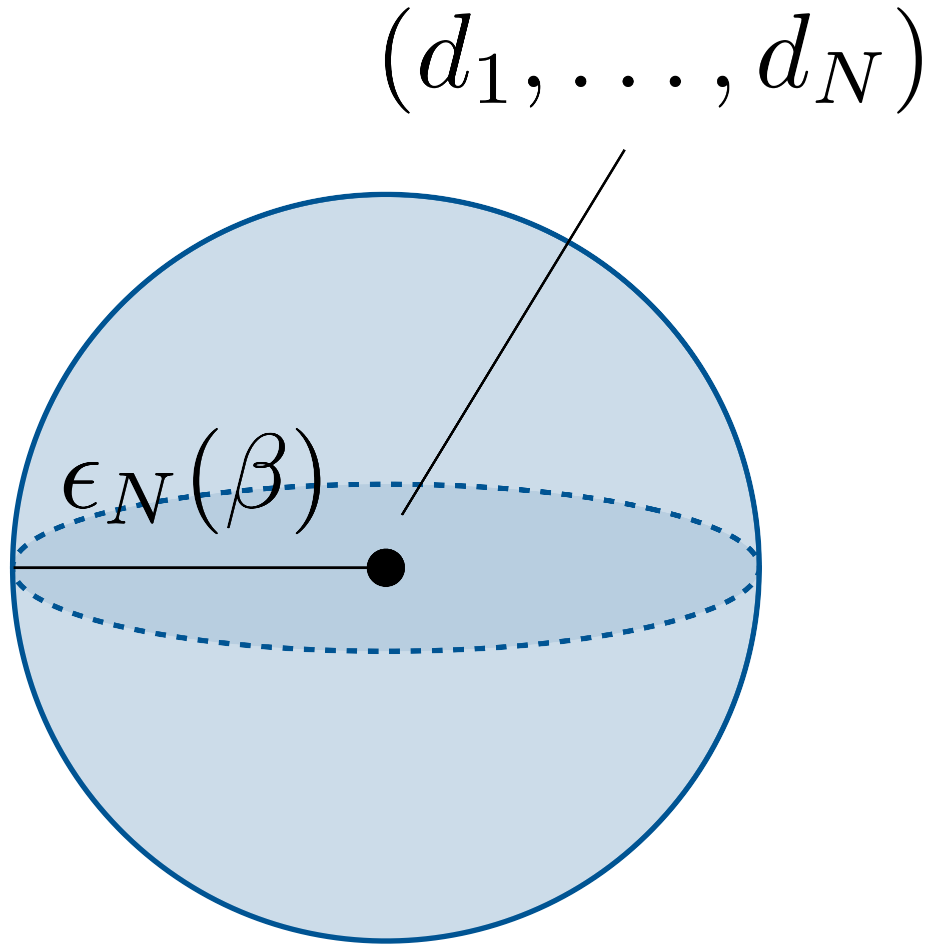
$$\mathbf{P}^N\left(\mathbf{E}(g(u,\hat{x}_N)) \leq 0\right) \geq \underbrace{1-\beta}$$

Satisfying the probabilistic guarantees

probability of constraint satisfaction

uncertainty set radius

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

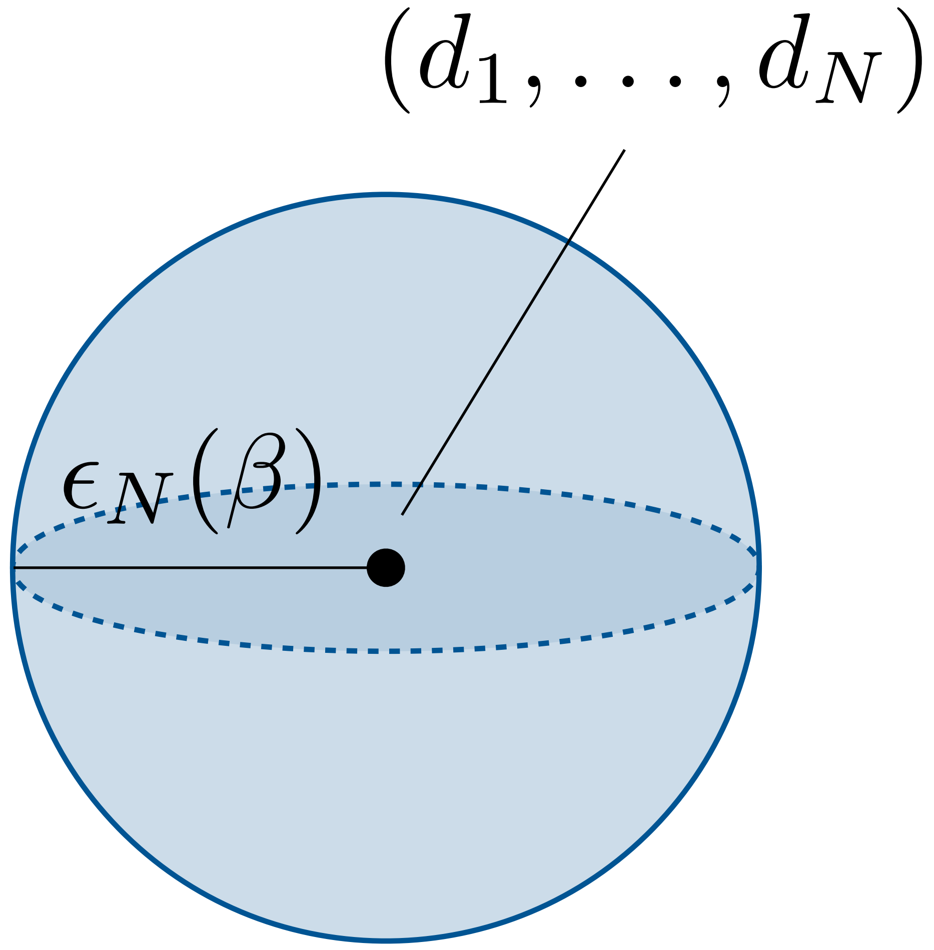


Satisfying the probabilistic guarantees

probability of constraint satisfaction

uncertainty set radius

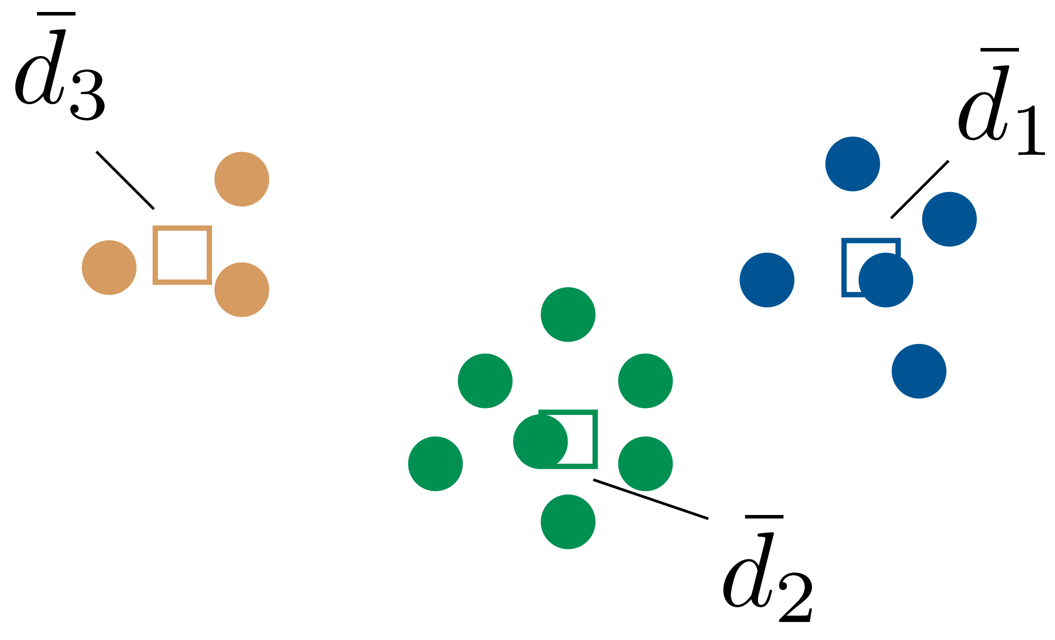
$$\mathbf{P}^N\left(\mathbf{E}(g(u,\hat{x}_N)) \le 0\right) \ge 1-\beta \xrightarrow{\text{light-tailed}} \mathcal{U}(N,\epsilon_N(\beta))$$



MRO clustering

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

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

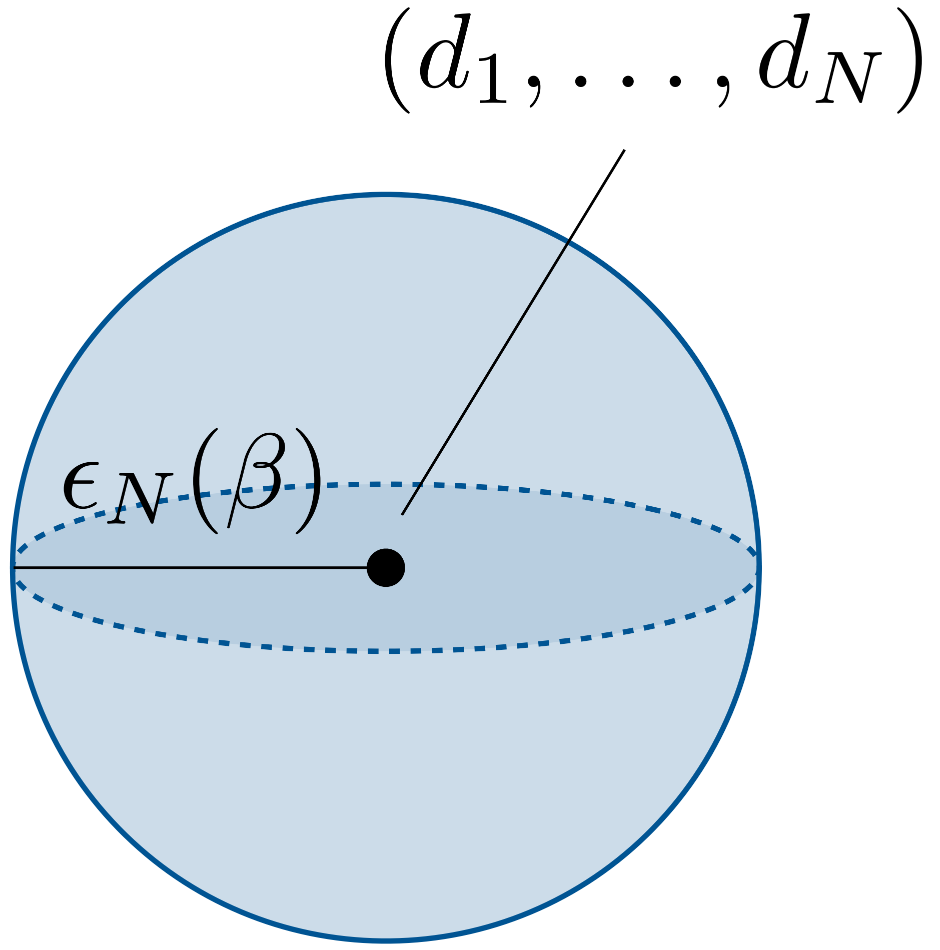


Satisfying the probabilistic guarantees

probability of constraint satisfaction

uncertainty set radius

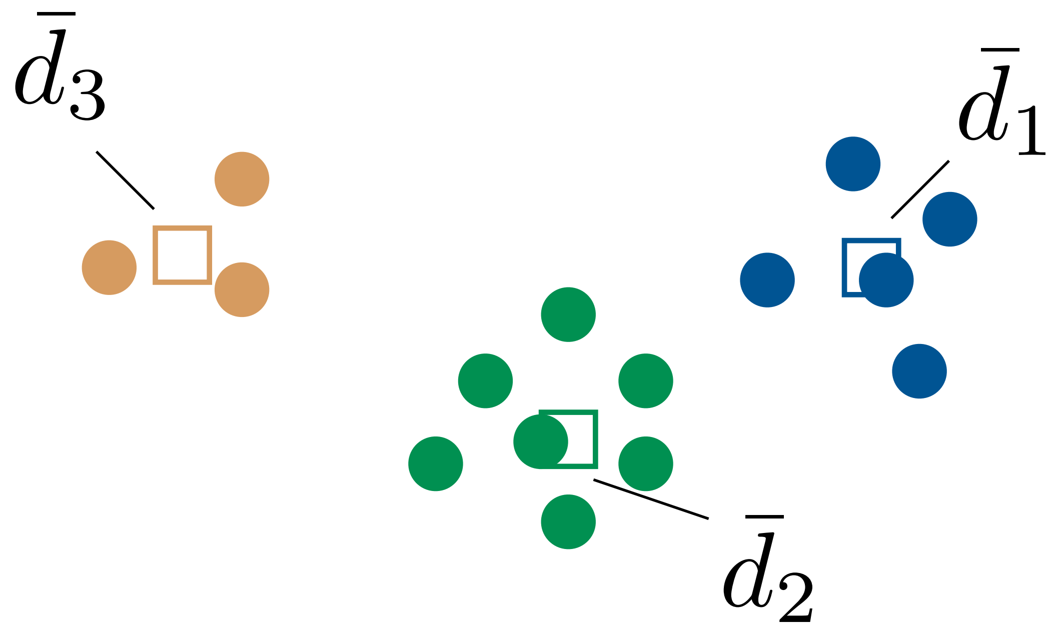
$$\mathbf{P}^N\left(\mathbf{E}(g(u, \hat{x}_N)) \le 0\right) \ge \boxed{1 - \beta} \xrightarrow{\text{light-tailed}} \mathcal{U}(N, \boxed{\epsilon_N(\beta)})$$



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

MRO constraint

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

Worst-case values

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

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

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

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

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

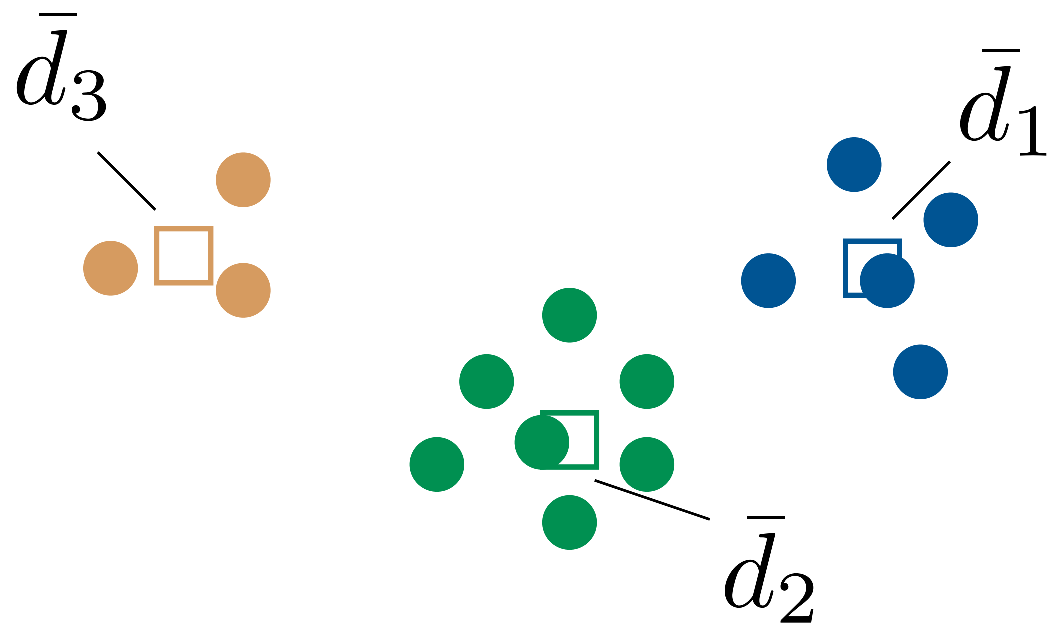
$$\bar{g}(u, x) \leq 0 \quad \forall u \in \mathcal{U}(K, \epsilon)$$

Theorem

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

clustering objective

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



Bounding the conservatism

MRO constraint

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

Worst-case values

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

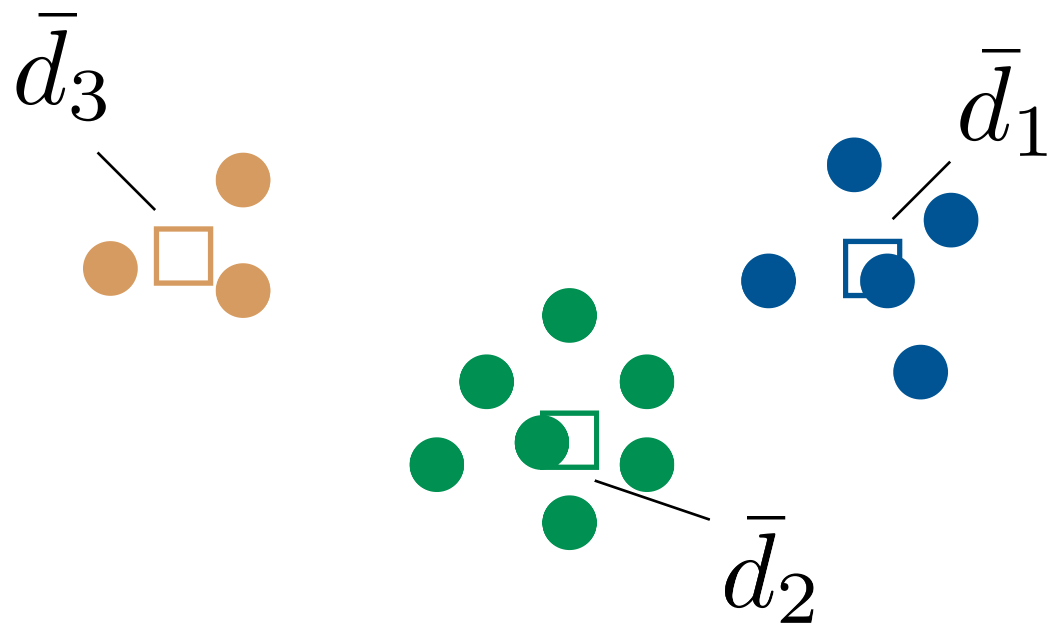
$$\bar{g}^{K}(x) = \underset{u \in \mathcal{U}(K, \epsilon)}{\text{maximize}} \quad \bar{g}(u, x)$$

Theorem

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

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

clustering objective



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

Example MRO with linear constraints

$p = \infty$ $\qquad (a + Pu)^T x \leq b \xrightarrow{\qquad} g(u, x) = (a + Pu)^T x - b$

Example MRO with linear constraints

$p = \infty$ $\quad (a + Pu)^T x \leq b \xrightarrow{\quad\quad} g(u, x) = (a + Pu)^T x - b$

$$[-g]^*(z_k, x) = \sup_u z_k^T u - (a + Pu)^T x + b = \begin{cases} a^T x - b & \text{if } z_k + P^T x = 0 \\ \infty & \text{otherwise} \end{cases}$$

Example MRO with linear constraints

$p = \infty$

$$(a + Pu)^T x \leq b \longrightarrow g(u, x) = (a + Pu)^T x - b$$

$$[-g]^*(z_k, x) = \sup_u z_k^T u - (a + Pu)^T x + b = \begin{cases} a^T x - b & \text{if } z_k + P^T x = 0 \\ \infty & \text{otherwise} \end{cases}$$

$$z_1 = z_2 = \cdots = z_k = -P^T x$$

Example MRO with linear constraints

$p = \infty$

$$(a + Pu)^T x \leq b \longrightarrow g(u, x) = (a + Pu)^T x - b$$

$$[-g]^*(z_k, x) = \sup_u z_k^T u - (a + Pu)^T x + b = \begin{cases} a^T x - b & \text{if } z_k + P^T x = 0 \\ \infty & \text{otherwise} \end{cases}$$

$$z_1 = z_2 = \cdots = z_k = -P^T x$$

Convex reformulation

minimize $\quad f(x)$

subject to $\quad a^T x - b + (P^T x)^T \sum_{k=1}^K w_k \bar{d}_k + \epsilon \|P^T x\|_* \leq 0$

Example MRO with linear constraints

$p = \infty$

$$(a+Pu)^T x \leq b \xrightarrow{\quad\quad} g(u,x) = (a+Pu)^T x - b$$

$$[-g]^*(z_k, x) = \sup_u z_k^T u - (a+Pu)^T x + b = \begin{cases} a^T x - b & \text{if } \boxed{z_k + P^T x = 0} \\ \infty & \text{otherwise} \end{cases}$$

$$z_1 = z_2 = \cdots = z_k = -P^T x$$

Convex reformulation

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & a^T x - b + (P^T x)^T \underbrace{\textstyle\sum_{k=1}^K w_k \bar{d}_k}_{\text{average } \bar{d}} + \epsilon \|P^T x\|_* \leq 0 \end{array}$$

(clustering makes no difference)

average $\bar{d}$

Example MRO with linear constraints

$p = \infty$

$$(a + Pu)^T x \leq b \longrightarrow g(u, x) = (a + Pu)^T x - b$$

$$[-g]^*(z_k, x) = \sup_u z_k^T u - (a + Pu)^T x + b = \begin{cases} a^T x - b & \text{if } \boxed{z_k + P^T x = 0} \\ \infty & \text{otherwise} \end{cases}$$

$$z_1 = z_2 = \cdots = z_k = -P^T x$$

Convex reformulation

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & a^T x - b + (P^T x)^T \boxed{\sum_{k=1}^K w_k \bar{d}_k} + \epsilon \|P^T x\|_* \leq 0 \end{array}$$

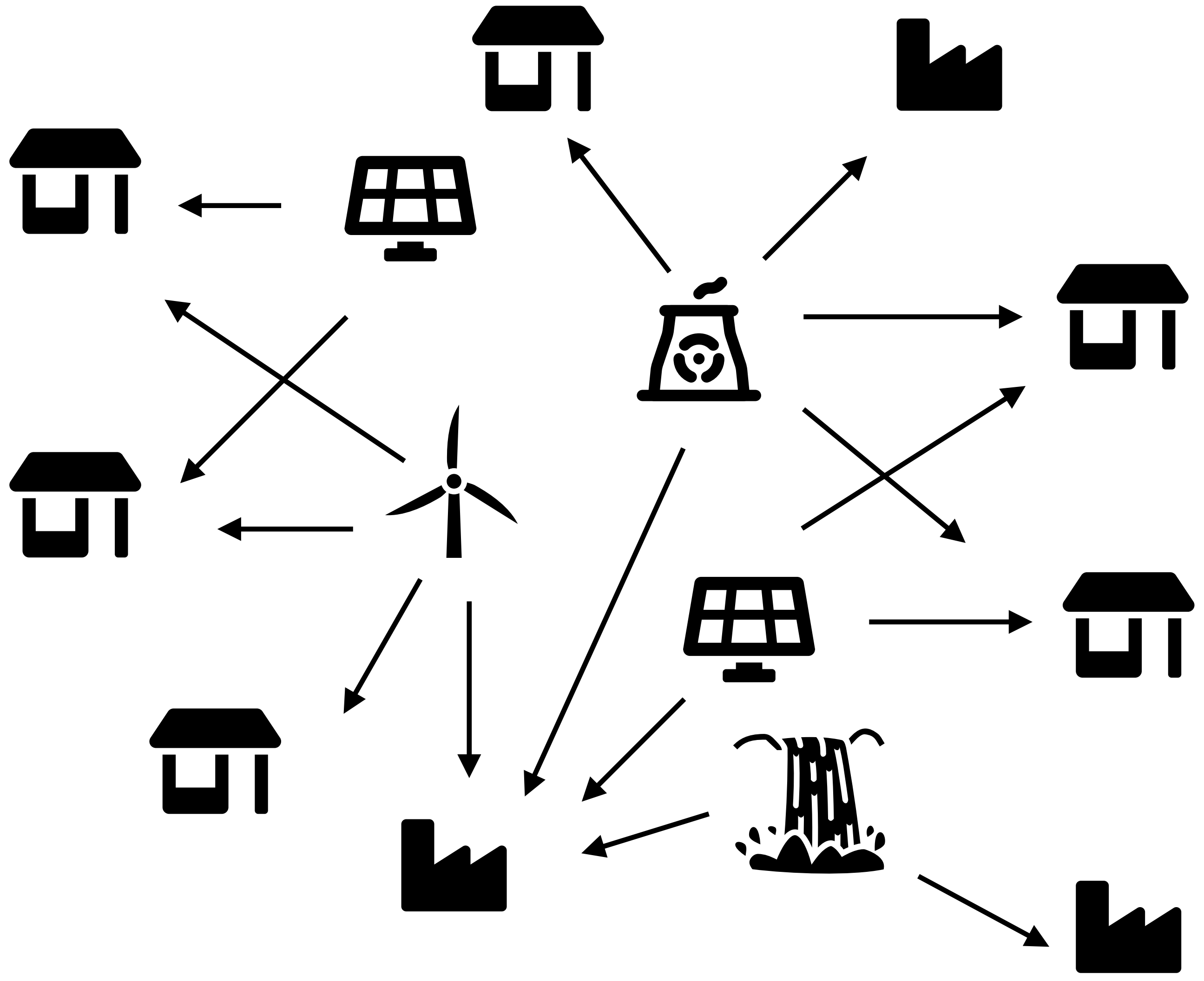
(clustering makes no difference)

average $\bar{d}$

Classical Robust Optimization reformulation

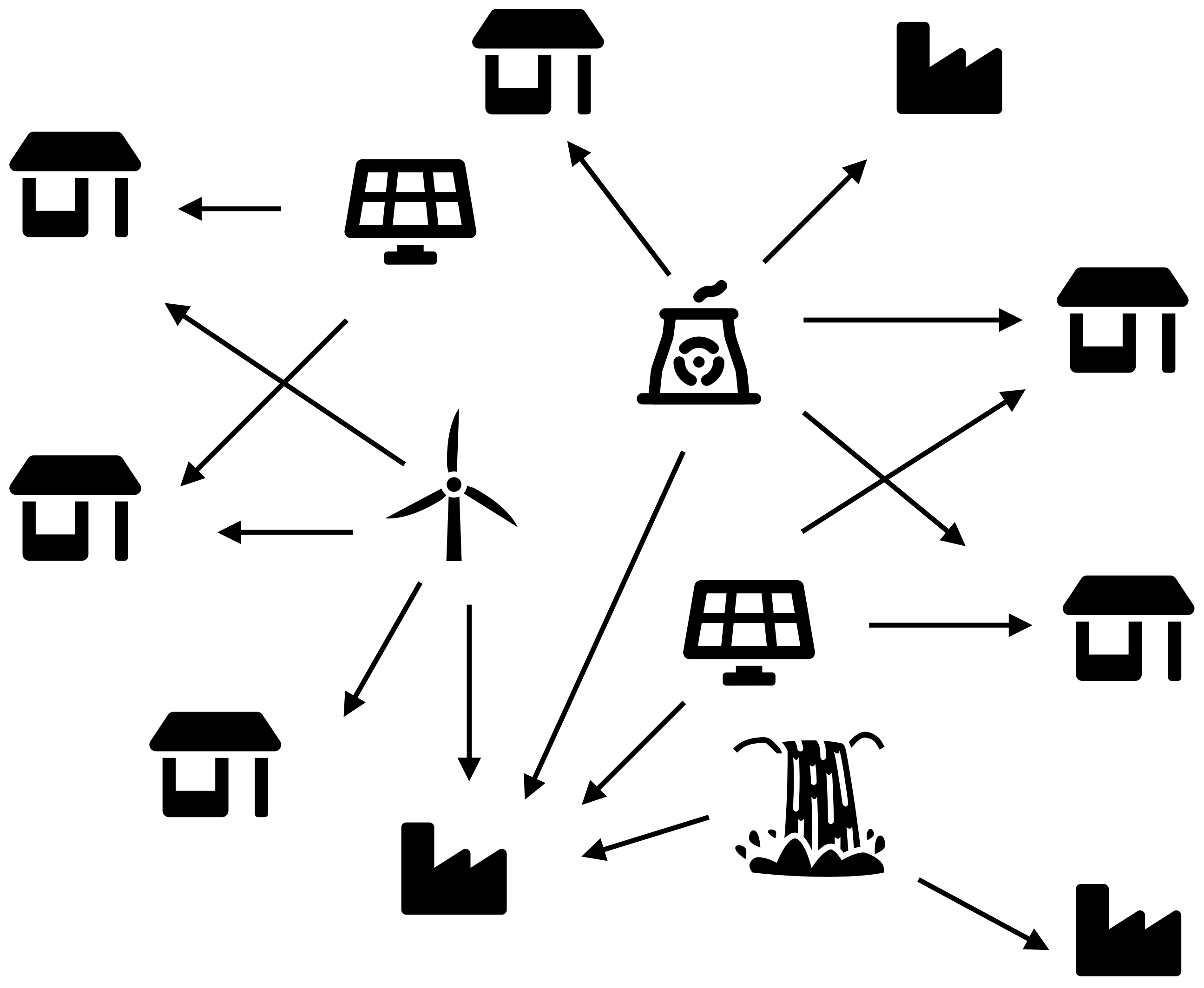
Numerical examples

Facility location example



$$\begin{array}{ll} \text{minimize} & c^T x + \mathbf{tr}(C^T X) \\ \text{subject to} & \mathbf{1}^T X_j = 1, \quad j = 1, \ldots, m \\ & (X^T)_i u \leq r_i x_i, \quad i = 1, \ldots, n \\ & x \in \{0,1\}^n, \quad X \in \mathbf{R}^{n \times m} \end{array}$$

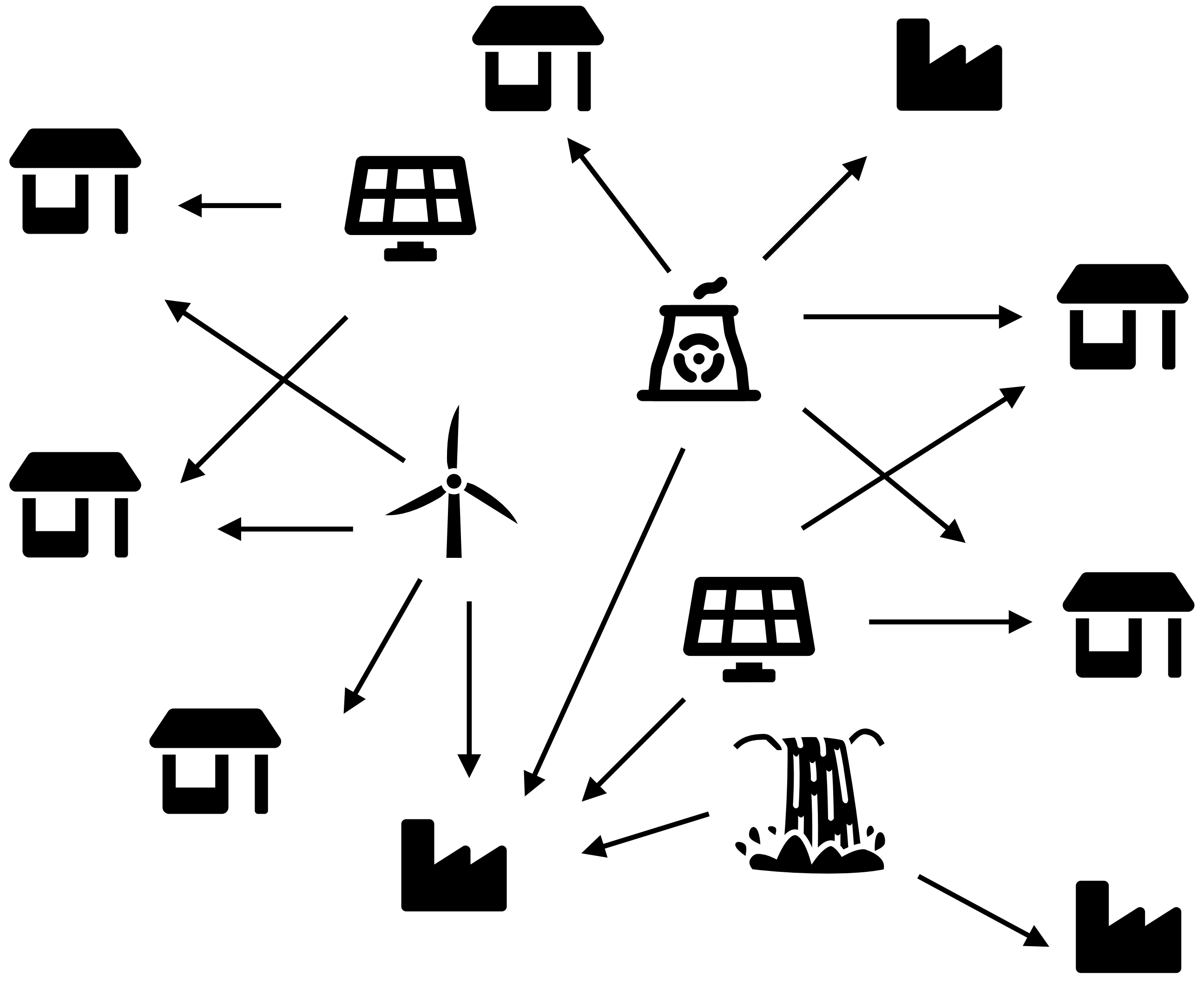
Facility location example



cost of
opening facilities

$$\begin{array}{ll} \text{minimize} & c^T x + \mathbf{tr}(C^T X) \\ \text{subject to} & \mathbf{1}^T X_j = 1, \quad j = 1, \dots, m \\ & (X^T)_i u \leq r_i x_i, \quad i = 1, \dots, n \\ & x \in \{0,1\}^n, \quad X \in \mathbf{R}^{n \times m} \end{array}$$

Facility location example

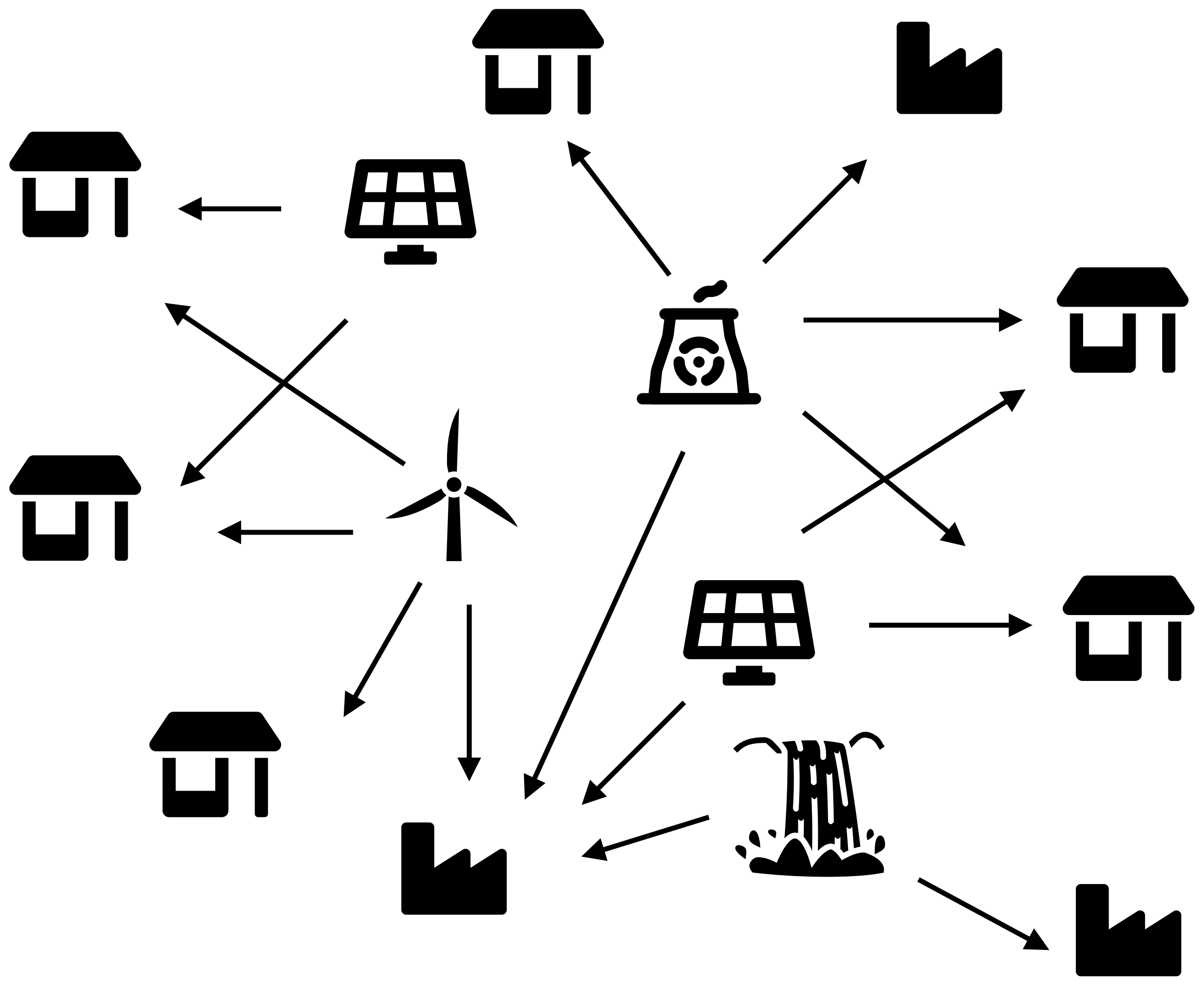


cost of
opening facilities

cost of
energy distribution

$$\begin{array}{ll}\text{minimize} & \underbrace{c^T x}_{} + \mathbf{tr}(C^T X) \\ \text{subject to} & \mathbf{1}^T X_j = 1, \quad j = 1, \dots, m \\ & (X^T)_i u \leq r_i x_i, \quad i = 1, \dots, n \\ & x \in \{0, 1\}^n, \quad X \in \mathbf{R}^{n \times m}\end{array}$$

Facility location example



cost of opening facilities

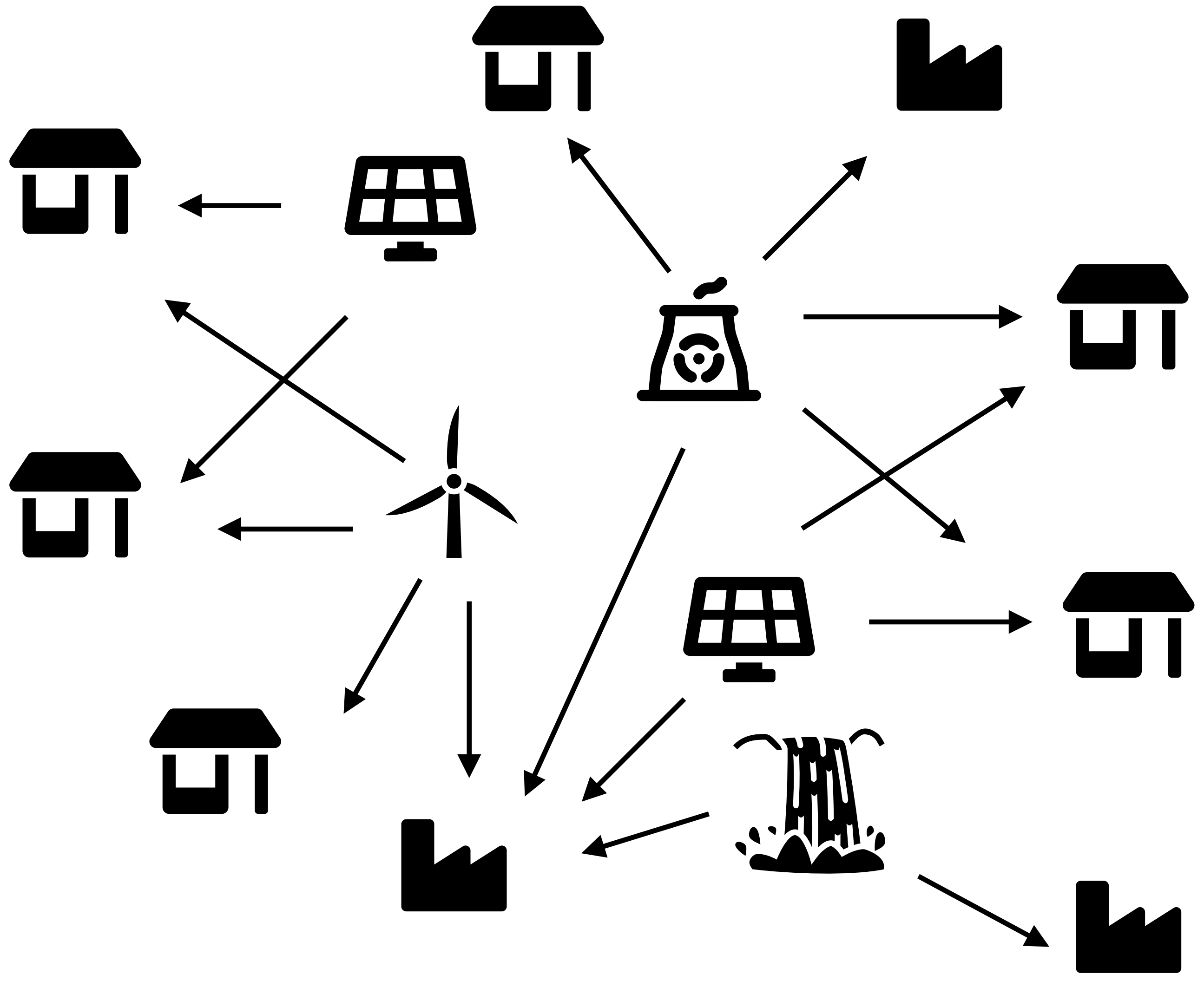
cost of energy distribution

$$\begin{array}{ll} \text{minimize} & c^T x + \mathbf{tr}(C^T X) \\ \text{subject to} & \mathbf{1}^T X_j = 1, \quad j = 1, \dots, m \\ & \boxed{(X^T)_i u \leq r_i x_i, \quad i = 1, \dots, n} \\ & x \in \{0,1\}^n, \quad X \in \mathbf{R}^{n \times m} \end{array}$$

capacity constraints

$g(u, x, X) \leq 0$

Facility location example



cost of
opening facilities

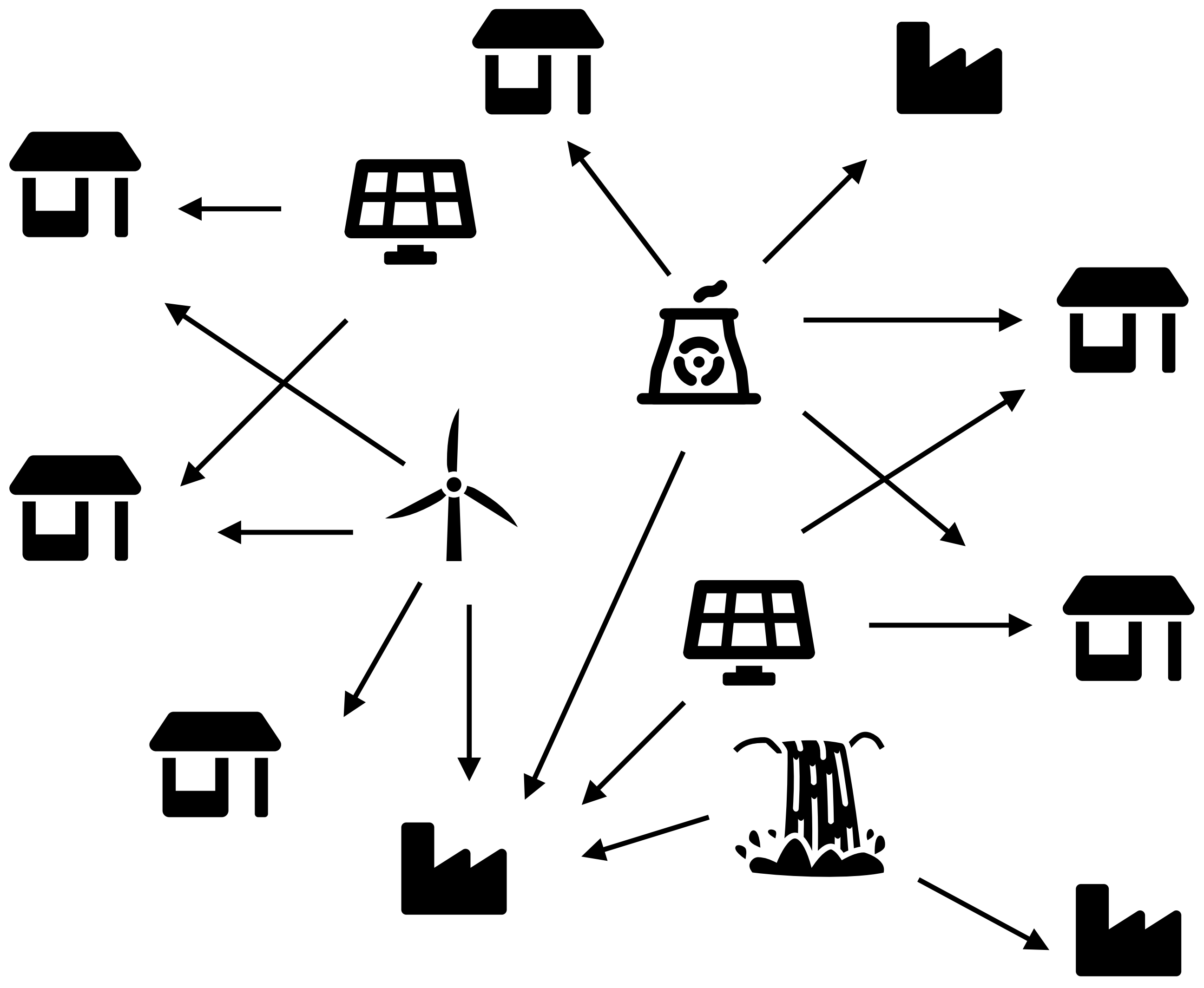
cost of
energy distribution

$$\begin{array}{ll} \text{minimize} & c^T x + \mathbf{tr}(C^T X) \\ \text{subject to} & \mathbf{1}^T X_j = 1, \quad j = 1, \ldots, m \\ & \boxed{(X^T)_i u \leq r_i x_i, \quad i = 1, \ldots, n} \\ & x \in \{0, 1\}^n, \quad X \in \mathbf{R}^{n \times m} \end{array}$$

vector of uncertain
energy demands

capacity
constraints
$g(u, x, X) \leq 0$

Facility location example



cost of
opening facilities

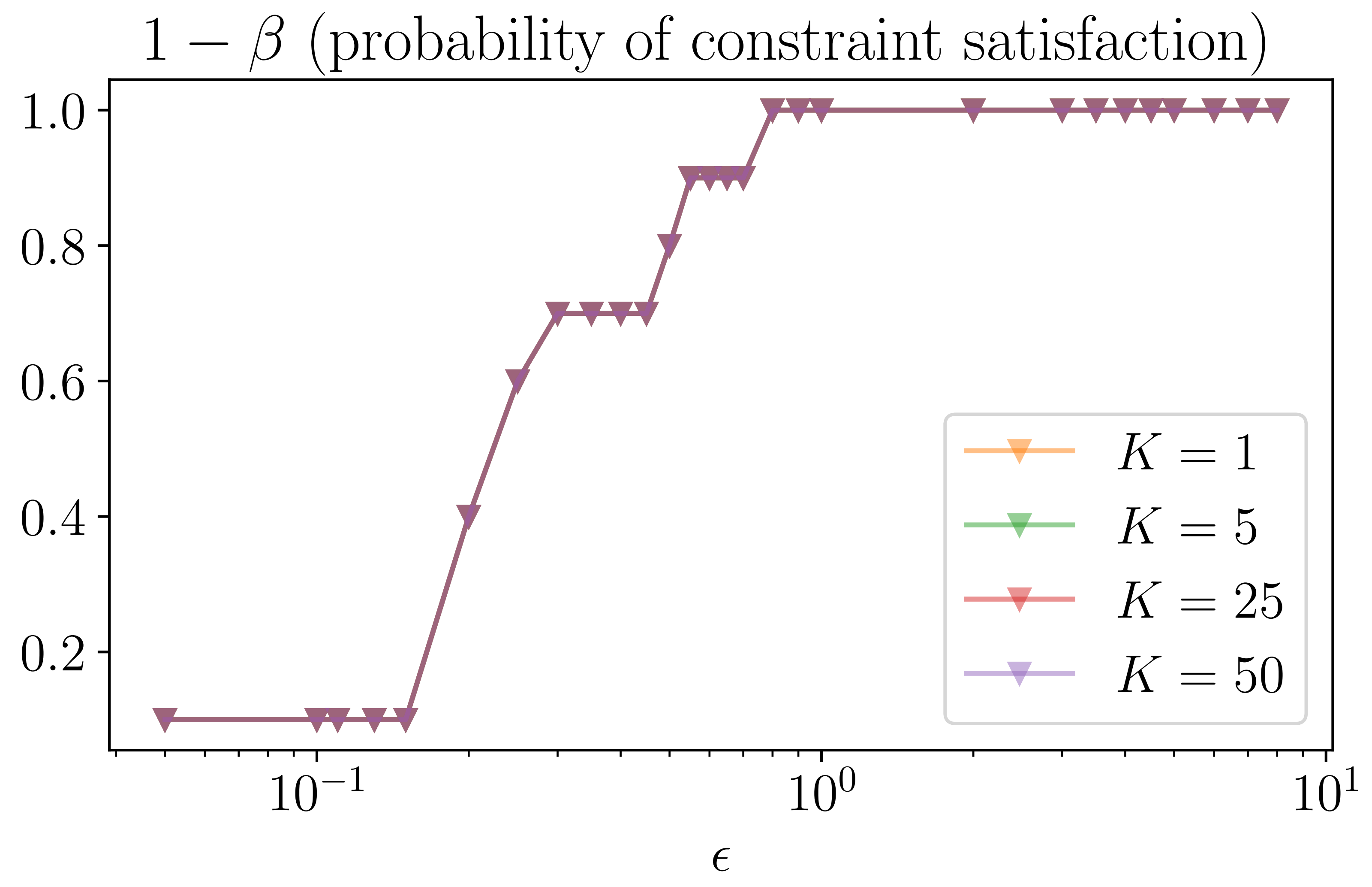
cost of
energy distribution

$$\begin{array}{ll} \text{minimize} & c^T x + \mathbf{tr}(C^T X) \\ \text{subject to} & \mathbf{1}^T X_j = 1, \quad j = 1, \dots, m \\ & \boxed{(X^T)_i u \le r_i x_i, \quad i = 1, \dots, n} \\ & x \in \{0,1\}^n, \quad X \in \mathbf{R}^{n \times m} \end{array}$$

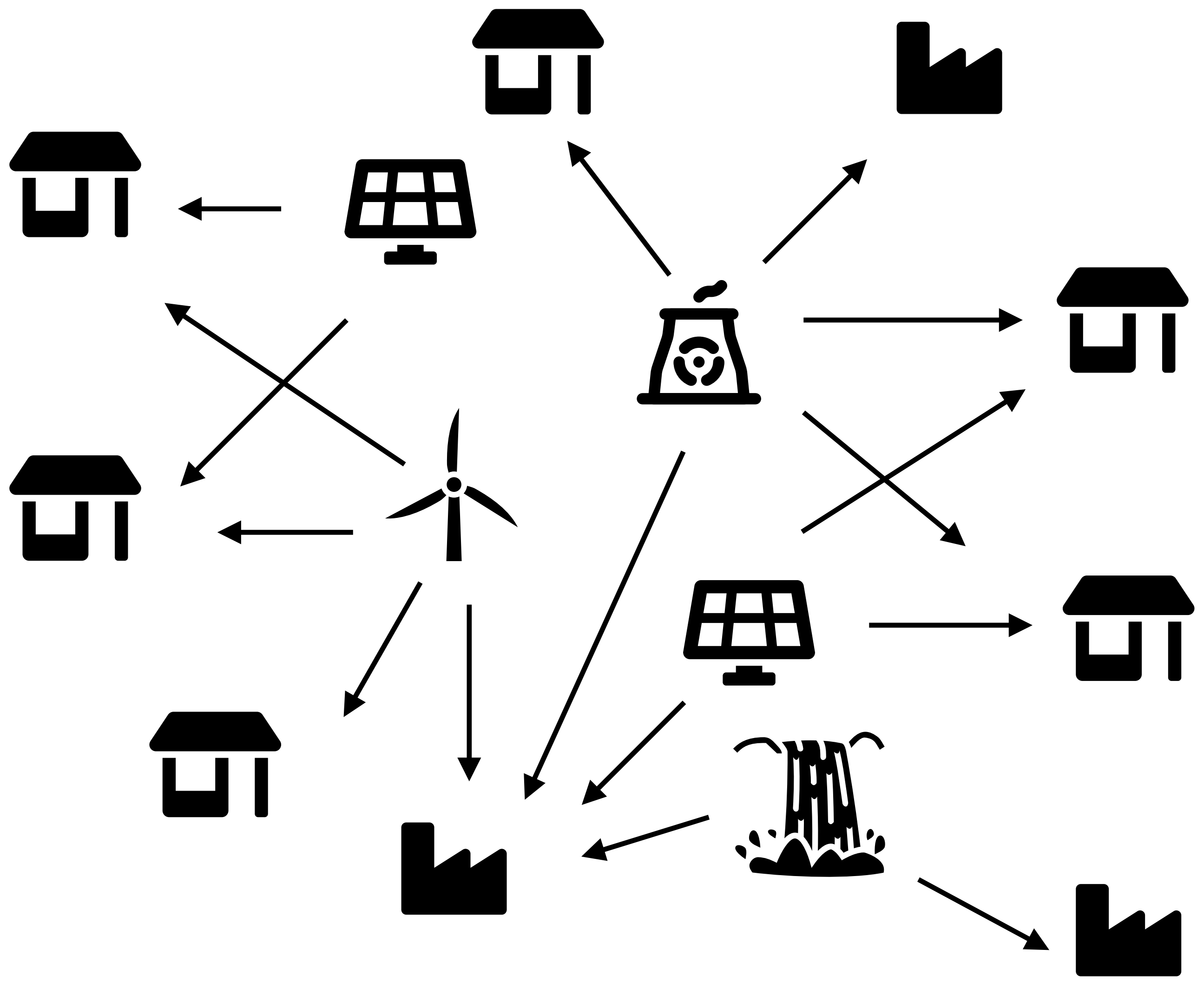
capacity
constraints
$g(u, x, X) \le 0$

vector of uncertain
energy demands

clustering does not affect constraint satisfaction



Facility location example



cost of opening facilities

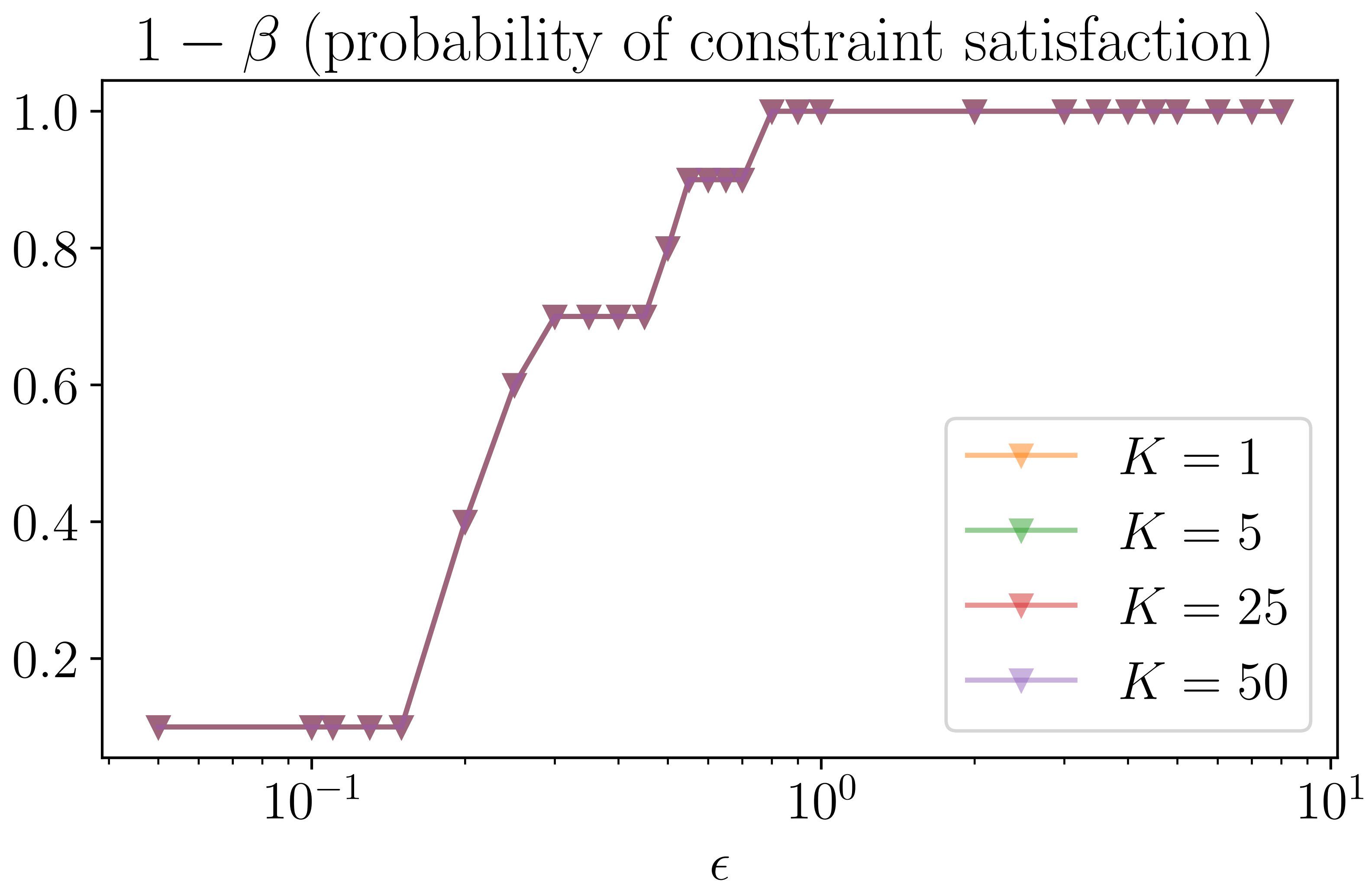
cost of energy distribution

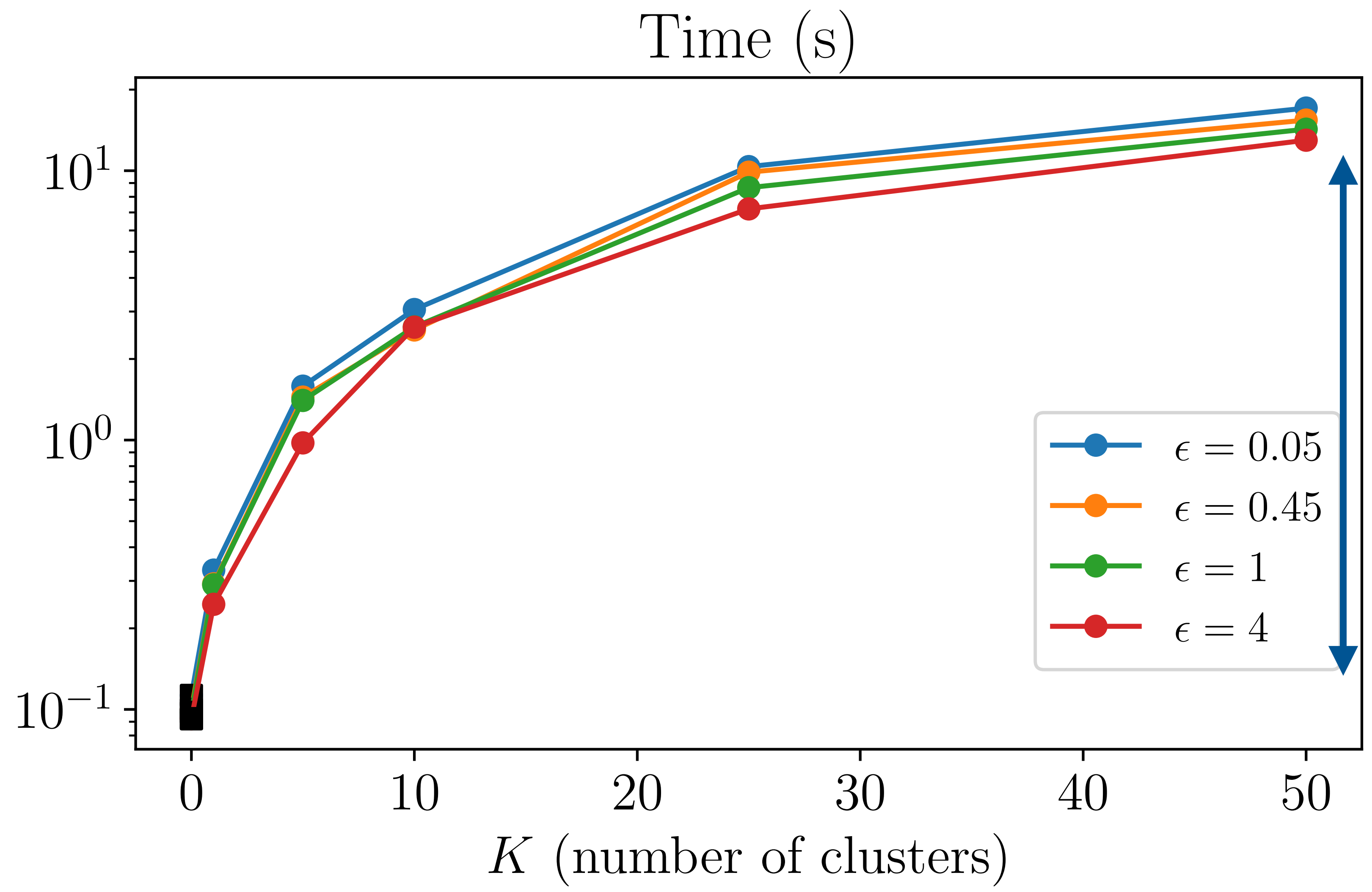
$$\begin{array}{ll} \text{minimize} & c^T x + \mathbf{tr}(C^T X) \\ \text{subject to} & \mathbf{1}^T X_j = 1, \quad j = 1, \ldots, m \\ & \boxed{(X^T)_i u \leq r_i x_i, \quad i = 1, \ldots, n} \\ & x \in \{0, 1\}^n, \quad X \in \mathbf{R}^{n \times m} \end{array}$$

vector of uncertain energy demands

capacity constraints $g(u, x, X) \leq 0$

clustering does not affect constraint satisfaction





100x speedups!

Capital budgeting example

Problem

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

Capital budgeting example

total
net present value (NPV)

Problem

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

Capital budgeting example

total
net present value (NPV)

Problem

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

budget
constraint

Capital budgeting example

Problem

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

Capital budgeting example

Problem

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

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

NPV of project j

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

cash
flow

discount
rate

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

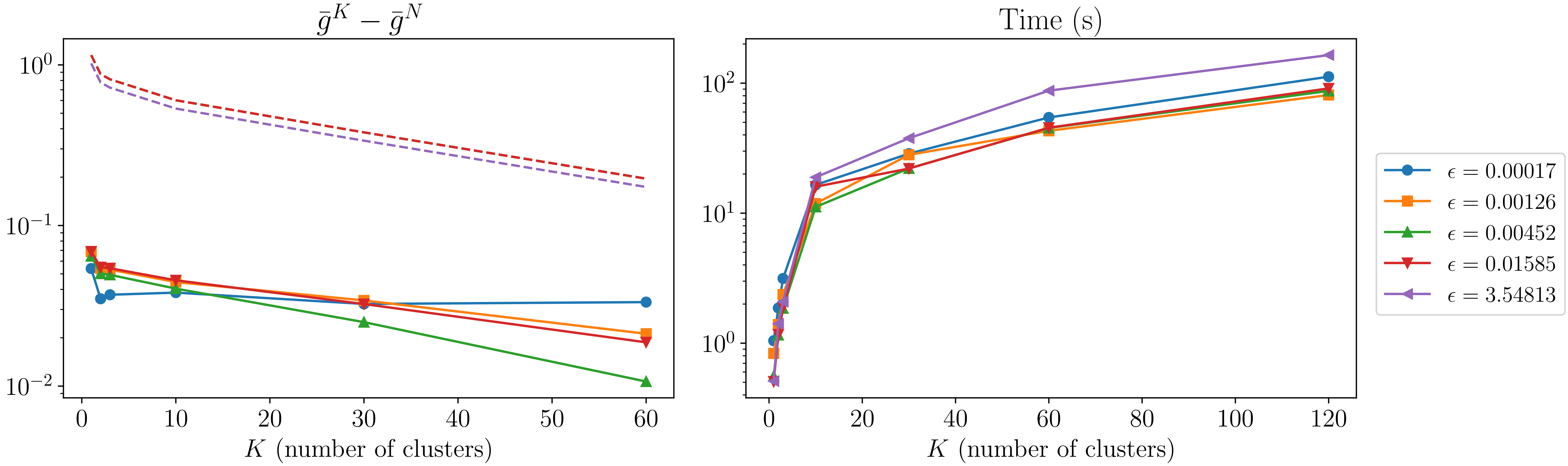
$$\eta_j(u) = \sum_{t=1}^{T} \frac{F_{jt}}{(1 + u_j)^t}$$

cash flow

discount rate

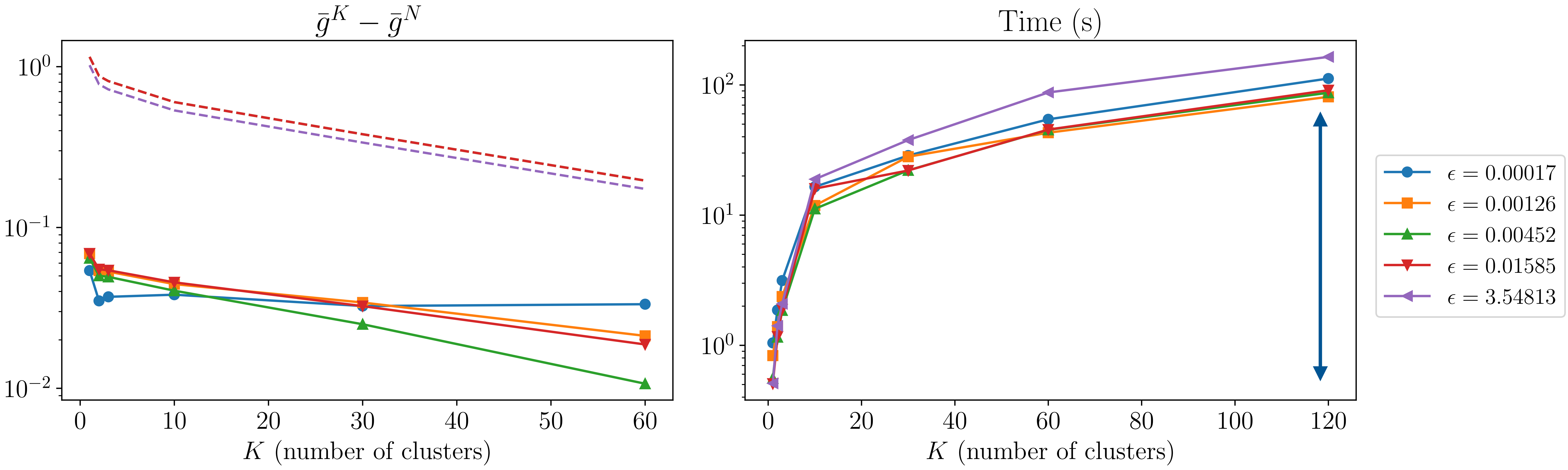
Capital budgeting results

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



Capital budgeting results

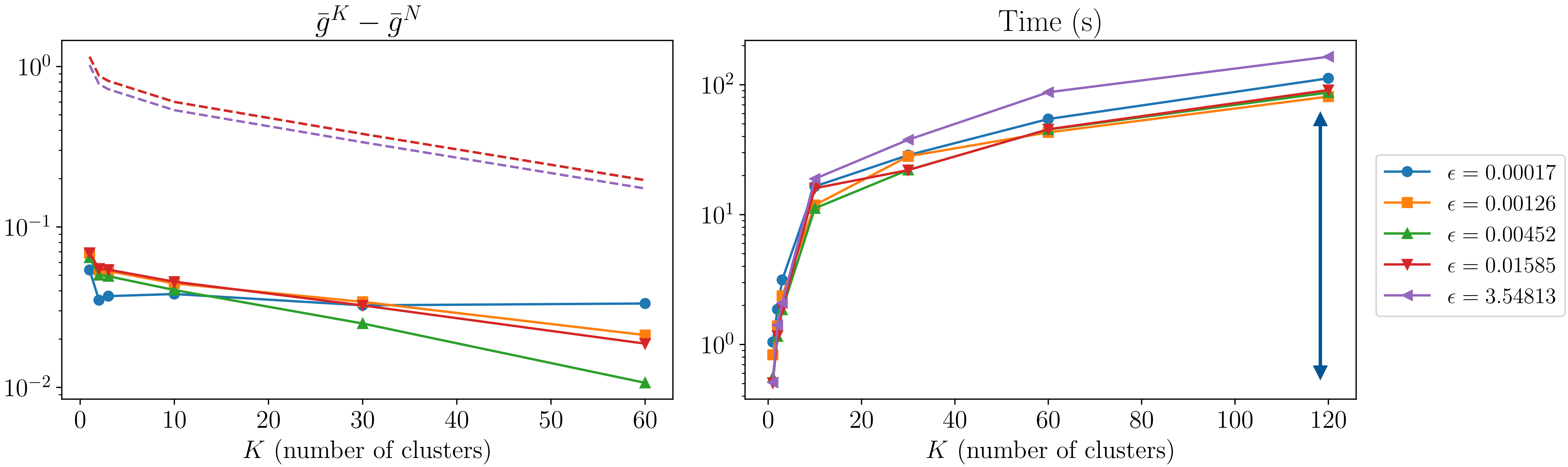
$n = 20, N = 120, T = 5$



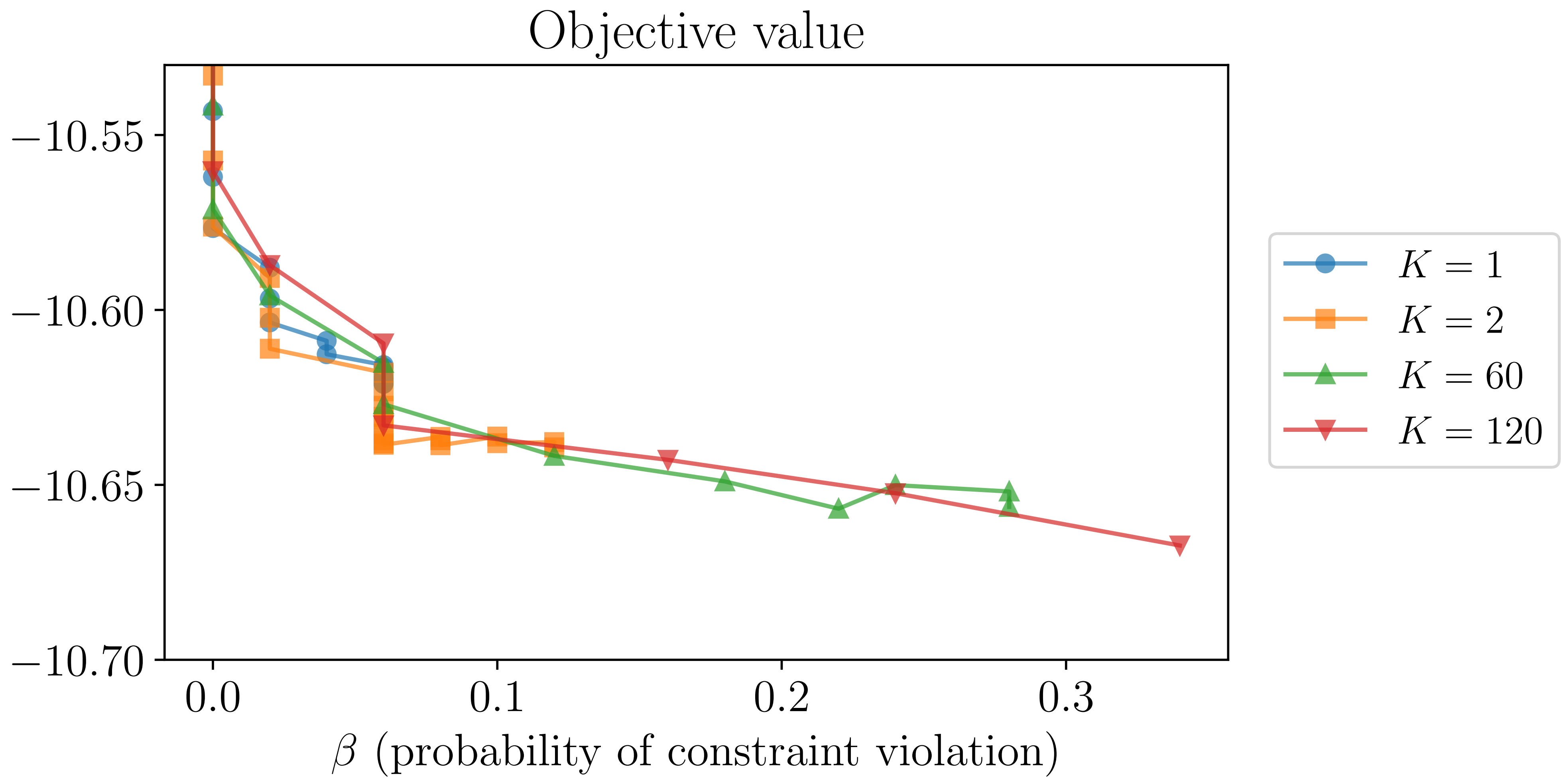
100x speedups!

Capital budgeting results

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

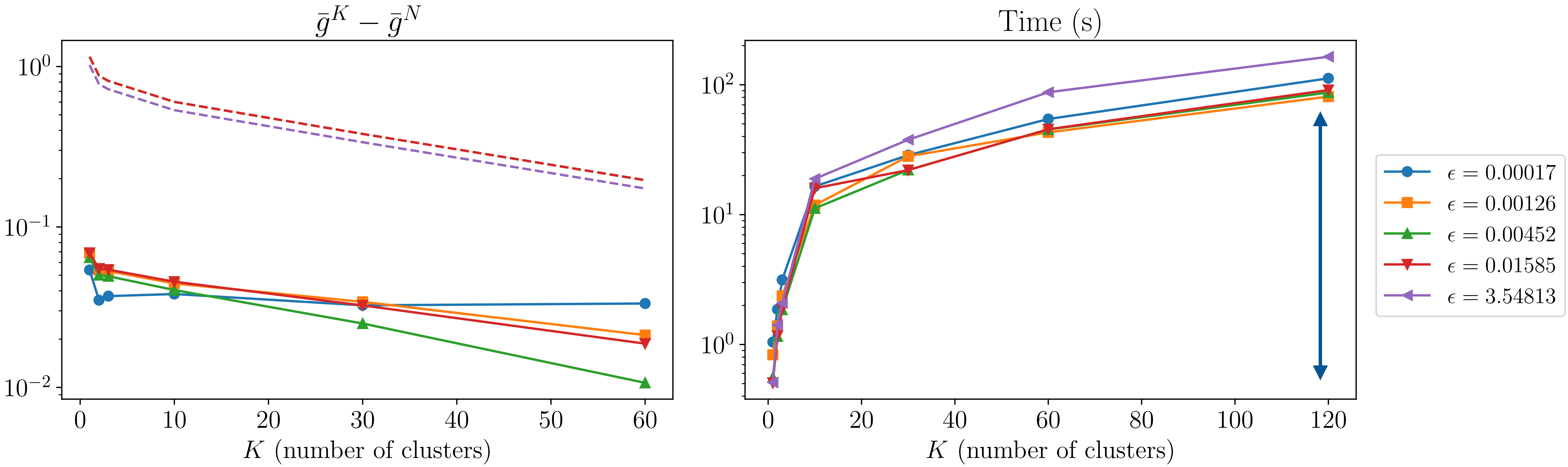


100x speedups!

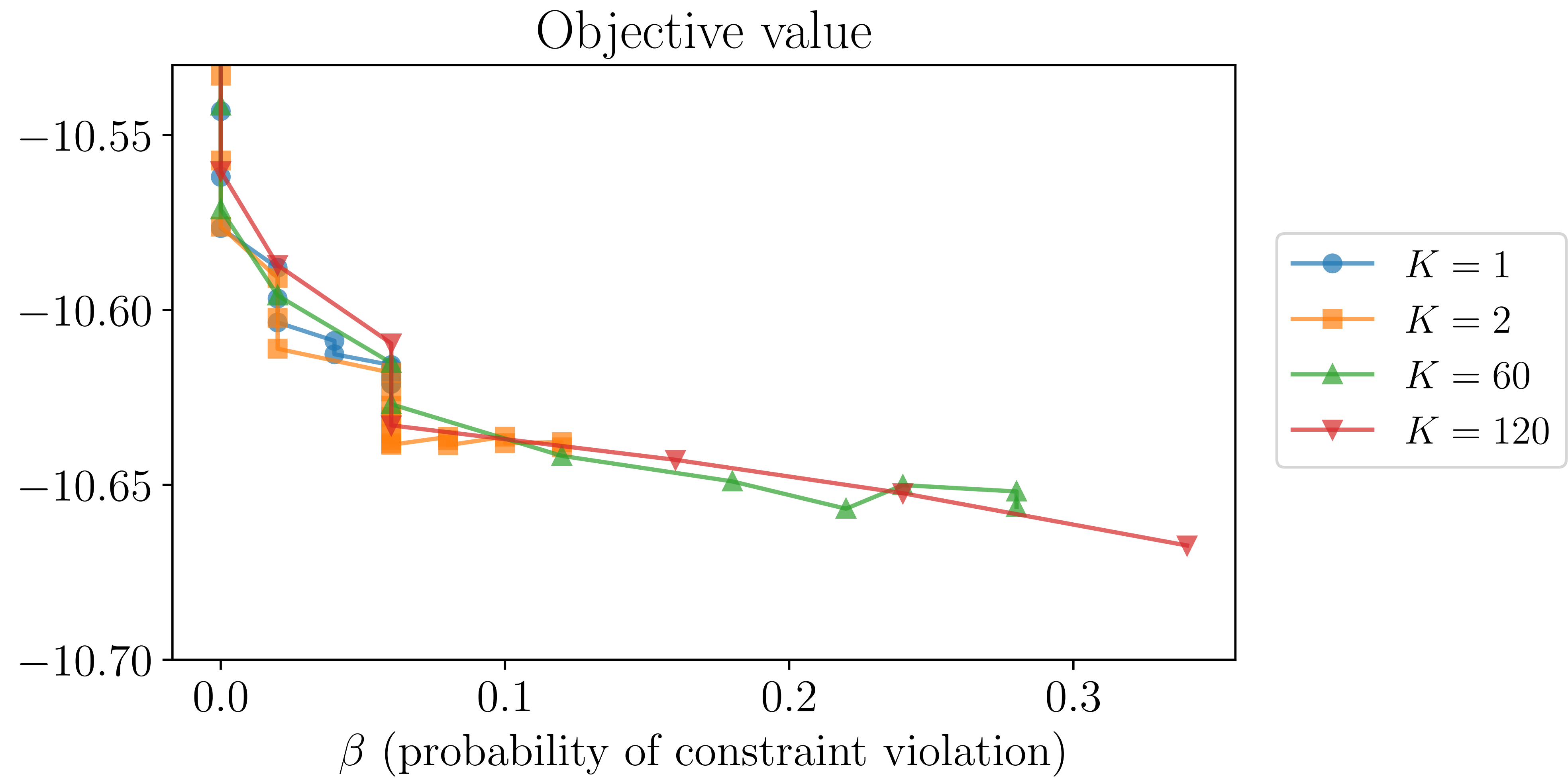


Capital budgeting results

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



100x speedups!



2 clusters give near-optimal performance

Conclusions

Acknowledgements

Irina Wang

Princeton

Cole Becker

Princeton

Bart Van Parys

MIT

Mean Robust Optimization

- Bridge RO and DRO

- Clustering effect

- Multiple orders of magnitude speedups

Mean Robust Optimization

- Bridge RO and DRO

- Clustering effect

- Multiple orders of magnitude speedups

Mean Robust Optimization

- **Bridge** RO and DRO

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

- Multiple **orders of magnitude** speedups

Mean Robust Optimization

- Bridge RO and DRO

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

- Multiple orders of magnitude speedups

Mean Robust Optimization

- Bridge RO and DRO

- Clustering effect
  - g affine in u $\longrightarrow$ zero clustering effect!
  - g concave in u $\longrightarrow$ performance bound

- Multiple orders of magnitude speedups

https://github.com/stellatogrp/mro_experiments

Mean Robust Optimization
I. Wang, C. Becker, B. Van Parys, and B. Stellato
arXiv e-prints:2207.10820, 2022

INFORMS Computing Society
Student Paper Award Winner

Takeaway message

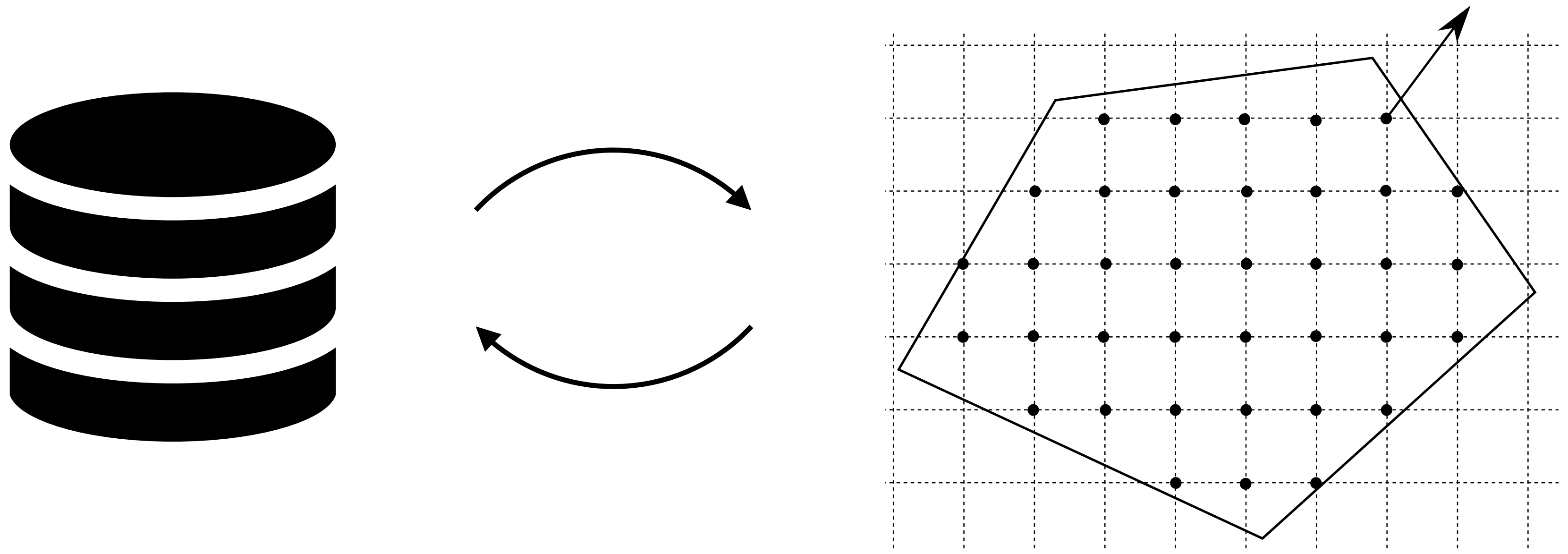
Machine Learning
can help us
formulate optimization problems

Takeaway message

Machine Learning
can help us
formulate optimization problems

Many opportunities ahead!

Data **Formulations**



stellato.io bstellato@princeton.edu @b_stellato