The Rare Eclipse Problem on Tiles: Quantized Embeddings of Disjoint Convex Sets

Valerio Cambareri Chunlei Xu Laurent Jacques



Image and Signal Processing Group,
ICTEAM/ELEN, University of Louvain (UCL), Belgium

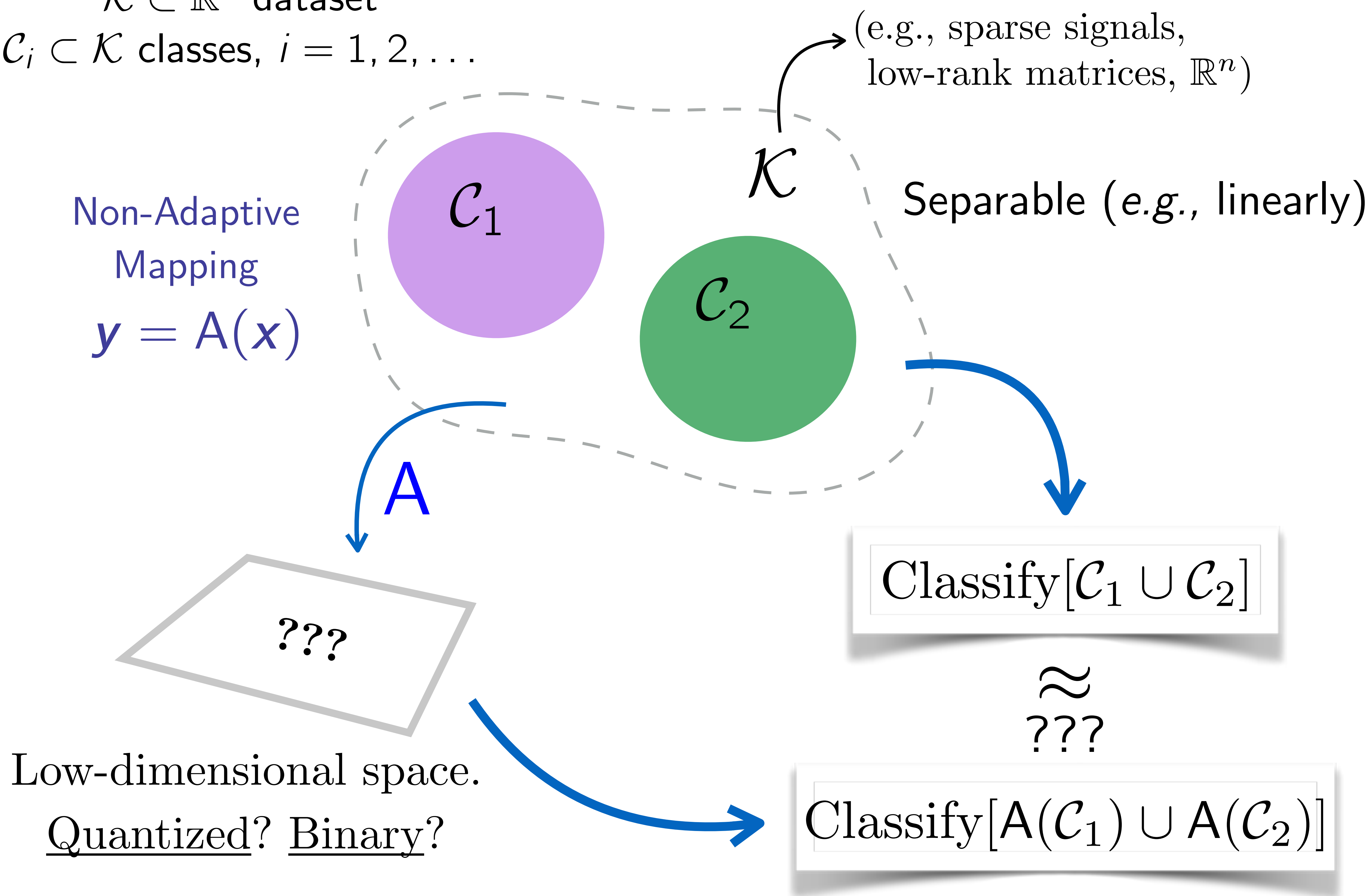
SampTA'17, Tallinn, Estonia

UCL Université catholique de Louvain icteam ISP Group fnrs LA LIBERTÉ DE CHERCHER

The Big Picture

$\mathcal{K} \subset \mathbb{R}^n$ dataset

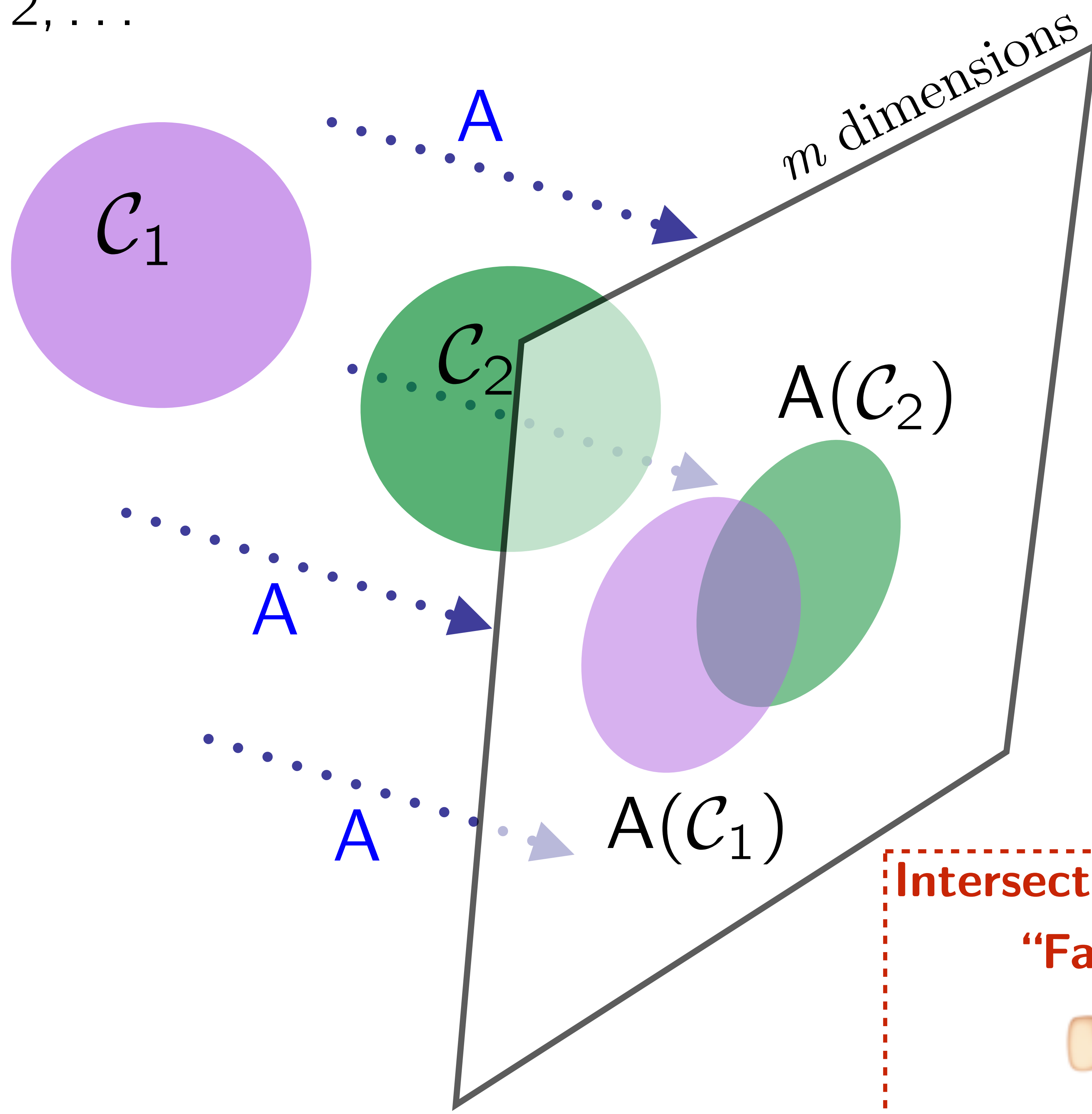
$\mathcal{C}_i \subset \mathcal{K}$ classes, $i = 1, 2, \ldots$



The Big Picture

$\mathcal{K} \subset \mathbb{R}^n$ dataset
$\mathcal{C}_i \subset \mathcal{K}$ classes, $i = 1, 2, \ldots$

Non-Adaptive Mapping
$\boldsymbol{y} = \mathsf{A}(\boldsymbol{x})$

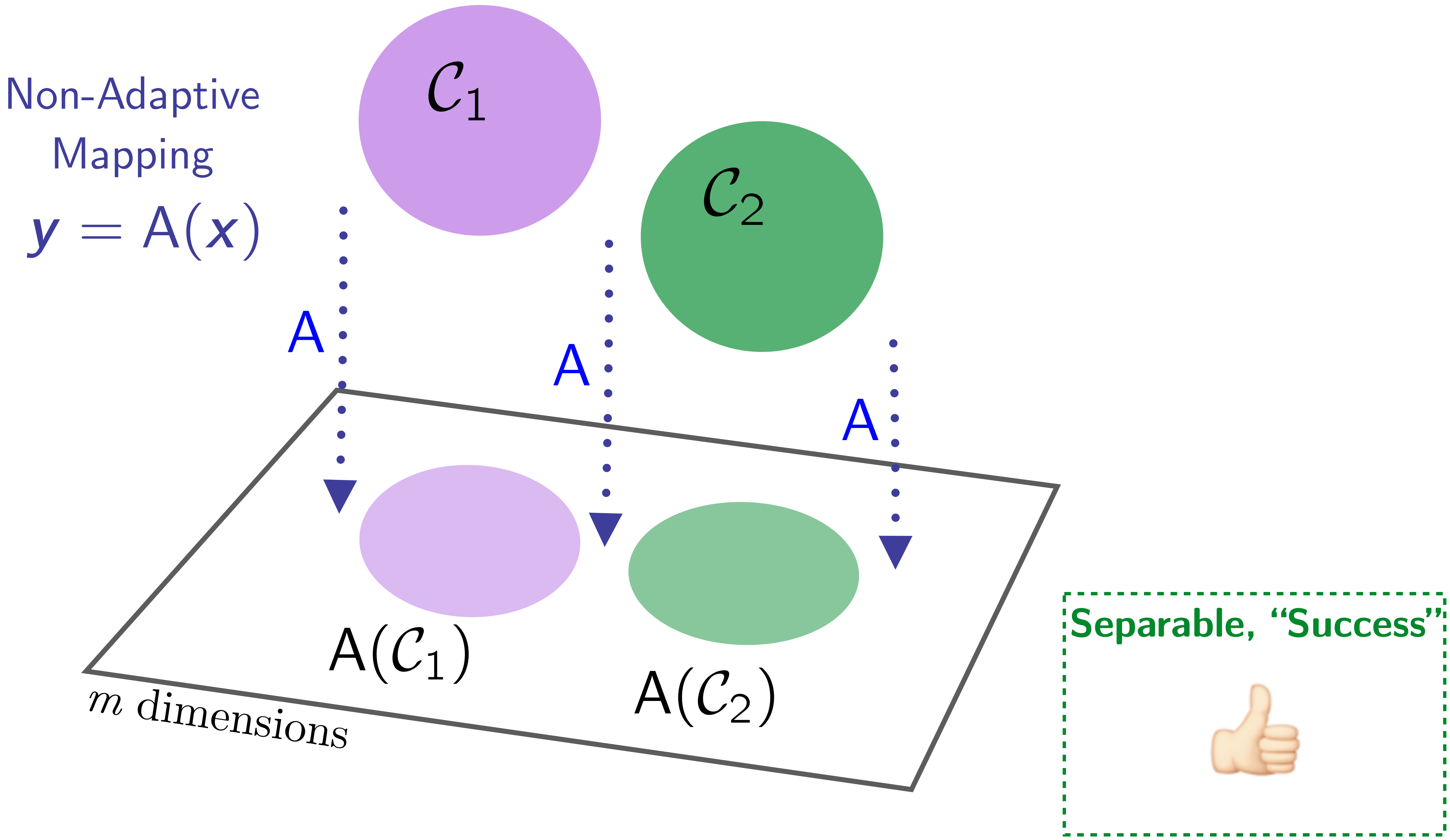


Intersecting/Eclipse
"Failure"

The Big Picture

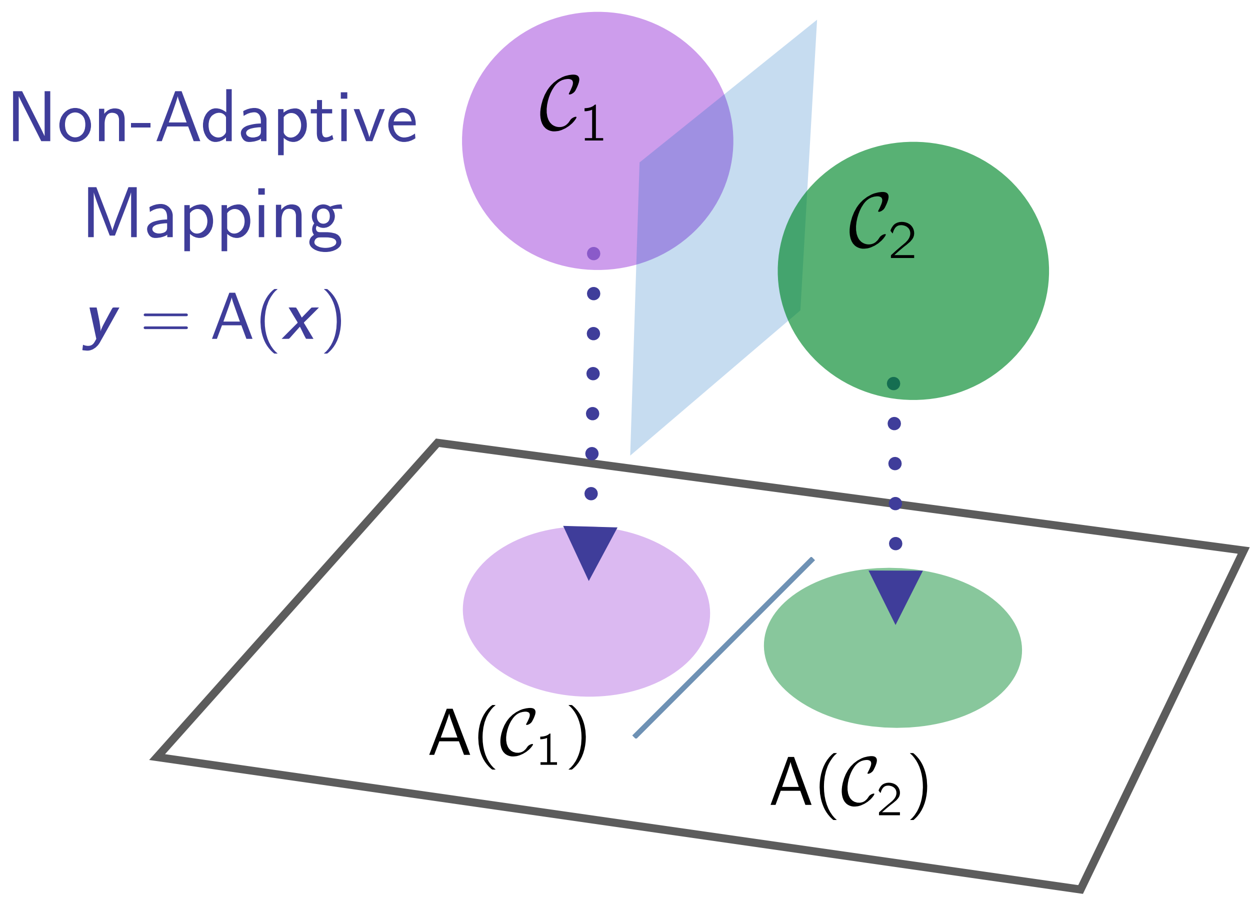
$\mathcal{K} \subset \mathbb{R}^n$ dataset

$\mathcal{C}_i \subset \mathcal{K}$ classes, $i = 1, 2, \ldots$



The Big Picture

How does the *probability of error* of a (generic) learning task depend on m, $\mathsf{A}, \mathcal{K}$? What if A is (*mildly*) non-linear?



Related works on "(linear) compressive classification":

Davenport *et al.* '07-'10, Haupt *et al.* '06,

Reboredo *et al.* '13-'16,

Bandeira, Mixon, Recht '14 [BMR '14]

The Linear Case

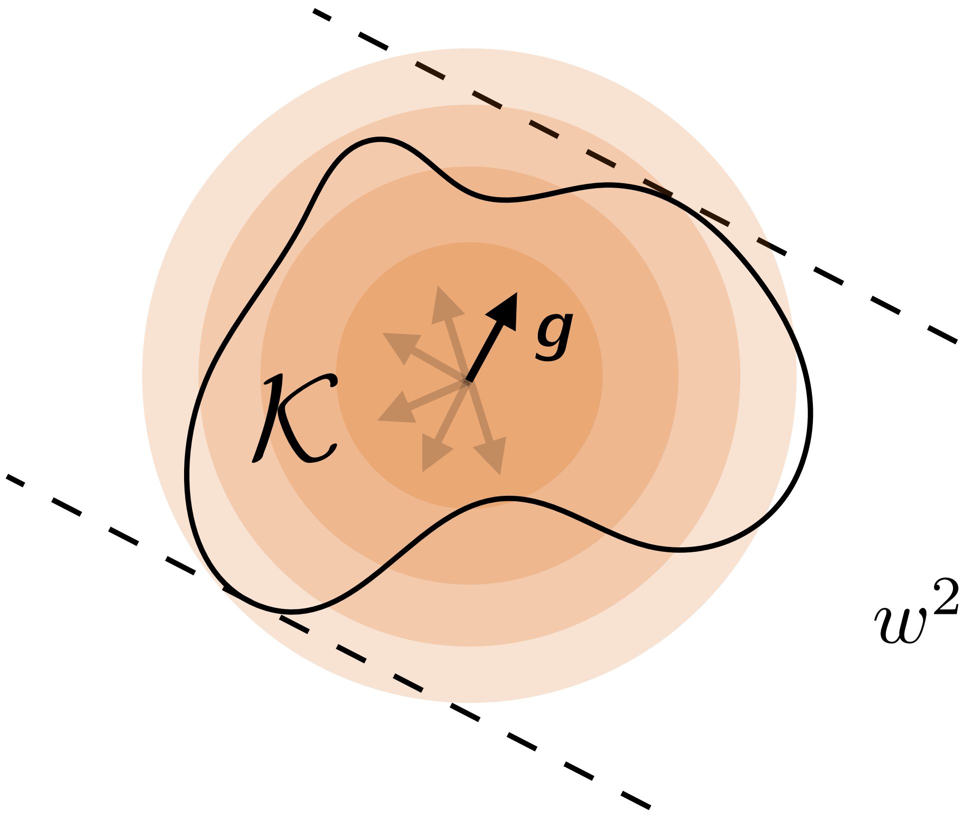
Bandeira, Mixon, Recht '14 [BMR '14]

Parenthesis: Useful Tool

Gaussian Width (GW):

Let $\mathcal{K} \subset \mathbb{R}^n$, $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$,

$$w(\mathcal{K}) = \mathbb{E}_{\boldsymbol{g}} \sup_{\boldsymbol{x} \in \mathcal{K}} |\langle \boldsymbol{x}, \boldsymbol{g} \rangle|$$



Examples:

$$w^2(\mathcal{K}) \lesssim \log |\mathcal{K}|$$
$$w^2(\mathbb{B}^n) \lesssim n$$
$$w^2(\Sigma_k^n \cap \mathbb{B}^n) \lesssim k \log(n/k)$$
$$w^2(\mathcal{M}_r \cap \mathbb{B}_F^{n \times n}) \lesssim rn$$
$$w^2(\cup_{i=1}^T \mathcal{K}_i) \lesssim \log T + \max_i w^2(\mathcal{K}_i)$$
$$\vdots$$

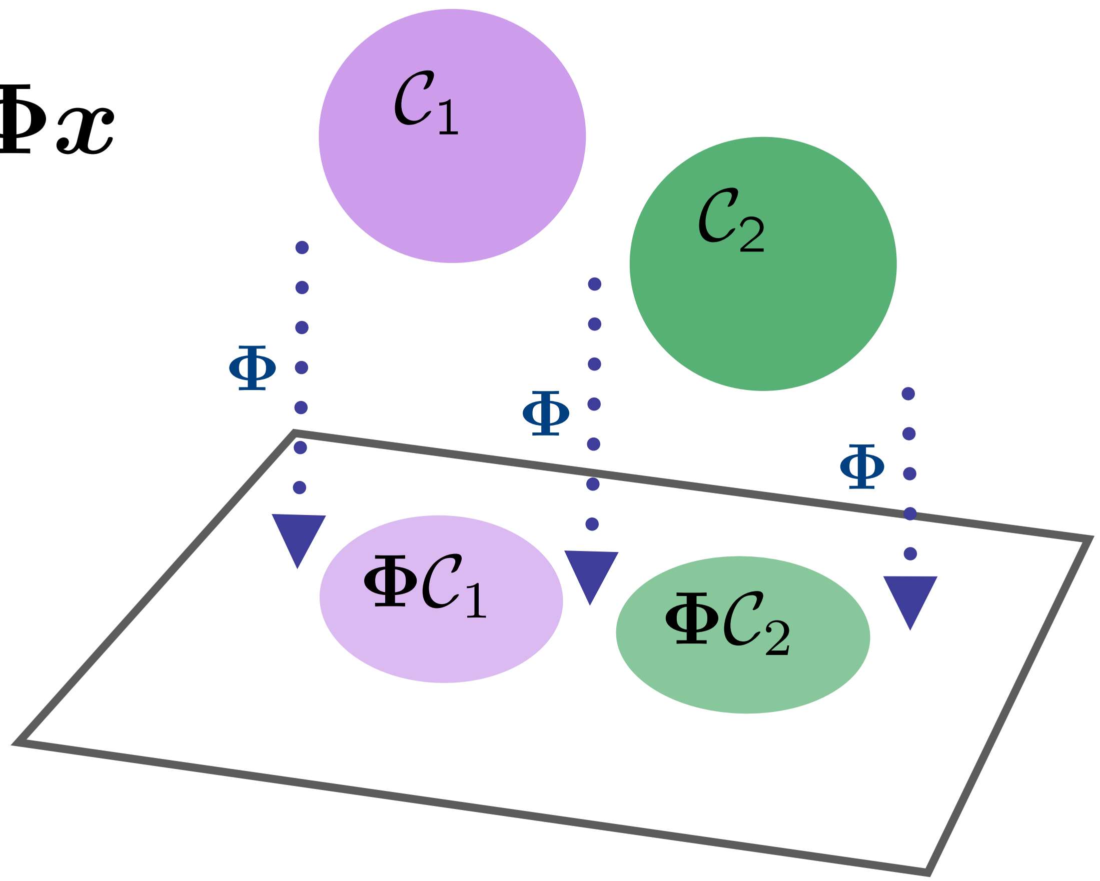
The Rare Eclipse Problem

Problem (Rare Eclipse Problem (Bandeira *et al.* '14)).

Let $\mathcal{C}_1, \mathcal{C}_2 \subset \mathbb{R}^n : \mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ be closed convex sets, $\mathbf{\Phi} \sim \mathcal{N}^{m \times n}(0, 1)$. Given $\eta \in (0, 1)$, find the smallest m so that

$$p_0 := \mathbb{P}_{\mathbf{\Phi}}[\mathbf{\Phi}\mathcal{C}_1 \cap \mathbf{\Phi}\mathcal{C}_2 = \emptyset] \geq 1 - \eta.$$

$$\mathsf{A}(\boldsymbol{x}) = \mathbf{\Phi}\boldsymbol{x}$$



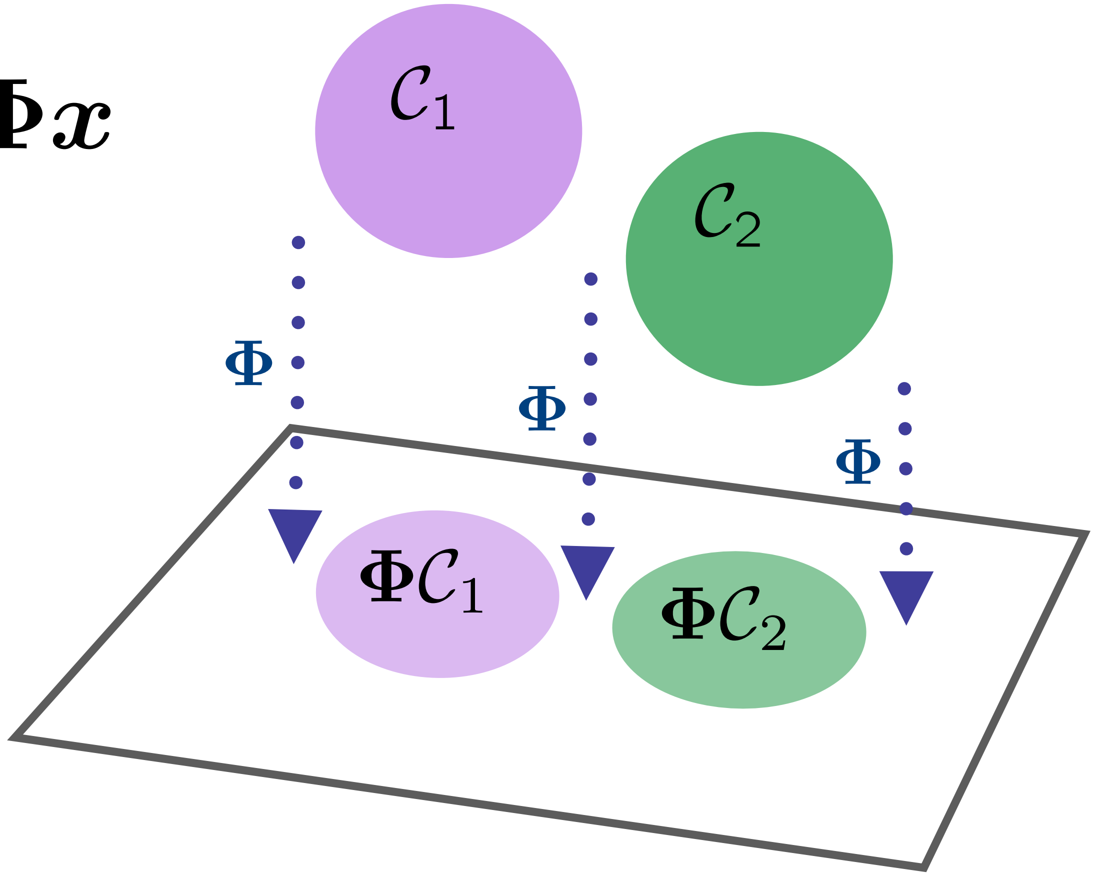
The Rare Eclipse Problem

Problem (Rare Eclipse Problem (Bandeira *et al.* '14)).

Let $\mathcal{C}_1, \mathcal{C}_2 \subset \mathbb{R}^n : \mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ be closed convex sets, $\mathbf{\Phi} \sim \mathcal{N}^{m \times n}(0,1)$.
Given $\eta \in (0,1)$, find the smallest m so that

$$p_0 = \mathbb{P}_{\mathbf{\Phi}}[\forall \boldsymbol{x}_1 \in \mathcal{C}_1, \forall \boldsymbol{x}_2 \in \mathcal{C}_2, \Phi(\boldsymbol{x}_1 - \boldsymbol{x}_2) \neq 0] \geq 1 - \eta.$$

$$\mathrm{A}(\boldsymbol{x}) = \mathbf{\Phi}\boldsymbol{x}$$

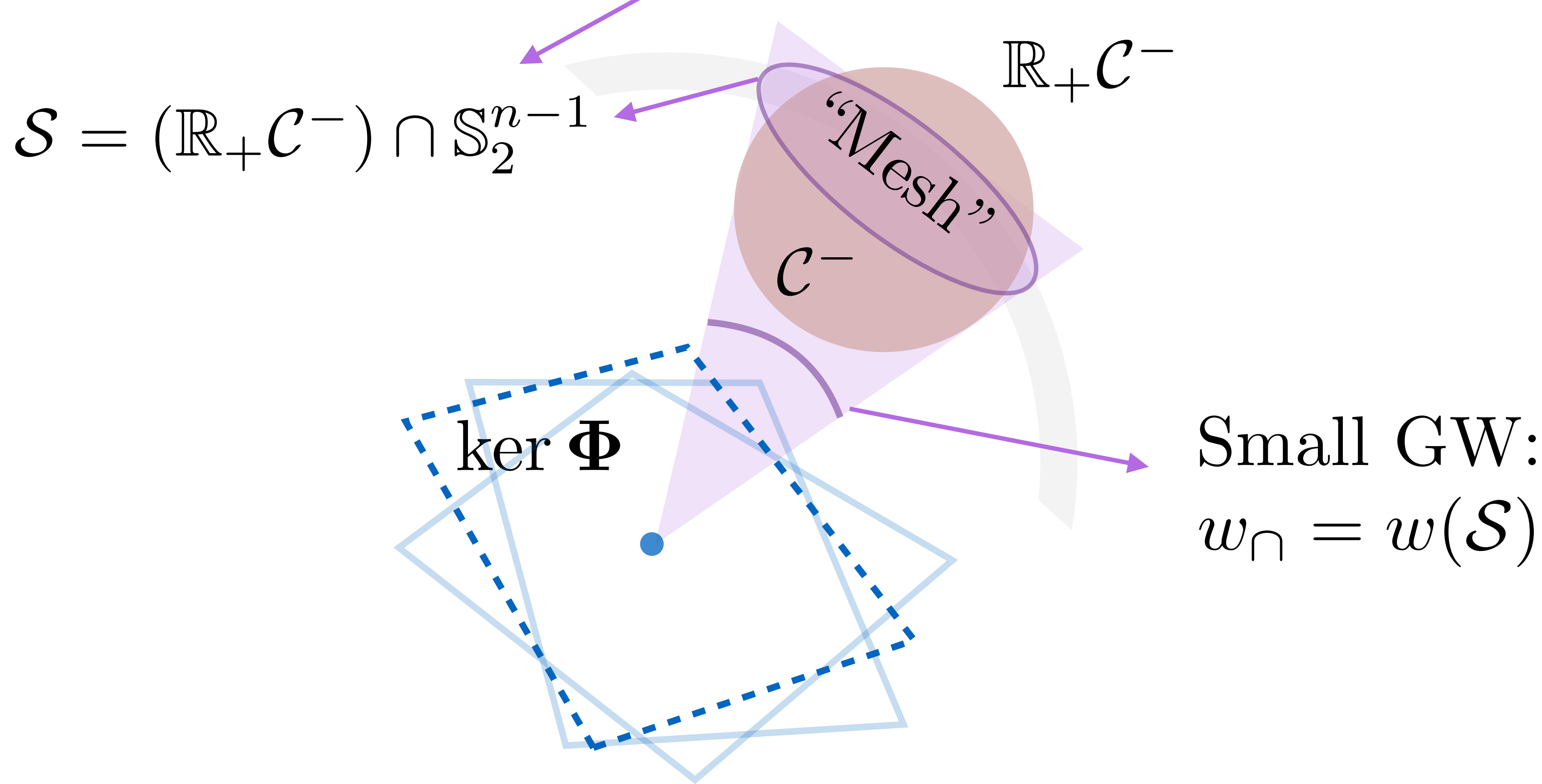


The Rare Eclipse Problem

Problem (Rare Eclipse Problem (Bandeira *et al.* '14)).

Let $\mathcal{C}_1, \mathcal{C}_2 \subset \mathbb{R}^n : \mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ be closed convex sets, $\boldsymbol{\Phi} \sim \mathcal{N}^{m \times n}(0,1)$. Given $\eta \in (0,1)$, find the smallest m so that, with $\mathcal{C}^- = \mathcal{C}_1 - \mathcal{C}_2$,

$$p_0 = \mathbb{P}_{\boldsymbol{\Phi}}[\mathcal{C}^- \cap \ker \boldsymbol{\Phi} = \emptyset] = \mathbb{P}_{\boldsymbol{\Phi}}[\mathcal{S} \cap \ker \boldsymbol{\Phi} = \emptyset] \geq 1 - \eta.$$



The Rare Eclipse Problem

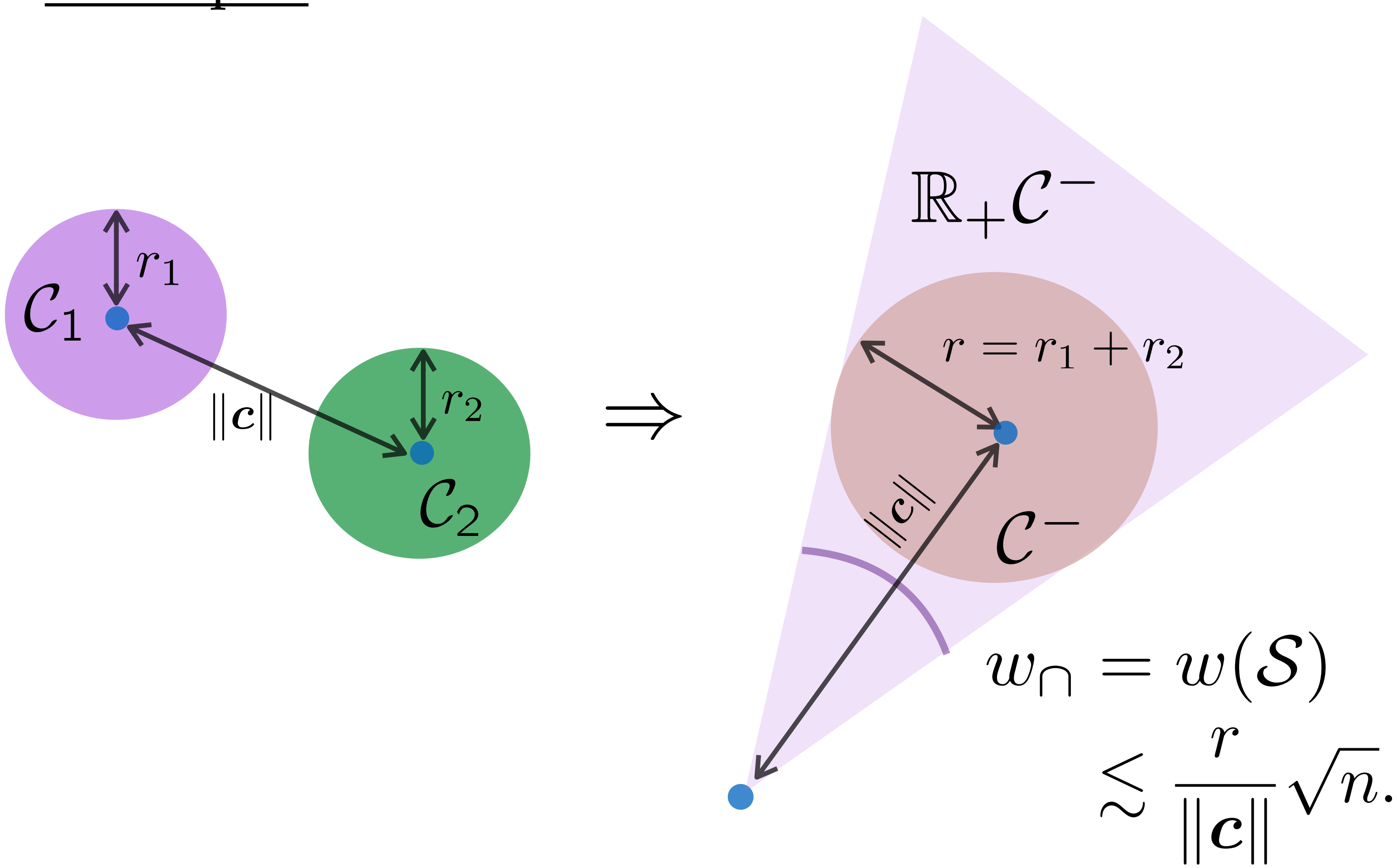
BMR '14: "Gordon's escape through a mesh" theorem

Proposition (Corollary 3.1 in BMR '14).

(& *really* tight [Amelunxen et al, 13])

Given $\eta \in (0,1)$, if $m > \left(w_\cap + \sqrt{2\log\frac{1}{\eta}}\right)^2 + 1$ then $p_0 \geq 1 - \eta$.

<u>Example:</u>



$\Rightarrow$

$$m \gtrsim \frac{r^2}{\|\boldsymbol{c}\|^2} n$$

The Rare Eclipse Problem (alternative)

Restricted Isometry Property: (ℓ_1, ℓ_2)-RIP$(\mathcal{K}, \epsilon)$

$$\forall \boldsymbol{x} \in \mathcal{K},\ (1-\epsilon)\,\|\boldsymbol{x}\| \leqslant \|\boldsymbol{\Phi x}\|_1 \leqslant (1+\epsilon)\,\|\boldsymbol{x}\|$$

[Schechtman, 06] [Plan, Vershynin, 14]

If $\mathcal{S} \subset \mathbb{S}^{n-1}$ and $m \gtrsim \epsilon^{-2} w^2(\mathcal{S})$, then, w.h.p*,

$$(1-\epsilon) \leqslant \sqrt{\tfrac{\pi}{2}}\,\tfrac{1}{m}\|\boldsymbol{\Phi u}\|_1 \leqslant (1+\epsilon)$$

for $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ and $\Phi_{ij} \sim_{\rm iid} \mathcal{N}(0,1)$.

*: *i.e.*, $\mathbb{P} \geqslant 1 - C\exp(-c\epsilon^2 m)$.

The Rare Eclipse Problem (alternative)

Restricted Isometry Property: $(\ell_1, \ell_2)\text{-RIP}(\mathcal{K}, \epsilon)$

$$\forall \boldsymbol{x} \in \mathcal{K},\ (1-\epsilon)\,\|\boldsymbol{x}\| \leqslant \|\boldsymbol{\Phi x}\|_1 \leqslant (1+\epsilon)\,\|\boldsymbol{x}\|$$

[Schechtman, 06] [Plan, Vershynin, 14]

If $\mathcal{S} \subset \mathbb{S}^{n-1}$ and $m \gtrsim \epsilon^{-2} w^2(\mathcal{S})$, then, w.h.p*,

$$(1-\epsilon) \leqslant \sqrt{\tfrac{\pi}{2}}\tfrac{1}{m}\|\boldsymbol{\Phi u}\|_1 \leqslant (1+\epsilon)$$

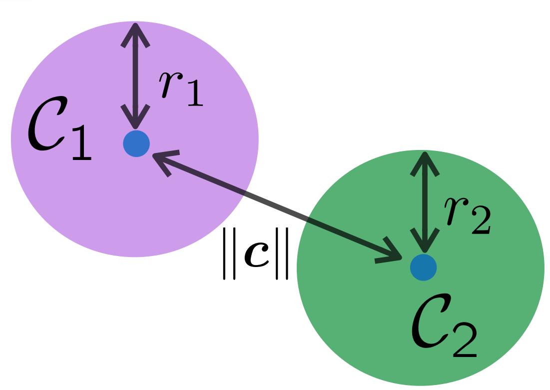
for $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ and $\Phi_{ij} \sim_{\text{iid}} \mathcal{N}(0,1)$.

<u>Therefore:</u>

For $\mathcal{S} = (\mathbb{R}^+\mathcal{C}^-) \cap \mathbb{S}^{n-1}$, if $m \gtrsim \epsilon^{-2} w_\cap^2$, w.h.p*, (P1)

$$(1-\epsilon)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \leqslant \sqrt{\tfrac{\pi}{2}}\,\tfrac{1}{m}\|\boldsymbol{\Phi x}_1 - \boldsymbol{\Phi x}_2\|_1 \leqslant (1+\epsilon)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|,$$

for all $\boldsymbol{x}_1 \in \mathcal{C}_1$ and all $\boldsymbol{x}_2 \in \mathcal{C}_2$.



This result also explains REP!
(but less sharply)

*: *i.e.*, $\mathbb{P} \geqslant 1 - C\exp(-c\epsilon^2 m)$.

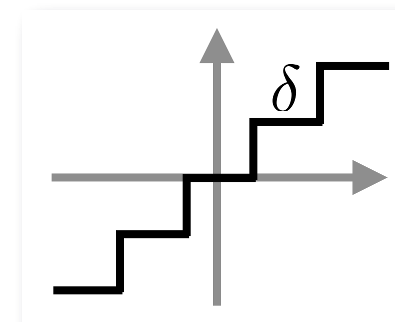
The Quantized Case

This work

Quantized Dithered Random Mapping

$$\begin{cases} \delta > 0 & \text{(a resolution)}, \\ \mathcal{Q}(\cdot) := \delta\lfloor \cdot/\delta \rfloor \quad \text{(applied componentwise)} & \text{(scalar quantizer)}, \\ \boldsymbol{\Phi} \text{ is } (\ell_1, \ell_2)\text{-RIP}(\mathcal{K}, \epsilon) & \text{(a well-behaved } \boldsymbol{\Phi}), \\ \text{and a \textit{dithering} } \boldsymbol{\xi} \in \mathbb{R}^m \text{ with } \xi_j \sim_{\text{iid}} \mathcal{U}([0,\delta]) & \text{(your friend)} \end{cases}$$

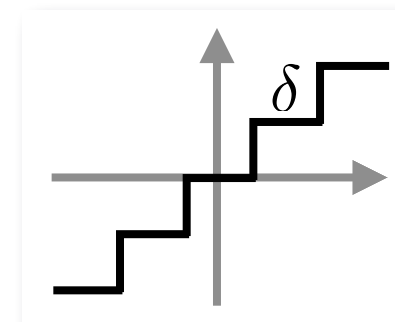
$$\text{QDRM:} \quad \boxed{\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\xi})}$$



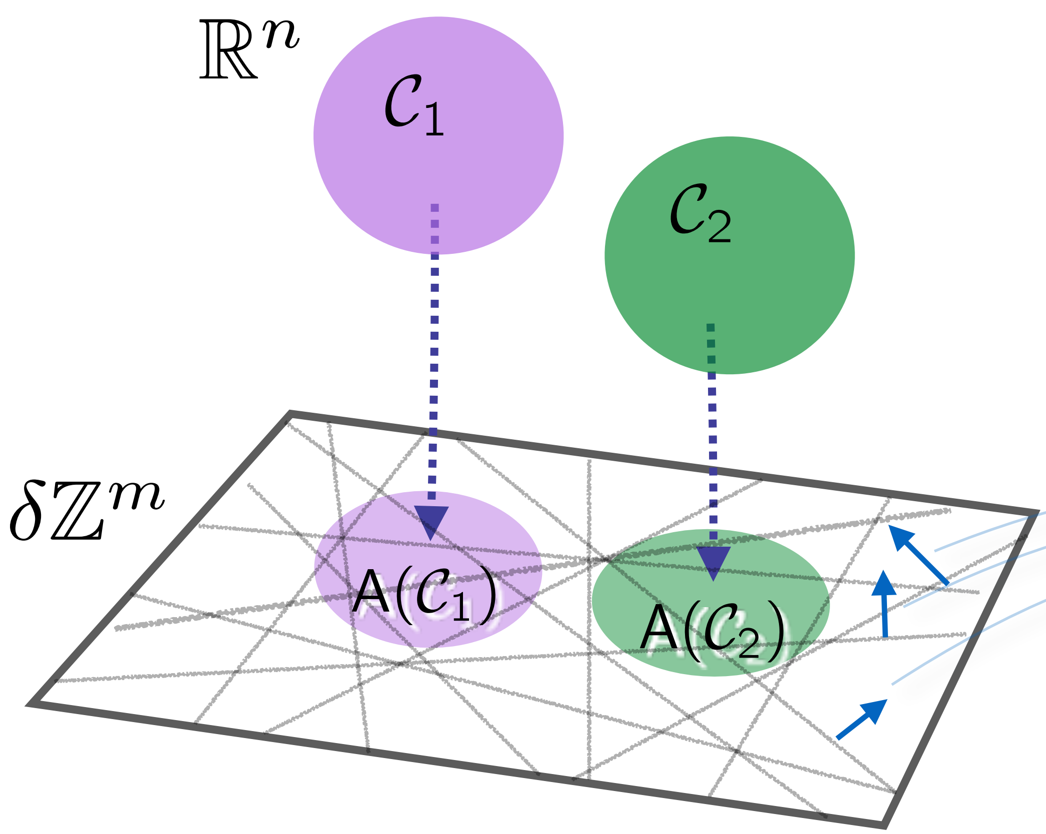
- QDRM = *compact* signatures for storage/transmission/processing.
- *Dithering* makes quantizer "transparent": $\mathbb{E}_{\xi\sim\mathcal{U}([0,1])}\lfloor \lambda + \xi \rfloor = \lambda$
- Results exist on *quasi-isometric embeddings* (LJ '15-'17) and "quantized" RIP (inherited from RIP, LJ, VC '17) with *additive* and *multiplicative* distortions decaying when *m* increases.

Quantized Dithered Random Mapping

$$
\begin{cases}
\delta > 0 & \text{(a resolution)},\\
\mathcal{Q}(\cdot) := \delta \lfloor \cdot / \delta \rfloor \quad \text{(applied componentwise)} & \text{(scalar quantizer)},\\
\boldsymbol{\Phi} \text{ is } (\ell_1, \ell_2)\text{-RIP}(\mathcal{K}, \epsilon) & \text{(a well-behaved } \boldsymbol{\Phi}\text{)},\\
\text{and a } \textit{dithering } \boldsymbol{\xi} \in \mathbb{R}^m \text{ with } \xi_j \sim_{\text{iid}} \mathcal{U}([0, \delta]) & \text{(your friend)}
\end{cases}
$$



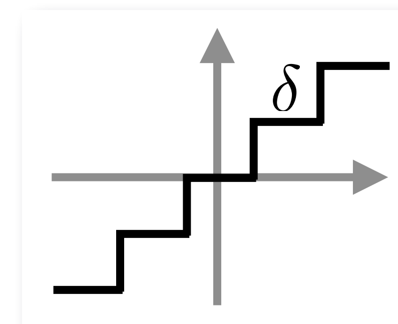
QDRM: $\boxed{\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi x} + \boldsymbol{\xi})}$



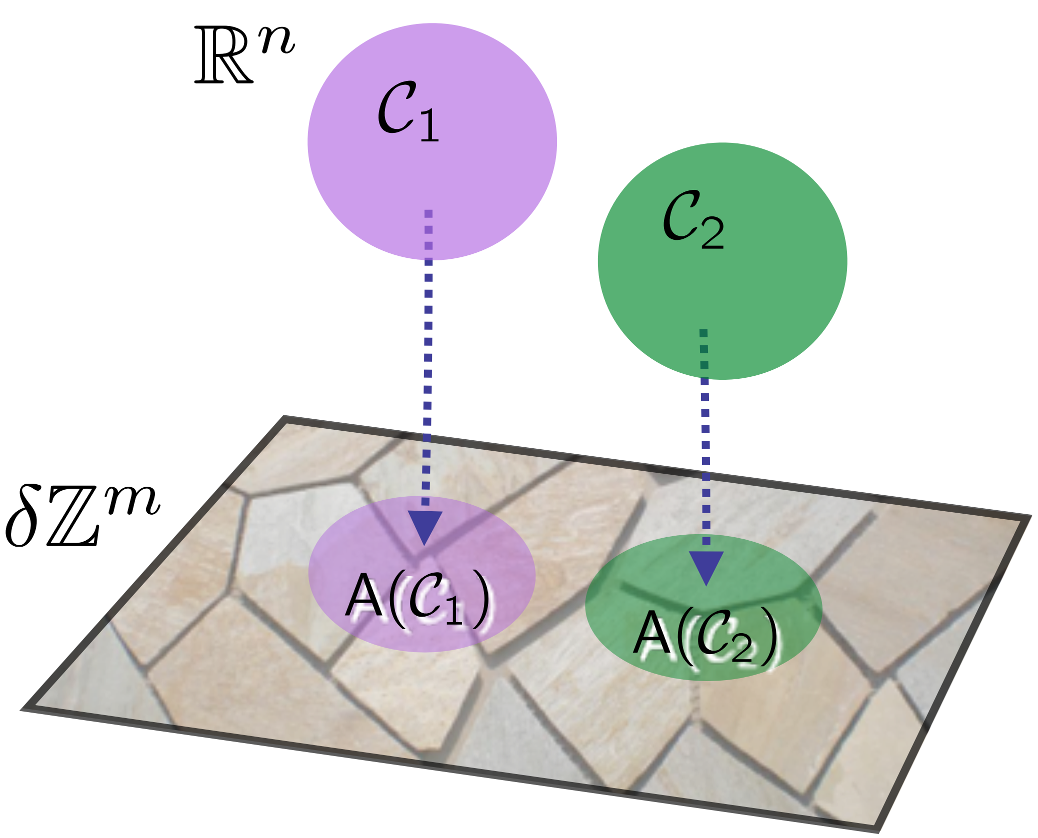
$\mathsf{A}_i(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{\varphi}_i^\top \boldsymbol{x} + \xi_i) \in \delta\mathbb{Z}$

Quantized Dithered Random Mapping

$$\begin{cases} \delta > 0 & \text{(a resolution),} \\ \mathcal{Q}(\cdot) := \delta\lfloor \cdot/\delta \rfloor \ \ \text{(applied componentwise)} & \text{(scalar quantizer),} \\ \boldsymbol{\Phi} \text{ is } (\ell_1, \ell_2)\text{-RIP}(\mathcal{K}, \epsilon) & \text{(a well-behaved } \boldsymbol{\Phi}\text{),} \\ \text{and a } \textit{dithering } \boldsymbol{\xi} \in \mathbb{R}^m \text{ with } \xi_j \sim_{\text{iid}} \mathcal{U}([0, \delta]) & \text{(your friend)} \end{cases}$$



QDRM: $$\boxed{\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\xi})}$$



"Rare Eclipse Problem on Tiles"

$\mathcal{C}_1$, $\mathcal{C}_2$, m and δ such that

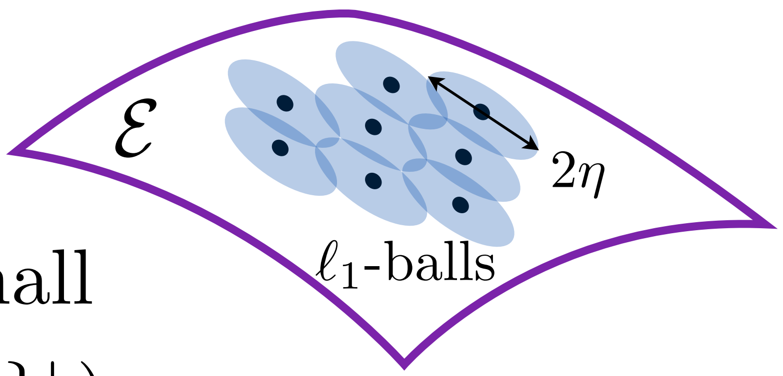
$\mathbb{P}[\mathsf{A}(\mathcal{C}_1) \cap \mathsf{A}(\mathcal{C}_2) = \emptyset] \geqslant 1 - \eta$?

Quantized Random Embeddings

(i.e., only thanks to the dithering!)

Embedding $\mathcal{E} \subset \mathbb{R}^m$ into $\delta\mathbb{Z}^m$ with

$$\mathsf{A}'(\boldsymbol{y}) := \underset{\delta}{\underset{\uparrow}{\mathcal{Q}}}(\boldsymbol{y} + \underset{\delta}{\underset{\uparrow}{\boldsymbol{\xi}}})$$



If $\mathcal{E}$ has small ℓ_1-Kolmogorov entropy, *i.e.*, small

$$\mathcal{H}_1(\mathcal{E}, \eta) = \log(\min_{\mathcal{S}} |\{\mathcal{S} \subset \mathcal{E} : \mathcal{S} + \eta\mathbb{B}_1^n \supset \mathcal{E}\}|)$$

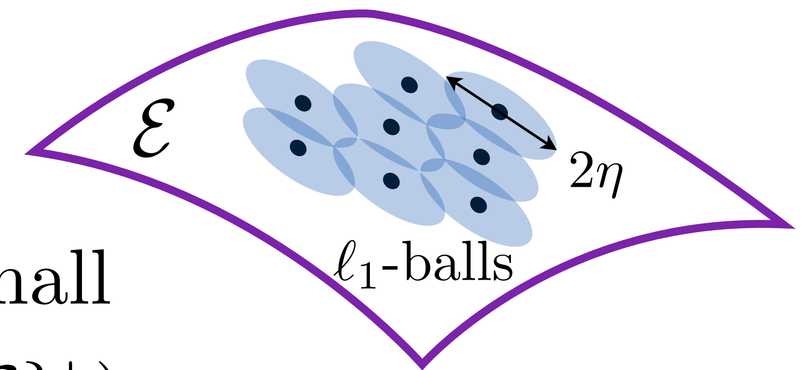
*: [LJ, Cambareri, '16] [Cambareri, Xu, LJ, '17]

Quantized Random Embeddings

(i.e., only thanks to the dithering!)

Embedding $\mathcal{E} \subset \mathbb{R}^m$ into $\delta\mathbb{Z}^m$ with

$$\mathsf{A}'(\boldsymbol{y}) := \underset{\delta}{\mathcal{Q}}(\boldsymbol{y} + \underset{\delta}{\boldsymbol{\xi}})$$



If $\mathcal{E}$ has small ℓ_1-Kolmogorov entropy, *i.e.*, small

$$\mathcal{H}_1(\mathcal{E}, \eta) = \log(\,\min_{\mathcal{S}} |\{\mathcal{S} \subset \mathcal{E} : \mathcal{S} + \eta\mathbb{B}_1^n \supset \mathcal{E}\}|\,)$$

Given $\epsilon > 0$, if $m \gtrsim \epsilon^{-2}\mathcal{H}_1(\mathcal{E}, \frac{m\delta\epsilon^2}{1+\epsilon})$, then, w.h.p*, for all $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{E}$ and some $c > 0$, (P2)

$$\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_1 - c\delta\epsilon \leqslant \tfrac{1}{m}\|\mathsf{A}'(\boldsymbol{y}_1) - \mathsf{A}'(\boldsymbol{y}_2)\|_1 \leqslant \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_1 + c\delta\epsilon.$$

Remarks: ∘ For $\mathcal{E} = \boldsymbol{\Phi}\mathcal{K}$ and $\boldsymbol{\Phi}$ an (ℓ_1, ℓ_2)-RIP$(\mathcal{K} - \mathcal{K}, \epsilon' < 1)$,

$$\mathcal{H}_1(\mathcal{E}, 2m\eta) \leqslant \mathcal{H}_2(\mathcal{K}, \eta) \quad \text{(P3)}$$

∘ $\mathcal{H}_2$ is bounded for sets (cones, convex spaces, ...)

*: [LJ, Cambareri, '16] [Cambareri, Xu, LJ, '17]

"Rare Eclipse Problem on Tiles"

Combining (P1), (P2) and (P3) (with $\mathsf{A} = \mathsf{A}' \circ \boldsymbol{\Phi}$) gives:

Given $\sigma := \min_{\boldsymbol{z} \in \mathcal{C}_-} \|\boldsymbol{z}\|$ and $w_\cap = w((\mathbb{R}_+ \mathcal{C}_-) \cap \mathbb{S}^{n-1})$.

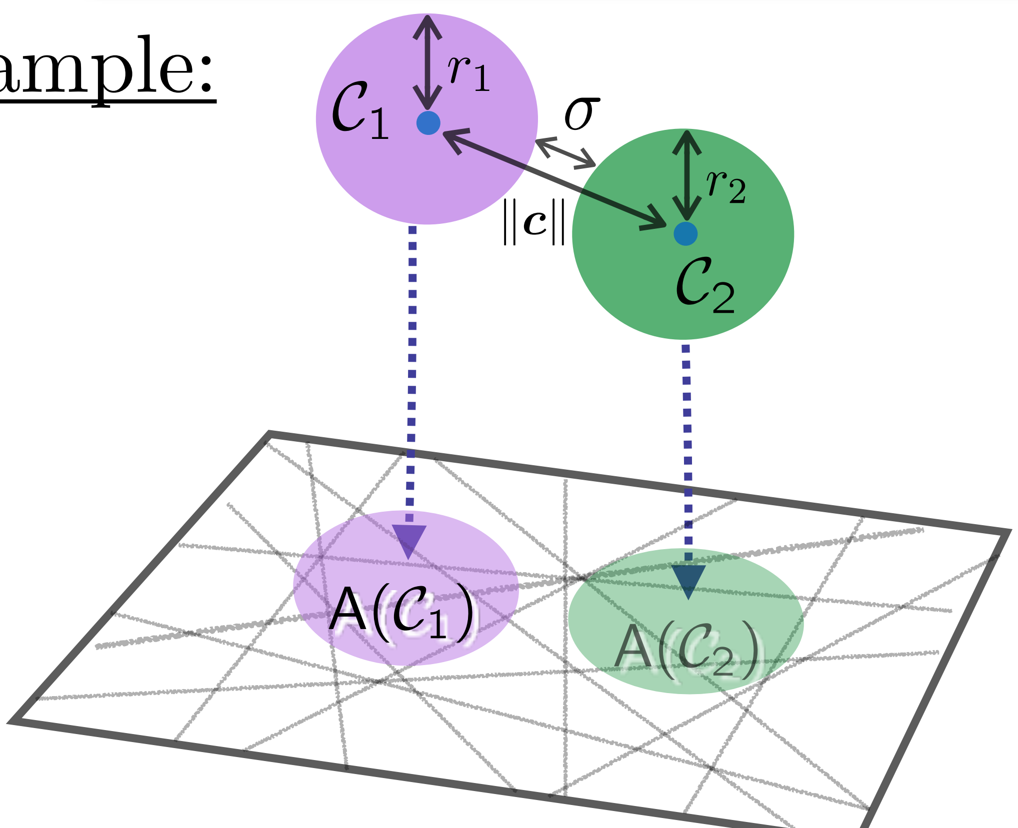
Provided

$$m \gtrsim \Big(\underbrace{w_\cap^2}_{\text{linear}} + \underbrace{n \tfrac{\delta^2}{\sigma^2}}_{\text{quantiz.}}\Big)\ \big(1 + \underbrace{\log\big(1 + \tfrac{rm}{\delta n}\big)}_{\text{proof artifact?}} + \underbrace{w_\cap^{-2} \log \tfrac{1}{\eta}}_{\text{linear}}\big),$$

we have

$$\mathbb{P}[\mathsf{A}(\mathcal{C}_1) \cap \mathsf{A}(\mathcal{C}_2) = \emptyset] \geqslant 1 - \eta.$$

<u>Example:</u>



$$\Rightarrow \quad m \gtrsim \Big(\frac{r^2}{\|\boldsymbol{c}\|^2} + \frac{\delta^2}{(\|\boldsymbol{c}\| - r)^2}\Big)\, n$$

Note: $\delta > \sigma$ is allowed (dithering effect!)

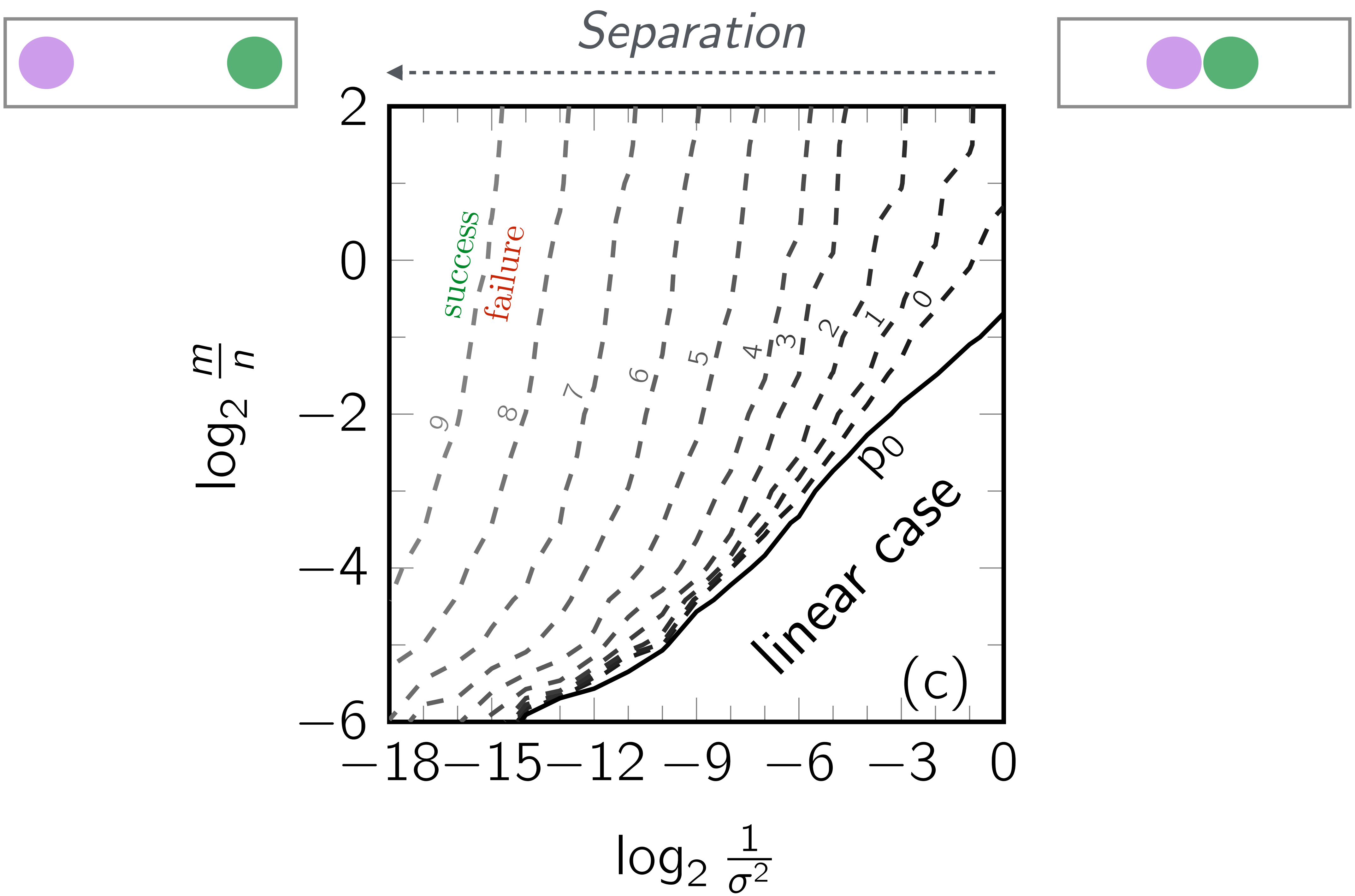
Note bis: $m > n$ not specially bad ($\delta\mathbb{Z}^m$).

Simulations: Phase transition study

Empirical evaluation of a crude lower bound (with no dithering)

128 trials, randomly drawn, $n = 64$, fixed $r = 1$, $\delta = \{1, 2, \ldots, 512\}$.

Represented: Phase transition level-curves (at 0.9).

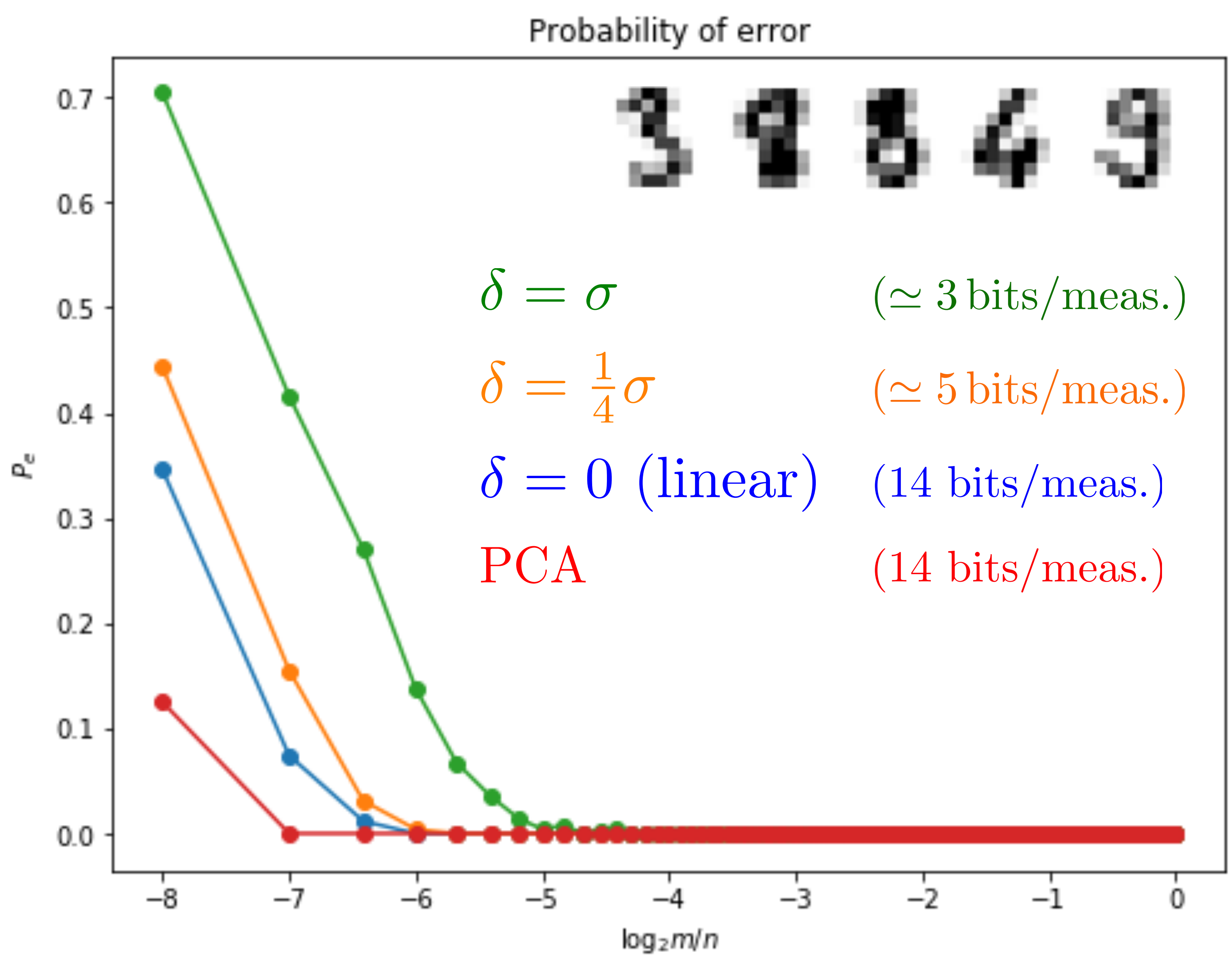


Simulations: Digit dataset (from scikit learn)

10 handwritten digits, 8x8 pixels (n=64), samples/class $\simeq$ 12.

Training/Test sets = 50%/50%. $\sigma = \min_{i,j:i\neq j} \min_{\boldsymbol{u}\in\mathcal{C}_i,\boldsymbol{v}\in\mathcal{C}_j} \|\boldsymbol{u}-\boldsymbol{v}\|$

Classification: 5-NN Classifier.



Try some code out here: **github.com/VC86/MLSPbox**

Conclusion

- **Take-away messages:**
  - QDRMs: non-adaptive dimensionality reduction,
    preserve geometry of datasets
  - Extension of BMR '14 to QDRMs,
    with sample complexity loss in $\frac{\delta^2}{\sigma^2} n$ (quantiz. impact)
- **Future work:**
  - Better bound and testable conditions for empirical tests.
  - Extension of result to $\text{RIP}_{2,2}$ matrices, "fast" random matrices
  - Extension of framework to other non-linear maps (ReLU?)
    $\rightarrow$ applicability to D/CNN with random weights [Giryes et al. 15]?

Thank you for your attention.

arXiv:1702.04664

Try some code out here: github.com/VC86/MLSPbox

For any question/suggestion, contact us at:
valerio.cambareri@uclouvain.be
chunlei.xu@uclouvain.be
laurent.jacques@uclouvain.be

FREEDOM TO RESEARCH

Université catholique de Louvain