

Learning to Reconstruct Signals From Binary Measurements



Julián Tachella
CNRS, Physics Laboratory
ENS Lyon, France



Laurent Jacques
INMA/ICTEAM
UCLouvain, Belgium

May 21, 2024

Inverse problems (IP)

Measurements in \mathbb{R}^m

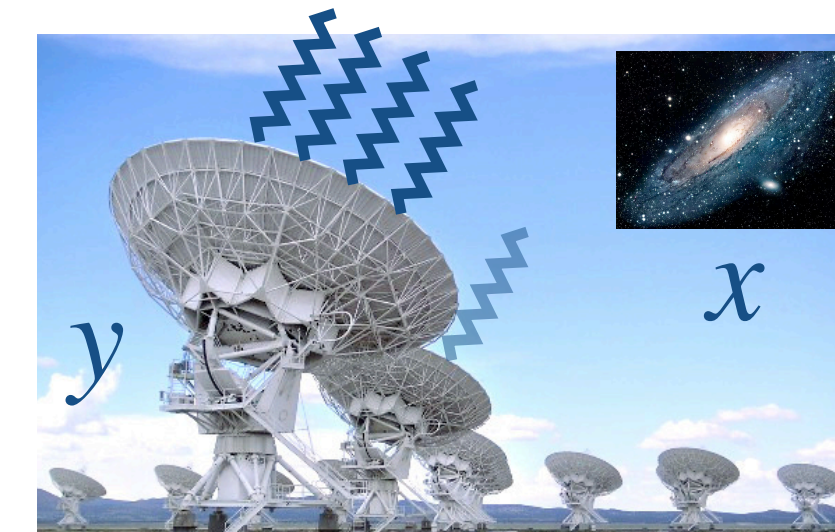
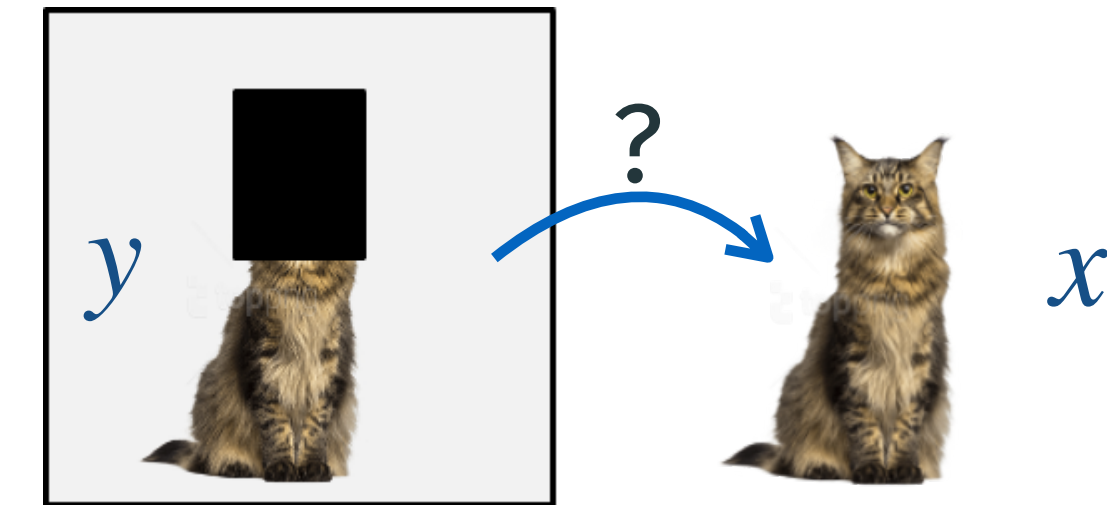
Measurements operator

$$y = A(x) + \varepsilon$$

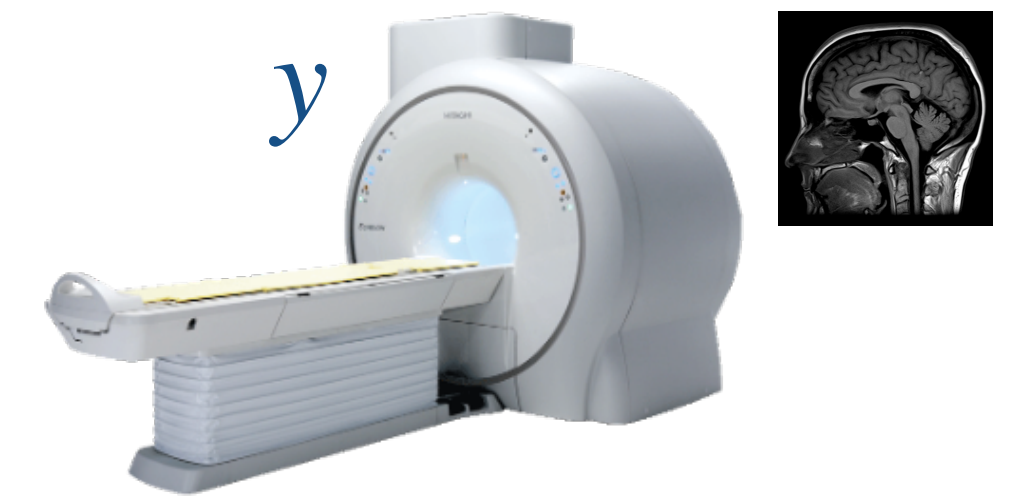
Signal in \mathbb{R}^n

noise in \mathbb{R}^m

Image inpainting



Radio Astronomy



MRI

Ill-posed:



many x consistent with y (e.g., if $m < n$)

Solution:

Restrict to a set of plausible signals $X \ni x$

Size? $\ll \mathbb{R}^n$?

Recommender system

Fruit	User 1	User 2	User 3
	★ ★	?	★
	★	★	?
	★	★ ★	★

Solving IP: regularised reconstruction

Idea: use a *prior* \equiv loss $\rho(x)$ to promote plausible reconstructions

$$\hat{x} \in \arg \min_x \rho(x) \text{ subject to } y \approx A(x)$$

Examples: wavelet/dictionary sparsity, total-variation, ...

Disadvantages:

- ▶ Loose description of true signal distribution $p(x)$
- ▶ Hard to define a good $\rho(x)$ in real world problems

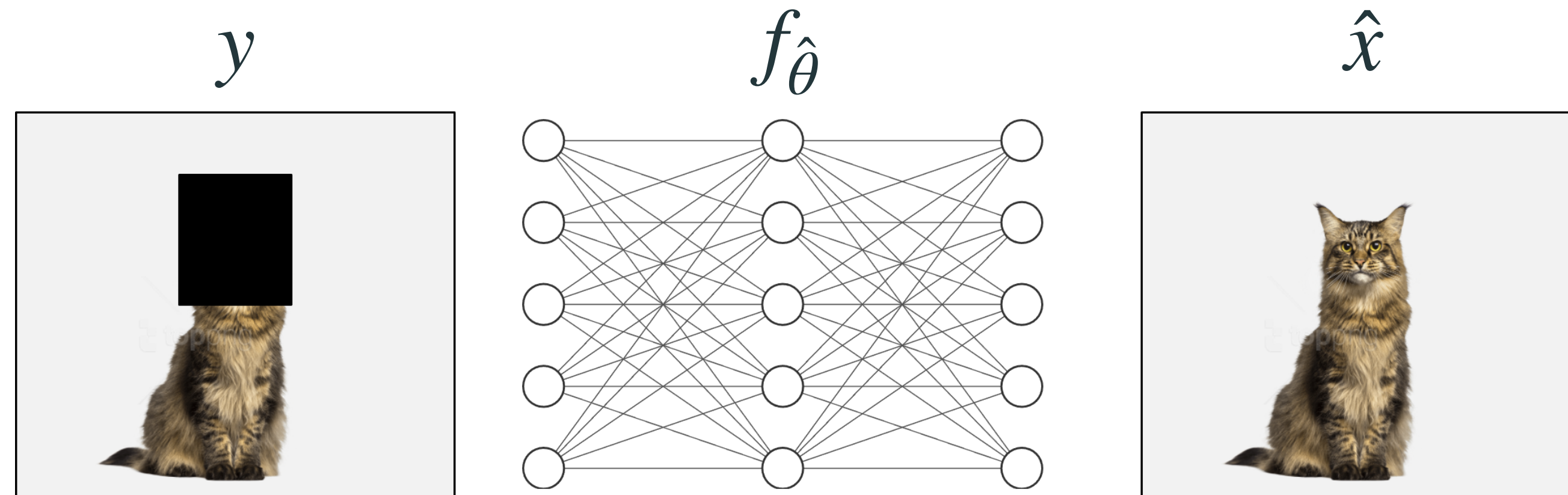


Btw, cat friendly presentation!

Solving IP: learning approach

Idea:

- ▶ use training pairs of **signals** and **measurements** $\{(x_i, y_i)\}_{i=1}^N$
- ▶ learn a parametric inversion function $y \rightarrow \hat{x} = f_{\hat{\theta}}(x)$



$$\hat{\theta} \in \arg \min_{\theta} \sum_{i=1}^N \|x_i - f_{\theta}(y_i)\|^2$$

where $f_{\theta} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is parameterized as a deep neural network.

Implicitly learn both the prior & distribution, and the reconstruction

Solving IP: learning approach

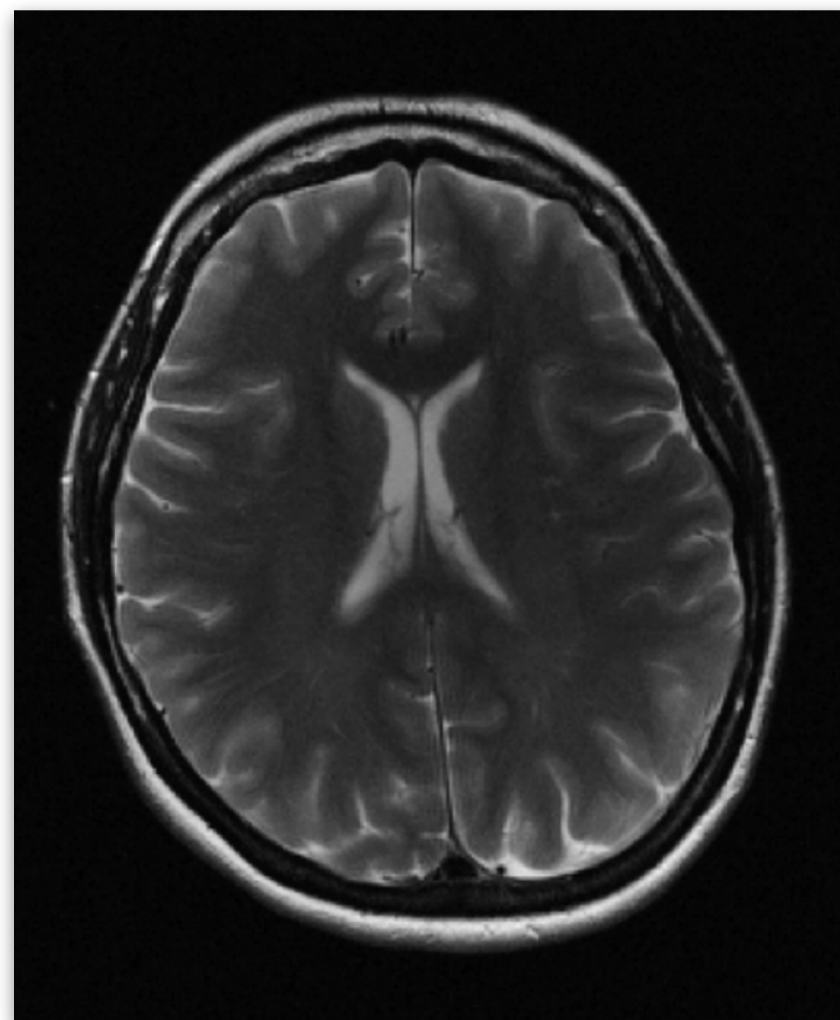
Advantages:

- ▶ State-of-the-art reconstructions
- ▶ “Once trained”, $f_{\hat{\theta}}$ is easy/fast to evaluate

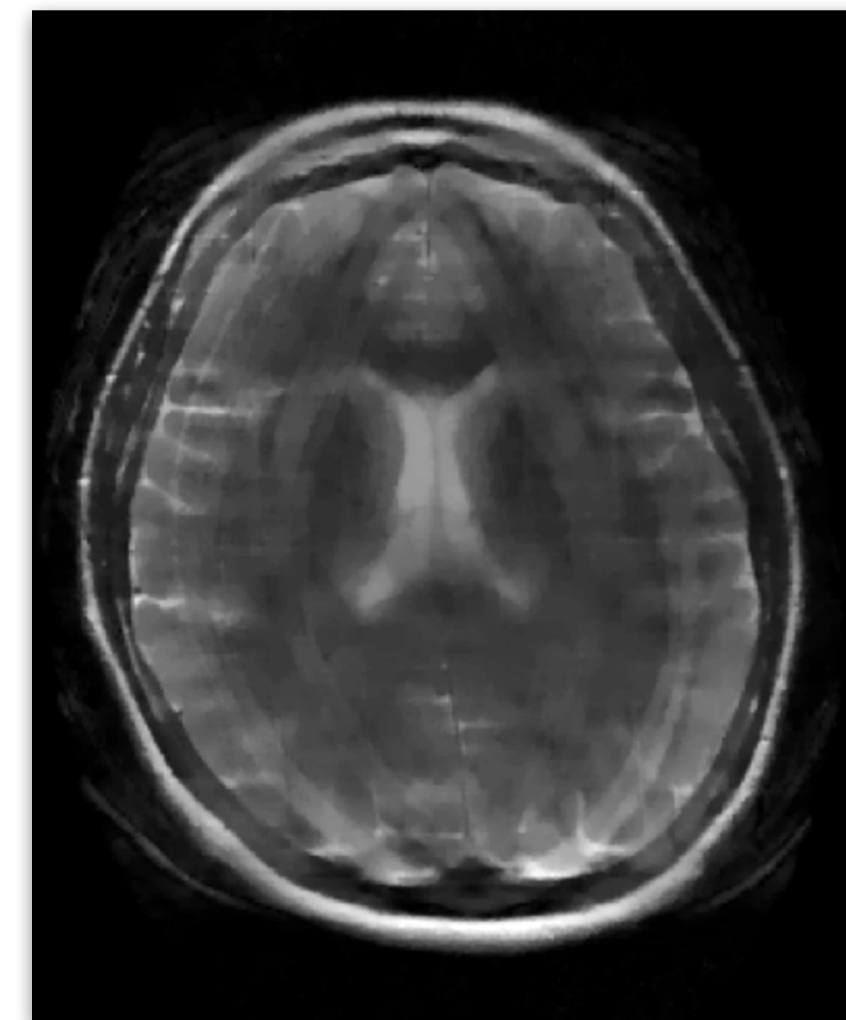
Example:

fastMRI

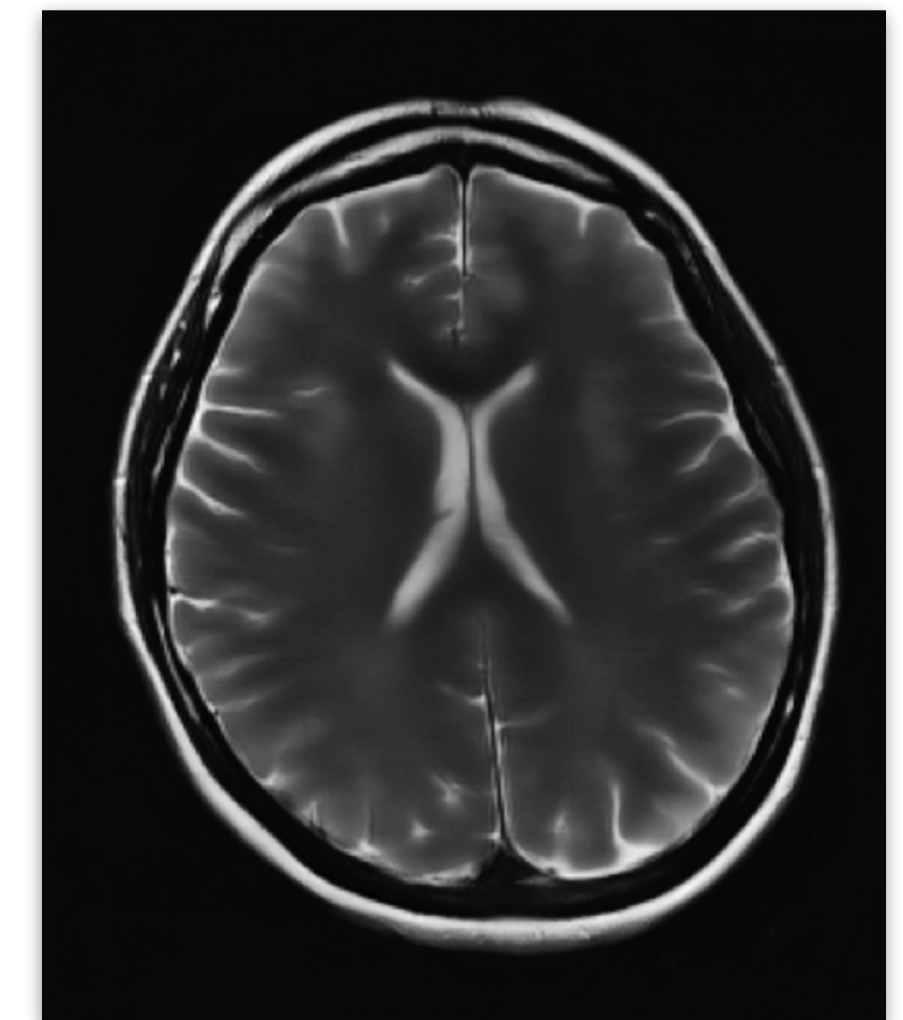
Accelerating MR Imaging with AI



Ground-truth



Total variation
(28.2 dB)



Deep network
(28.2 dB)

→ × 8 accelerated MRI [Zbontar et al., 2019]

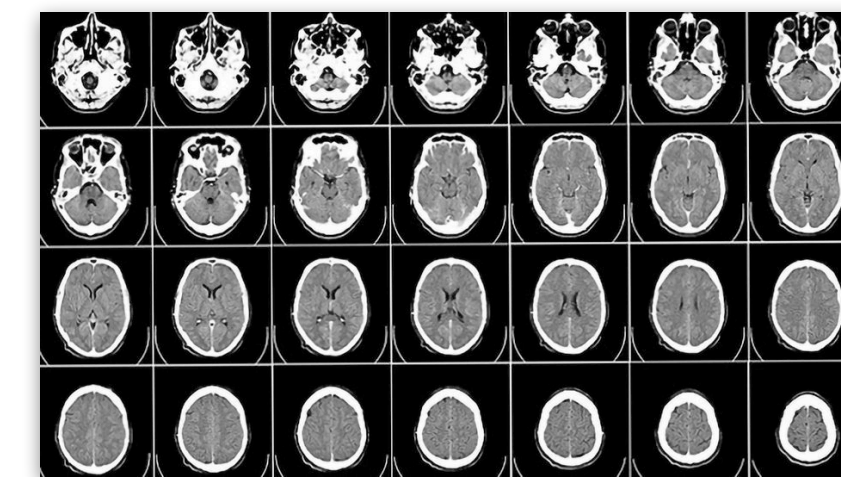
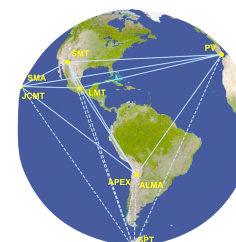
Solving IP: learning approach

Main disadvantage:

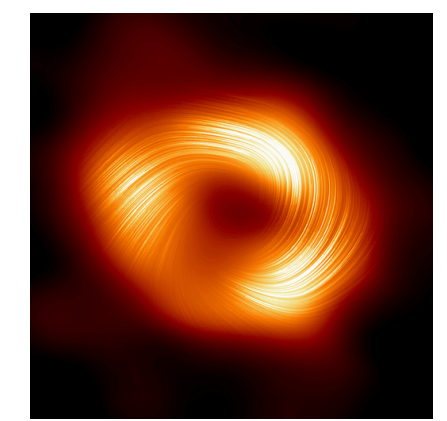
Obtaining training signals $\{x_i\}_{i=1}^N$ can be expensive/impossible.

For instance:

- ▶ Biomedical sciences (e.g., CT, MRI)
- ▶ Astronomical imaging (e.g., EHT)



EHT '22

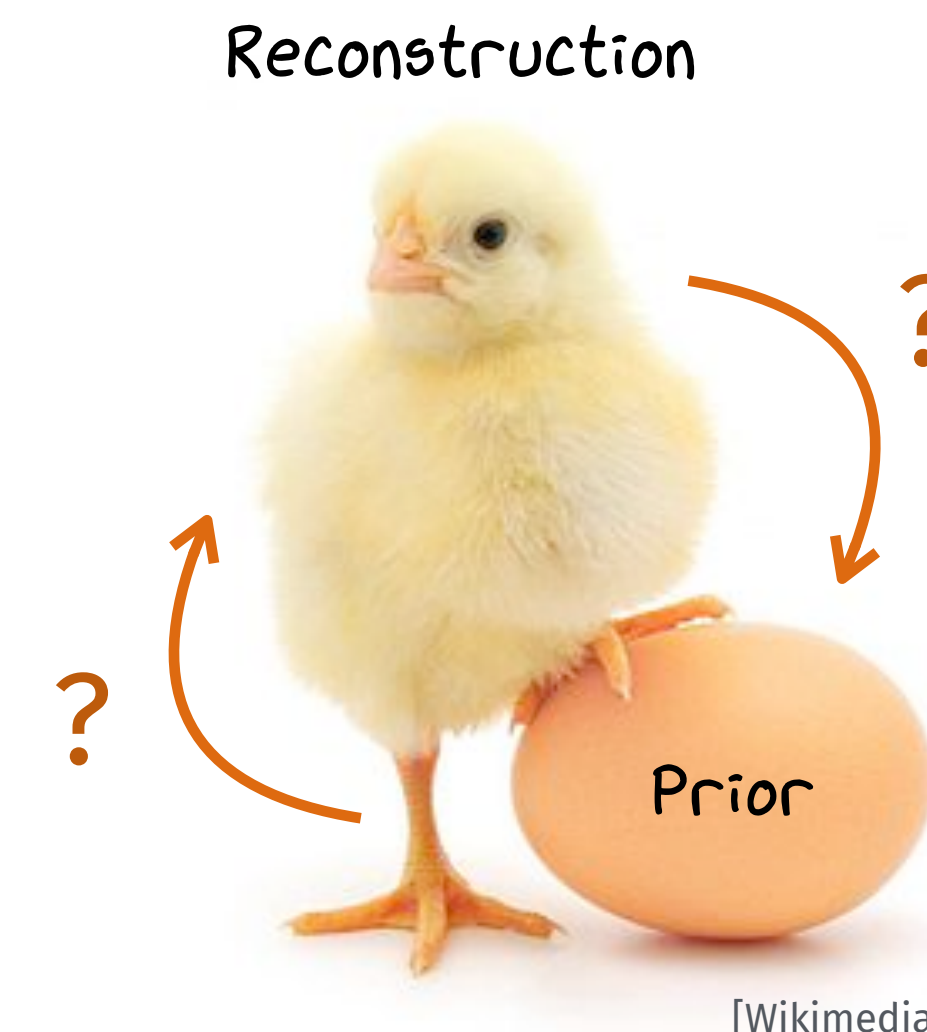


EHT '24

Consequence:

- ▶ Risk to solve expected solution (off-distribution problem)

Prior or reconstruction, which comes first?



[Wikimedia]

Measurement-Driven Computational Imaging

Unsupervised context:

Can we learn to reconstruct signals
from measurement data alone $\{y_i\}_{i=1}^N$?

Linear inverse problems: $y = A(x) + \epsilon \rightarrow \text{Yes}$

If signal set \mathcal{X} is **low-dimensional**, and, either **multiple operators** $\{A_i\}_{i=1}^G$,
or \mathcal{X} invariant to groups of transformations.

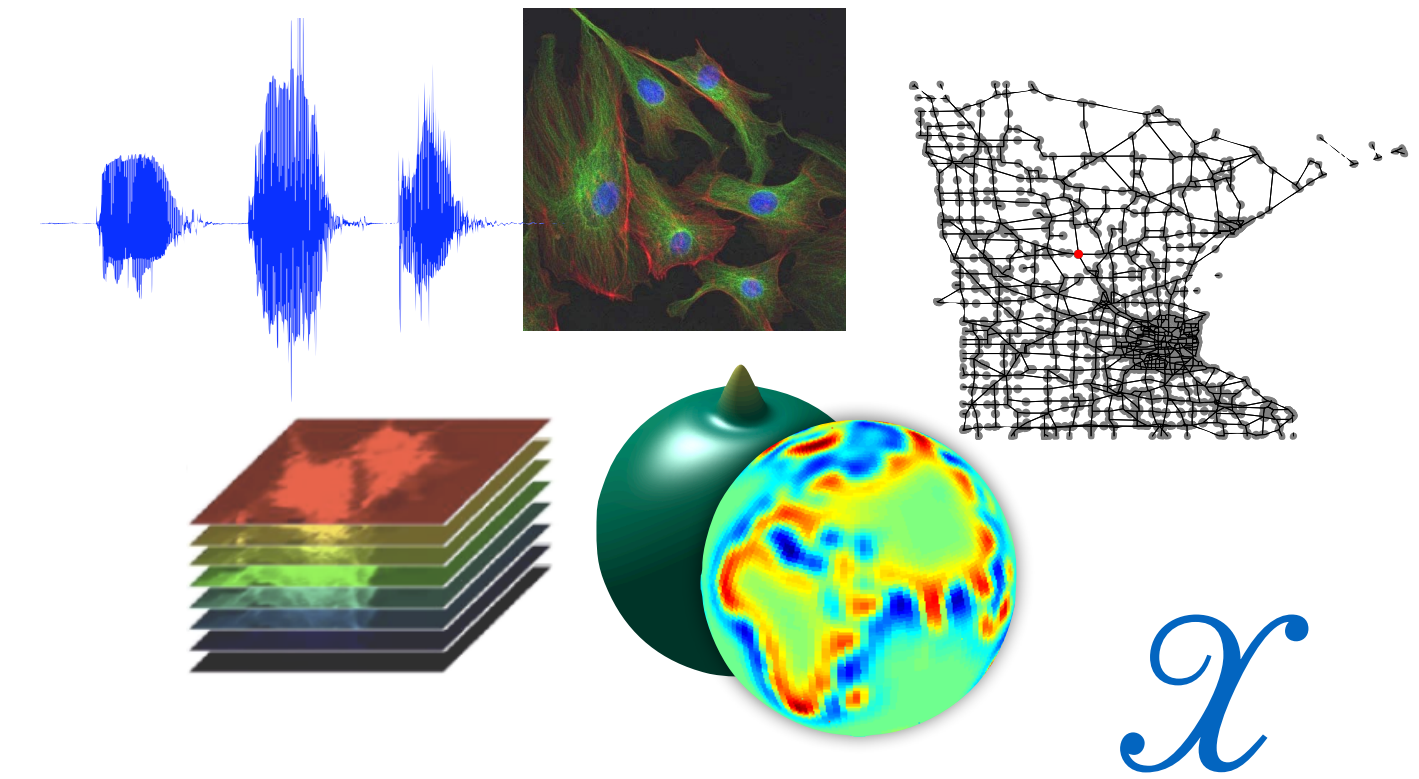
- ▶ Theory [T., Chen and Davies, JMLR, 2023]
- ▶ Algorithms [Chen, T., Davies, CVPR, ICCV, NeurIPS, 2022]

Non-linear inverse problems: $y = f \circ A(x) + \epsilon \rightarrow \text{Today}$
(with $f = \text{sign}$, binary measurements)

Why binary measurements?

Sensing model Given $A = (a_1, \dots, a_m)^\top, a_i \in \mathbb{R}^n$,
we observe a “signal” $x \in \mathcal{X} \subset \mathbb{R}^n$ with m binary measurements :

$$y_i = \text{sign}(a_i^\top x) \in \{\pm 1\} \quad \rightarrow \quad y = \text{sign}(Ax) \in \{\pm 1\}^m$$



Contexts

- ▶ Binary compressive sensing (1-bit CS) \rightarrow Can we estimate x from y ?
- ▶ Binary/quantized dimensionality reduction \rightarrow Do $\text{sign}(A\mathcal{X})$ capture the geometry of \mathcal{X} ?
- ▶ Machine learning \rightarrow Can we classify two signals from their binary measurements?

Interests

- ▶ Compression at acquisition, for signals or datasets
- ▶ Interesting questions related to information theory and high-dimensional statistics

Purpose of this talk

Learning to reconstruct from binary measurements?

Theoretical analysis: given N binary observations & G operators

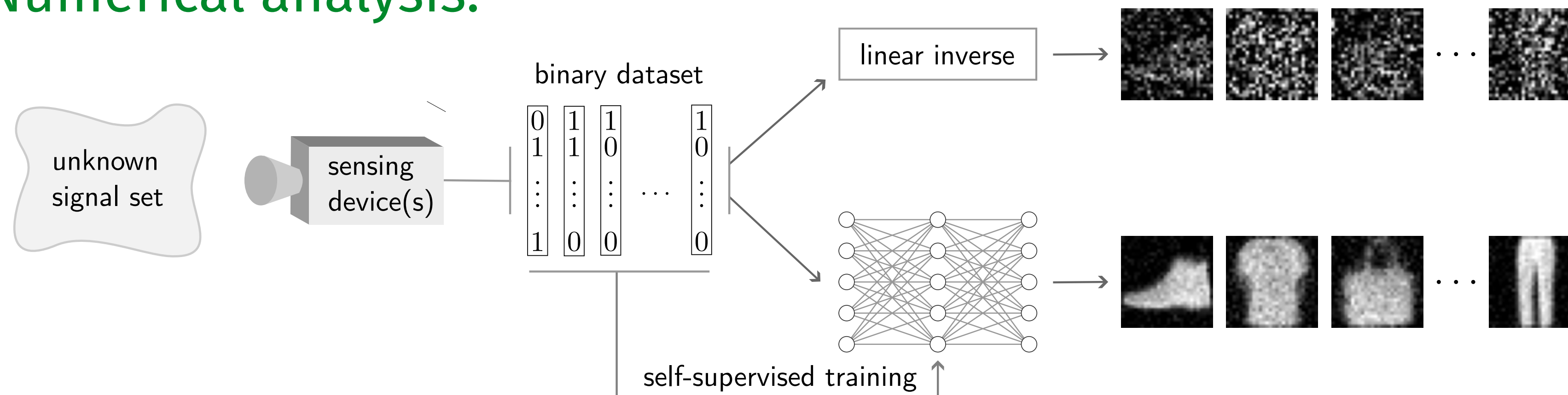
$$y_i = \text{sign}(A_{g_i} x_i), \text{ with } 1 \leq i \leq N \text{ and } g_i \in \{1, \dots, G\}$$

\downarrow \downarrow
known unknown

Estimate of $\mathcal{X} \supset \{x_i\}_{i=1}^N$? Estimation error?

Condition on m & G ? On the sample complexity N ?

Numerical analysis:

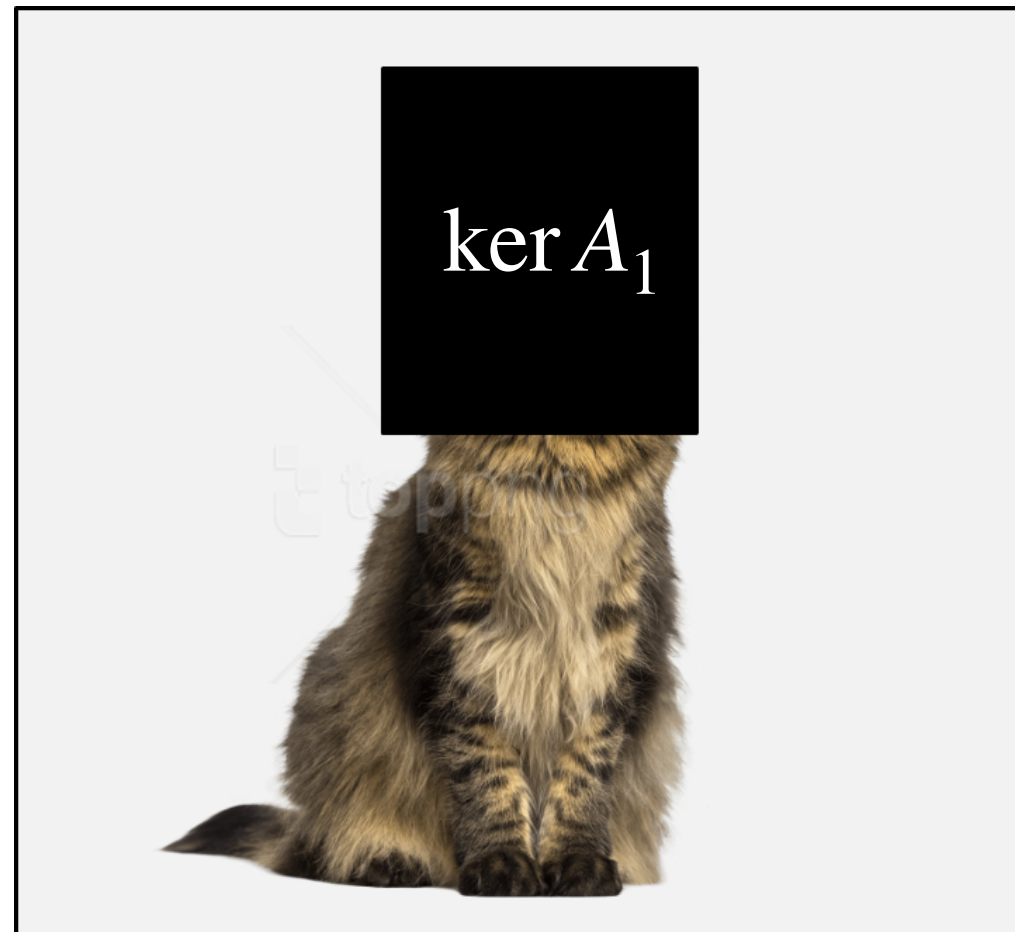


Cost function? Network architecture? Comparison to linear case?

Sensing scenario 1: Multiple Operators

Measurements might be associated to $G \geq 1$ forward operators

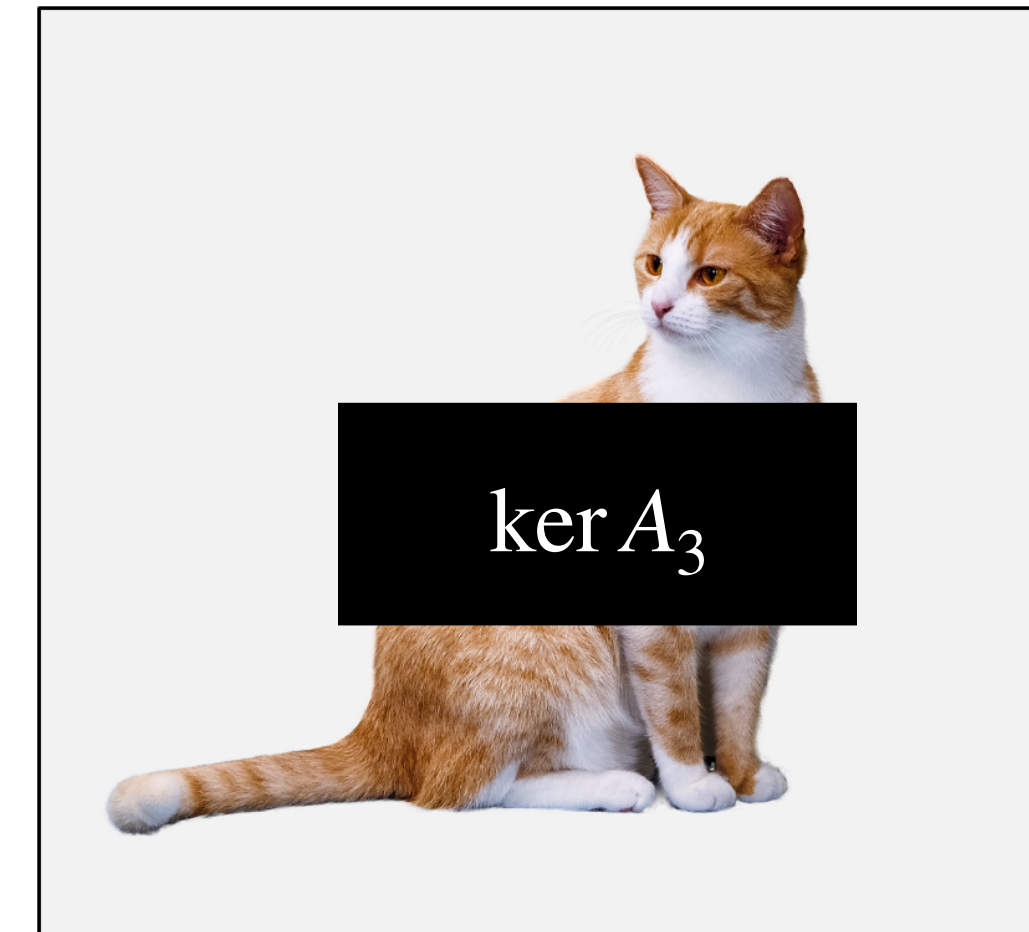
$$A_1 x_1$$



$$A_2 x_2$$



$$A_3 x_3$$



Examples:

- ▶ different access ratings for recommendation systems with distinct users
- ▶ dynamic sensors: $\{A_t : t = n\Delta_T\}$, multi-coil MRI, radio-astronomy ...

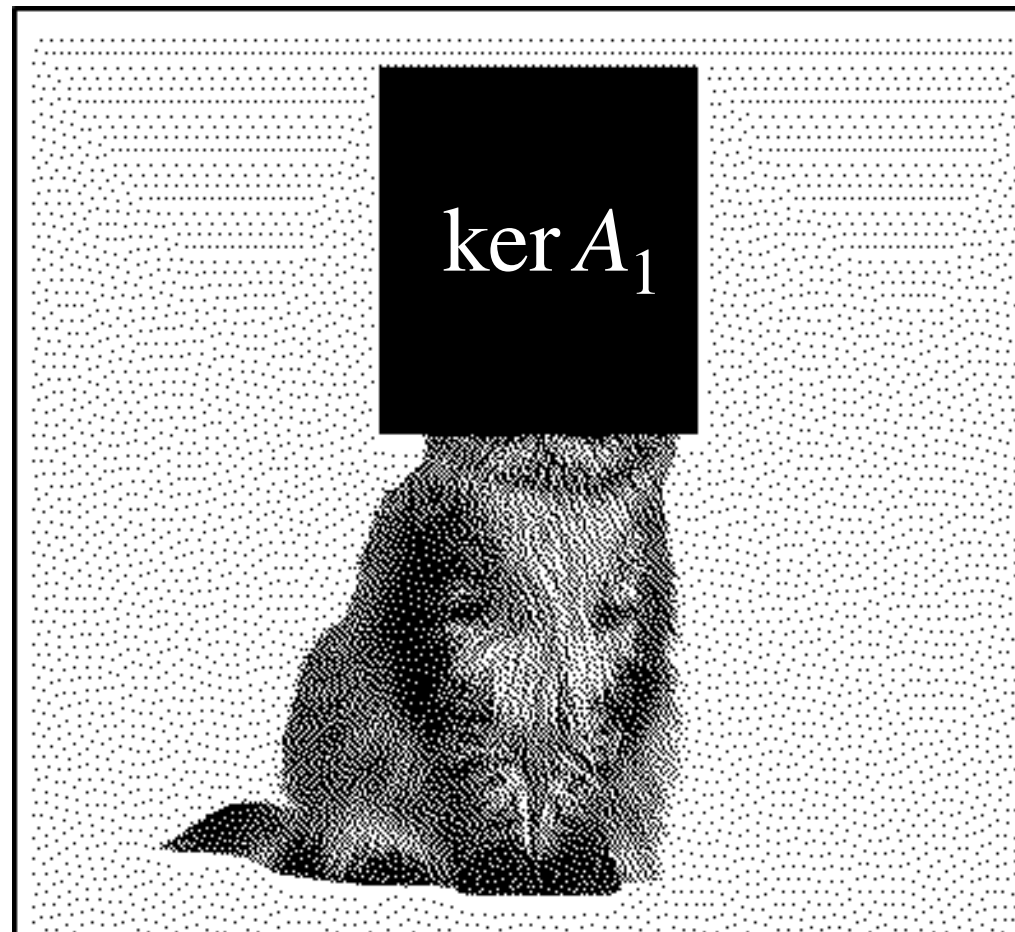
Principle (linear case):

Learning \mathcal{X} is possible if operators don't have the same kernel

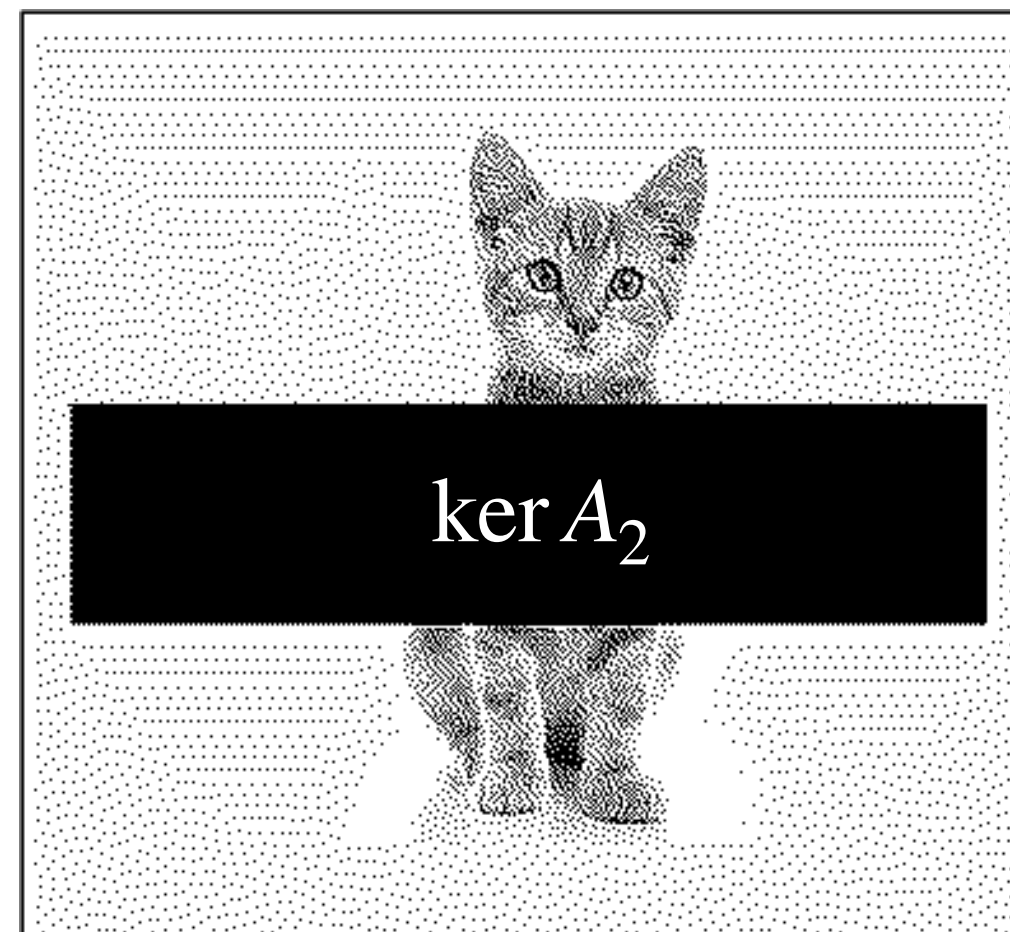
Sensing scenario 1: Multiple Operators

Measurements might be associated to $G \geq 1$ forward operators

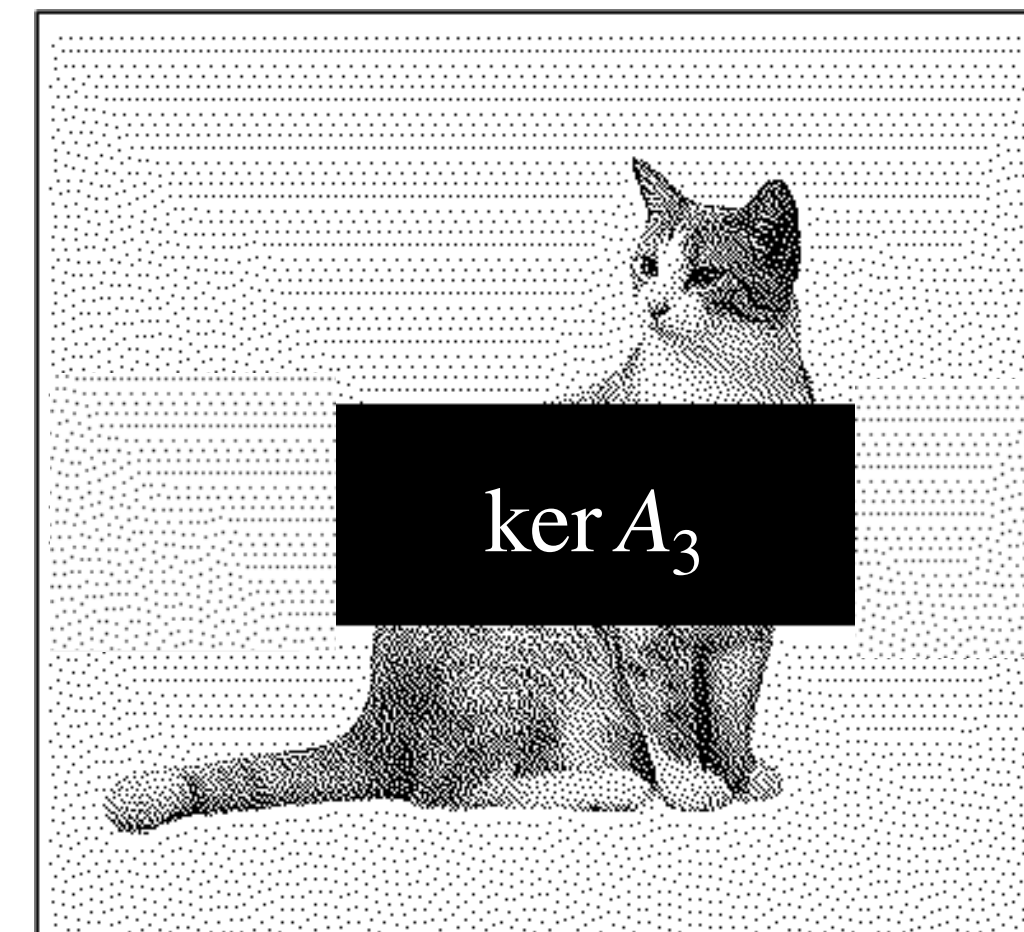
$$\text{sign}(A_1 x_1)$$



$$\text{sign}(A_2 x_2)$$



$$\text{sign}(A_3 x_3)$$



Examples:

- ▶ different access ratings for recommendation systems with distinct users
- ▶ dynamic sensors: $\{A_t : t = n\Delta_T\}$, multi-coil MRI, radio-astronomy ...

Principle (binary case):

Learning \mathcal{X} is possible if operators don't have the same kernel

Sensing scenario 2: Single operator & invariance

Most signals sets are invariant to groups of transformations:

$$\forall x \in \mathcal{X}, \quad \forall g \in \{1, \dots, G\}, \quad x' = T_g^{-1}x \in \mathcal{X}$$

(geometric prior)

For all $g \in \{1, \dots, G\}$ we have

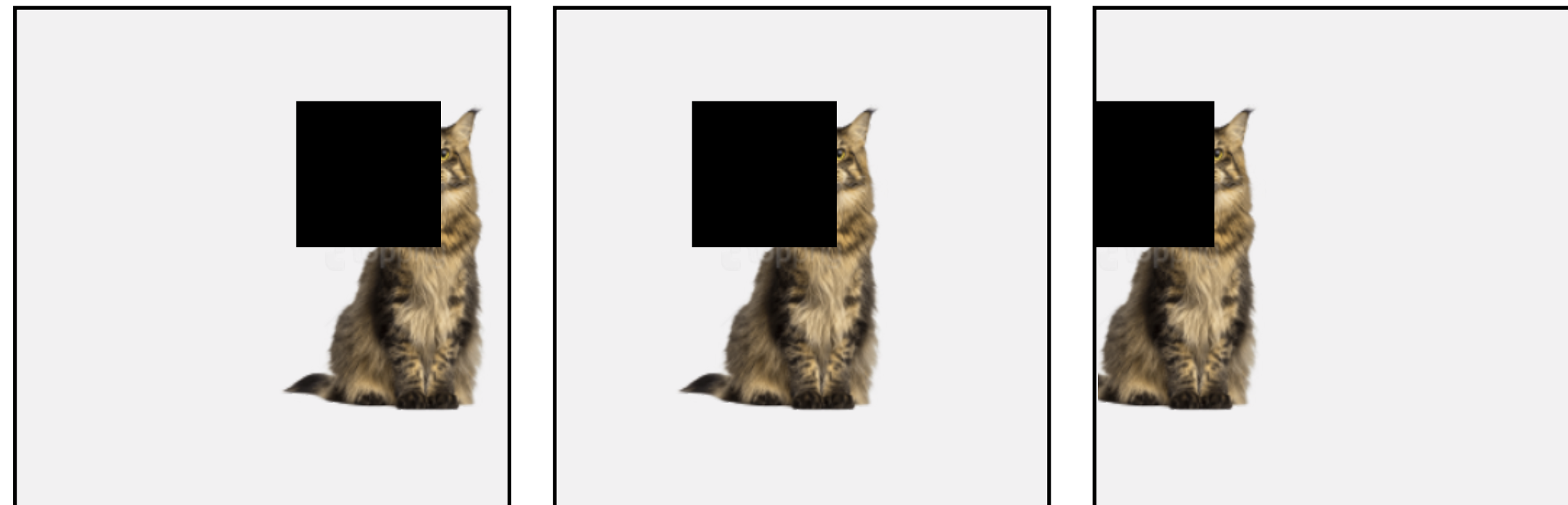
(linear case)

$$y = Ax = \underbrace{AT_g}_{A_g} \underbrace{T_g^{-1}x}_{x'} = A_g x'$$

AT_g for different g

Implicit access to
multiple operators

$$A_g = AT_g$$



Necessary condition

A is not equivariant:

$$AT_g \neq \tilde{T}_g A$$

for some \tilde{T}_g .

Sensing scenario 2: Single operator & invariance

Most signals sets are invariant to groups of transformations:

$$\forall x \in \mathcal{X}, \quad \forall g \in \{1, \dots, G\}, \quad x' = T_g^{-1}x \in \mathcal{X}$$

(geometric prior)

For all $g \in \{1, \dots, G\}$ we have

(binary case)

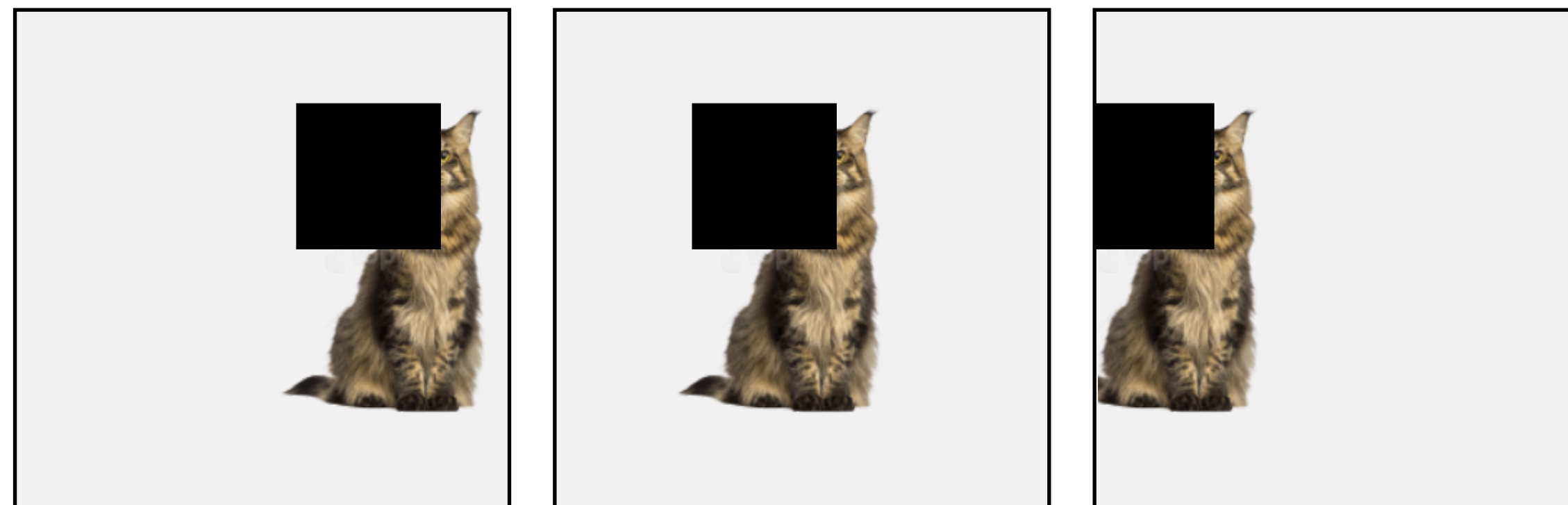
$$y = \text{sign}(Ax) = \text{sign}(AT_g T_g^{-1}x) = \text{sign}(A_g x')$$

\downarrow \downarrow
 A_g x'

AT_g for different g

Implicit access to
multiple operators

$$A_g = AT_g$$



Necessary condition

A is not equivariant:

$$AT_g \neq \tilde{T}_g A$$

for some \tilde{T}_g .

Model identification: the problem

Assumption: enough points of \mathcal{X} have been observed for all operators.
 ∞ for now (More on this later)

Question: Given the observed sets

$$\{\mathcal{Y}_g := \text{sign}(A_g \mathcal{X})\}_{g=1}^G$$

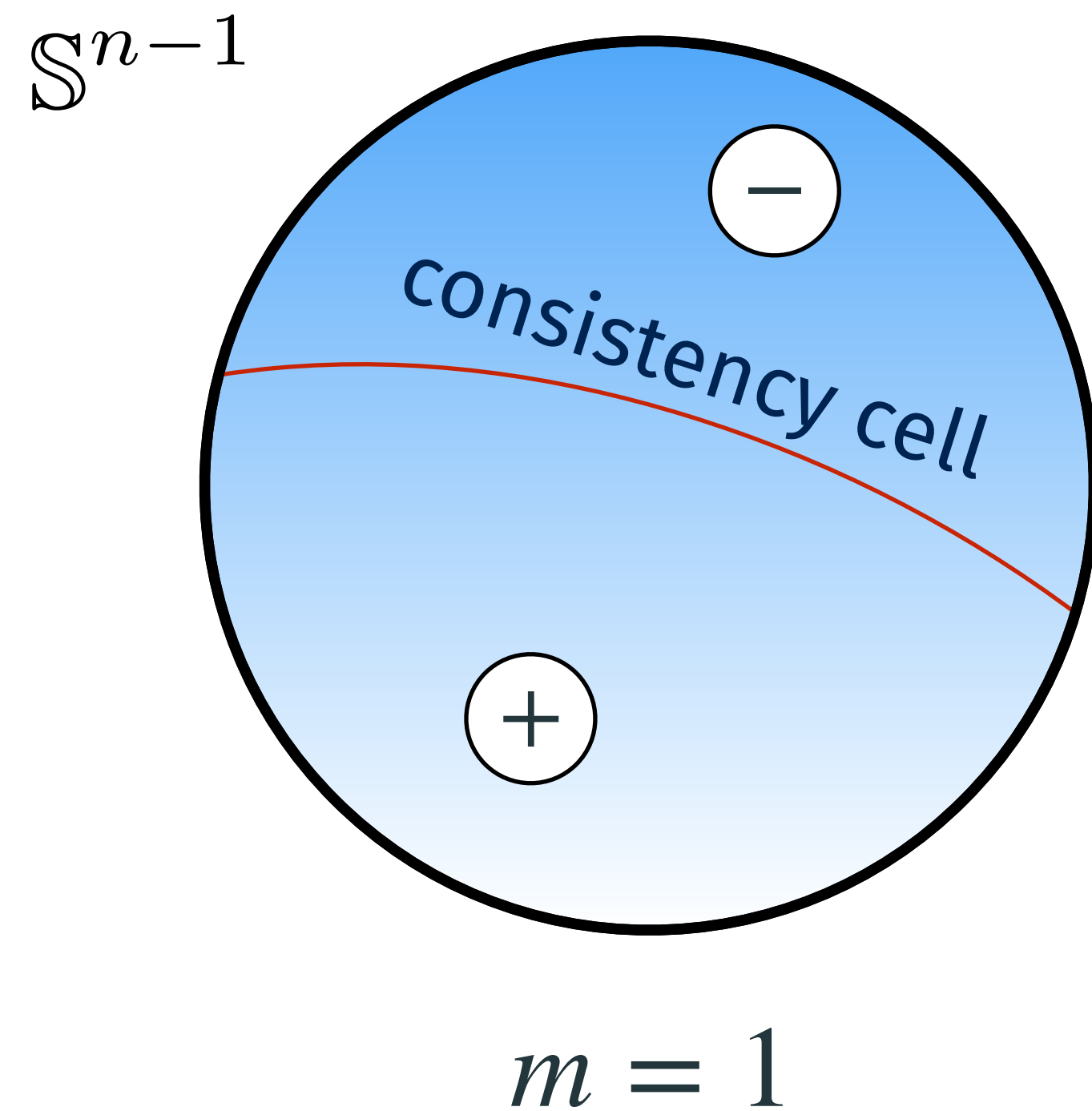
What's the best approximation $\hat{\mathcal{X}}$ of the signal set \mathcal{X} ?
meaning?

→ From $\hat{\mathcal{X}}$, a consistent (ideal) decoder reads:

$$f(y) \in \{x \in \mathbb{S}^{n-1} \mid \underbrace{\text{sign}(Ax) = y}_{\text{consistency}} \text{ and } \underbrace{x \in \hat{\mathcal{X}}}_{\text{approx. prior}}\}$$

Model identification: geometric intuition

Toy example: $n = 3$, $m \times n$ matrix A_g has Gaussian iid entries : $(A_g)_{ij} \sim_{\text{iid}} \mathcal{N}(0,1)$

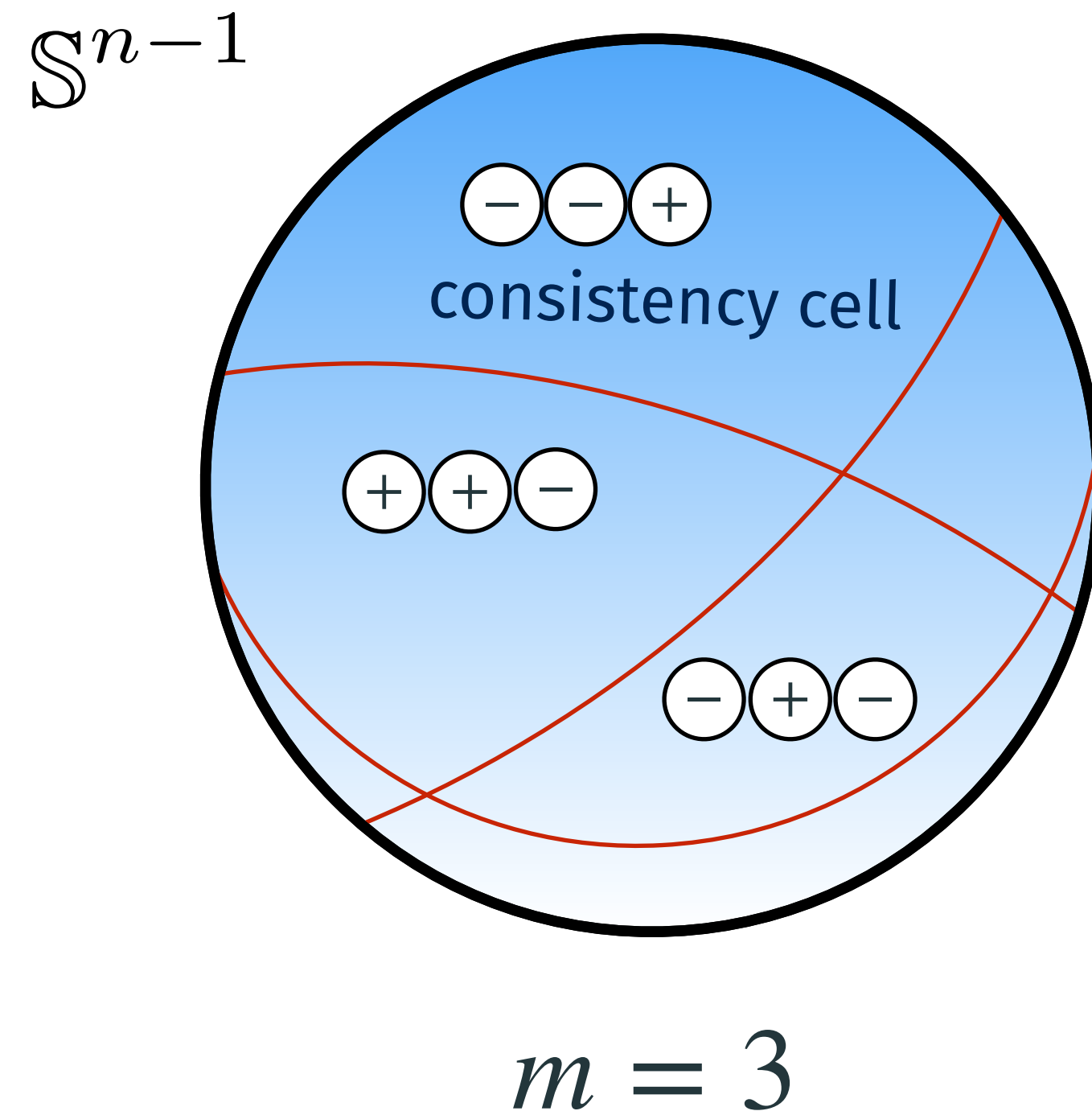


$$\Phi_g(\cdot) := \text{sign}(A_g \cdot) \text{ tessellates } \mathbb{S}^{n-1}$$

Growing number of consistency cells as $m \uparrow$
(bounded by 2^m)

Model identification: geometric intuition

Toy example: $n = 3$, $m \times n$ matrix A_g has Gaussian iid entries : $(A_g)_{ij} \sim_{\text{iid}} \mathcal{N}(0,1)$

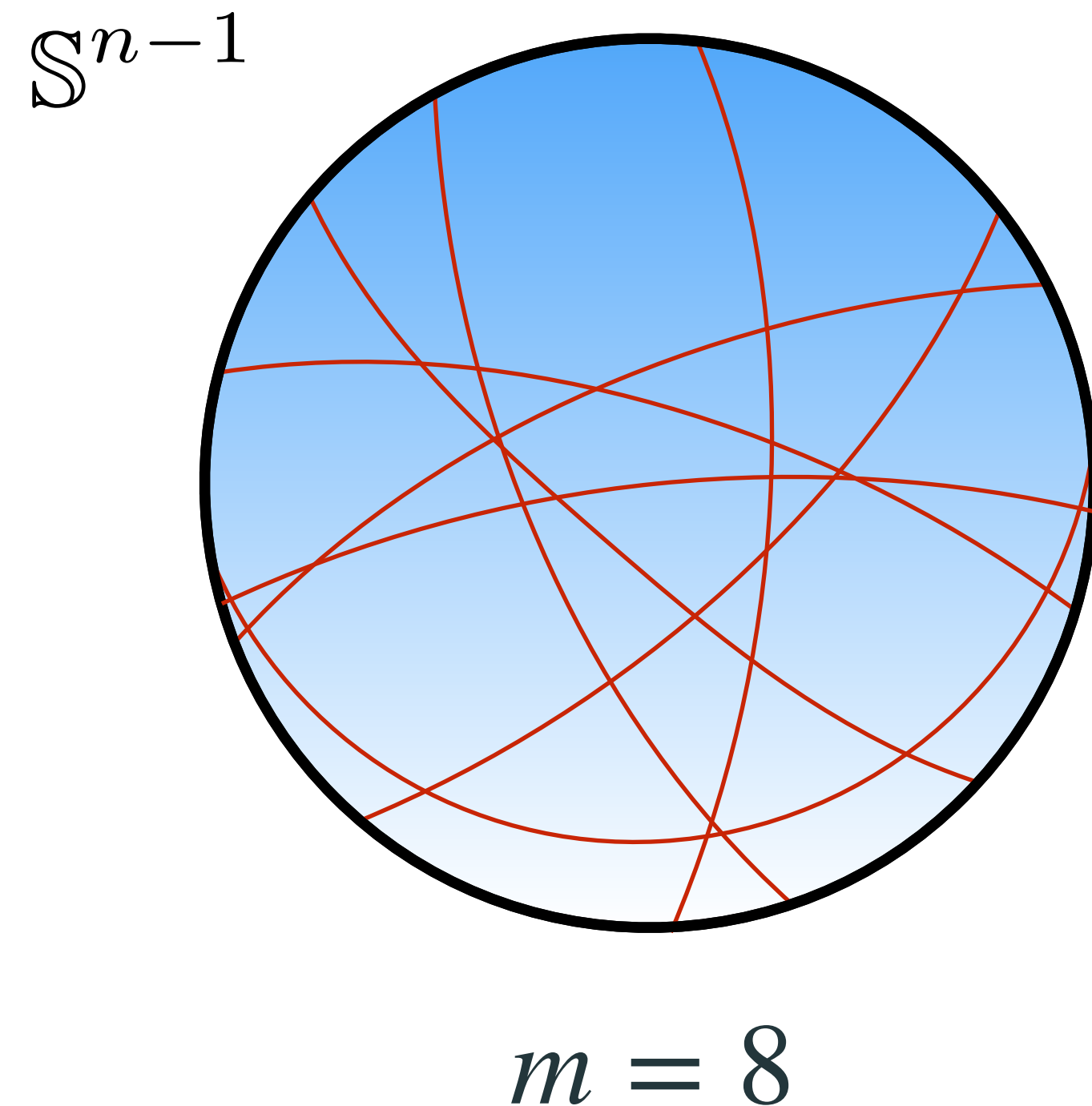


$$\Phi_g(\cdot) := \text{sign}(A_g \cdot) \text{ tessellates } \mathbb{S}^{n-1}$$

Growing number of consistency cells as $m \uparrow$
(bounded by 2^m)

Model identification: geometric intuition

Toy example: $n = 3$, $m \times n$ matrix A_g has Gaussian iid entries : $(A_g)_{ij} \sim_{\text{iid}} \mathcal{N}(0,1)$

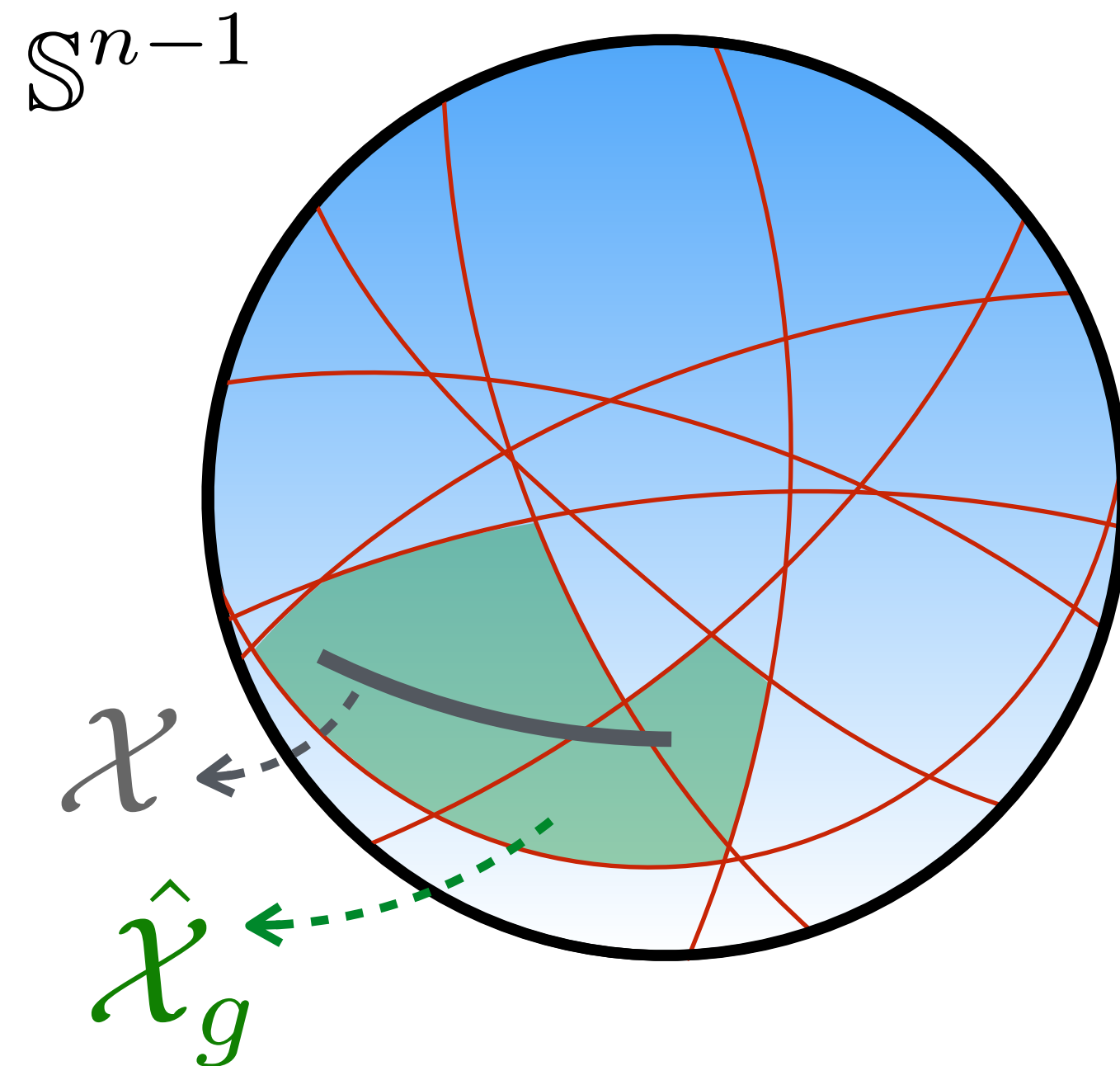


$$\Phi_g(\cdot) := \text{sign}(A_g \cdot) \text{ tessellates } \mathbb{S}^{n-1}$$

Growing number of consistency cells as $m \uparrow$
(bounded by 2^m)

Model identification: geometric intuition

Toy example: $n = 3$, $m \times n$ matrix A_g has Gaussian iid entries : $(A_g)_{ij} \sim_{\text{iid}} \mathcal{N}(0,1)$



$$\Phi_g(\cdot) := \text{sign}(A_g \cdot) \text{ tessellates } \mathbb{S}^{n-1}$$

Growing number of consistency cells as $m \uparrow$
(bounded by 2^m)

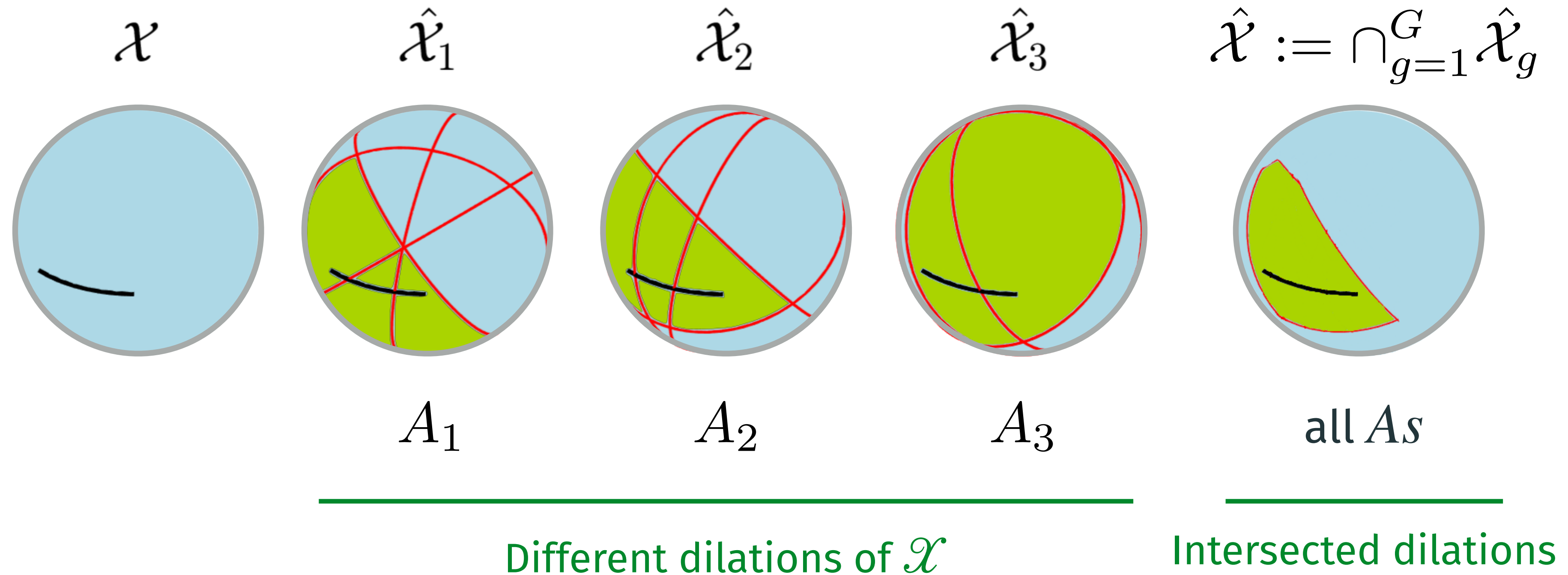
Let us define the biggest set binary consistent (wrt A_g) with \mathcal{X} :

$$\hat{\mathcal{X}}_g = \Phi_g^{-1}(A_g \mathcal{X}) = \{v \in \mathbb{S}^{n-1} \mid \exists x \in \mathcal{X}, \text{sign}(A_g v) = \text{sign}(A_g x)\}$$

→ dilation of \mathcal{X} by the “uncertainty” of $\text{sign} \circ A_g$

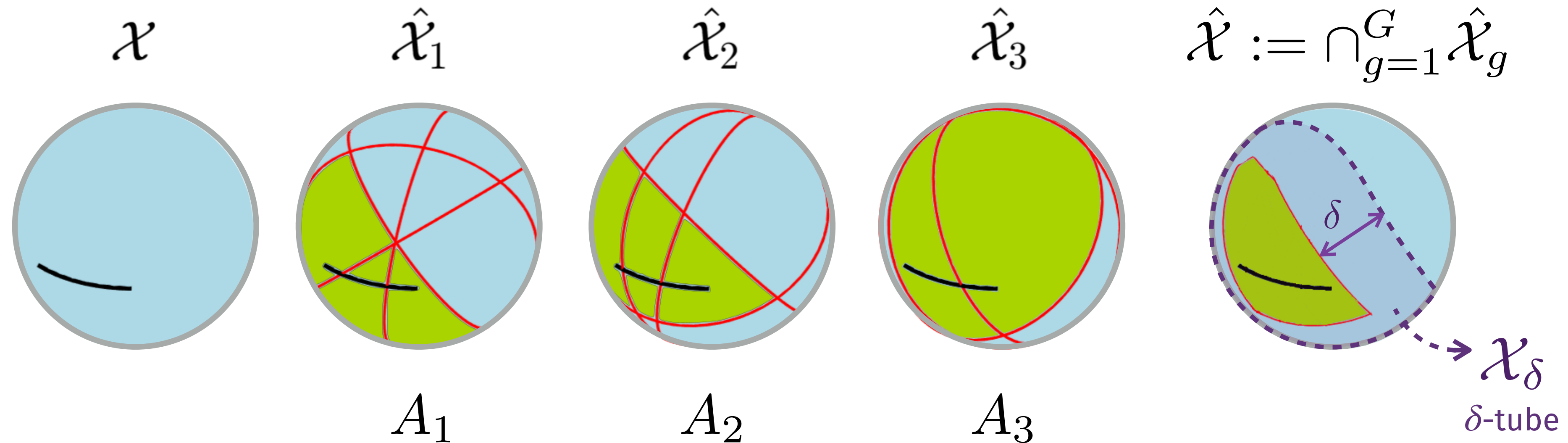
Model identification: geometric intuition

Toy example: $n = 3$, $G = 3$, $m = 4$, \mathcal{X} = black line



Model identification: geometric intuition

Toy example: $n = 3$, $G = 3$, $m = 4$, \mathcal{X} = black line



Identification error (definition)

Identify signal set up to global error $\delta \rightarrow \hat{\mathcal{X}}$ is in a δ -tube \mathcal{X}_δ

$$\hat{\mathcal{X}} \subseteq \mathcal{X}_\delta := \{v \in \mathbb{S}^{n-1} : \|x - v\| \leq \delta, x \in \mathcal{X}\}$$

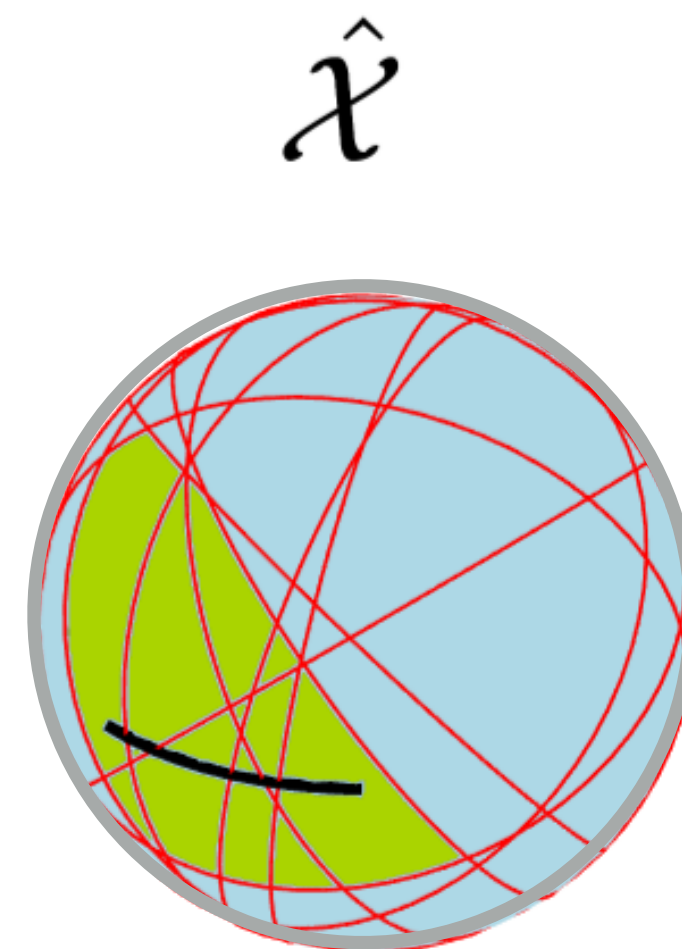
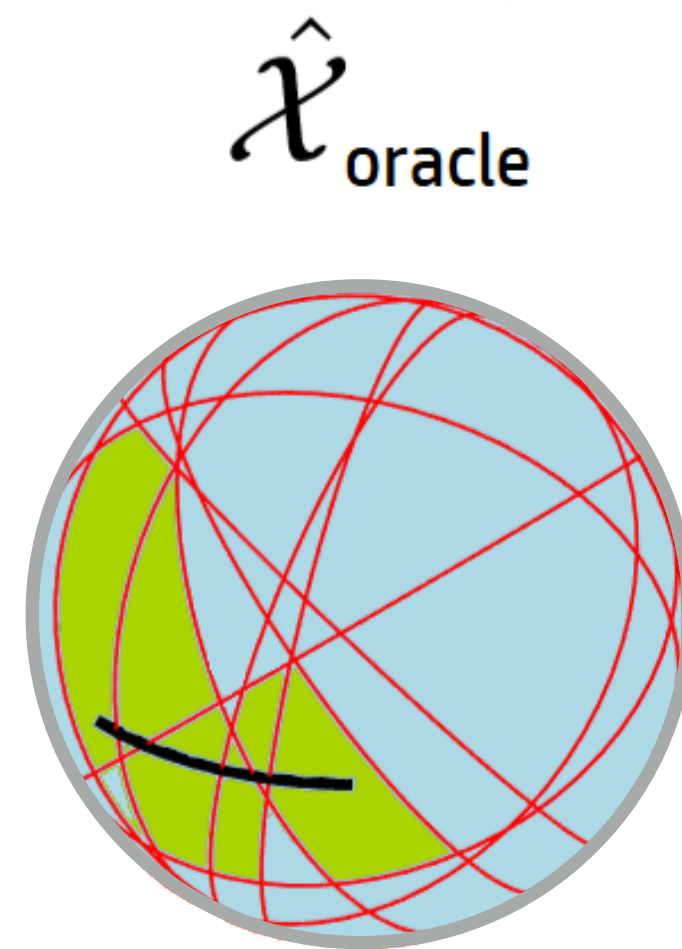
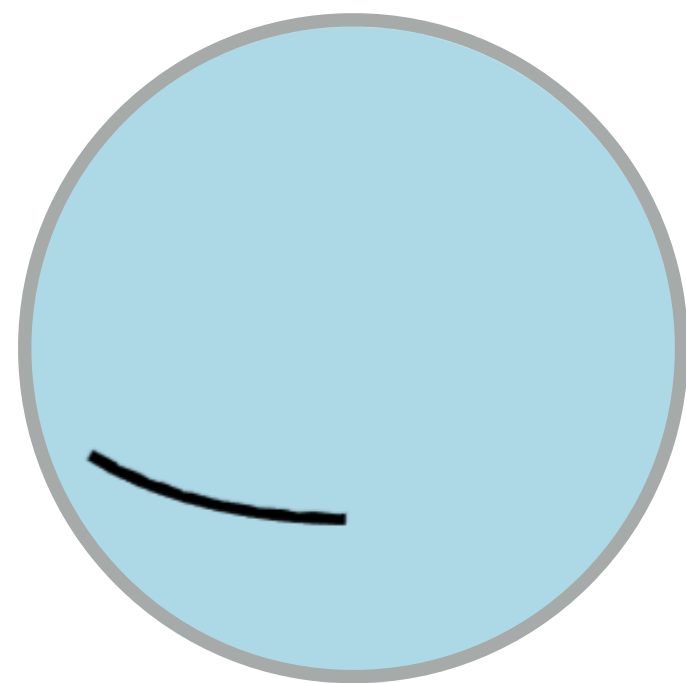
Upper/Lower bound on δ ? Sample complexity?

Lower bound on δ (via an *oracle* standpoint)

Oracle estimation: We access to G observations of each $x \in \mathcal{X}$

$$(\text{sign}(A_1 x), \dots, \text{sign}(A_G x)) \leftrightarrow \text{sign}(\bar{A}x) \in \{\pm 1\}^{mG}, \text{ with } \bar{A} = \begin{pmatrix} A_1 \\ \vdots \\ A_G \end{pmatrix} \in \mathbb{R}^{mG \times n}$$

$$\hat{\mathcal{X}}_{\text{oracle}} = \{v \in \mathbb{S}^{n-1} \mid \exists x \in \mathcal{X}, \text{sign}(\bar{A}v) = \text{sign}(\bar{A}x)\}$$



$$\hat{\mathcal{X}}_{\text{oracle}} = \bar{\Phi}^{-1}(\bar{A}\mathcal{X})$$

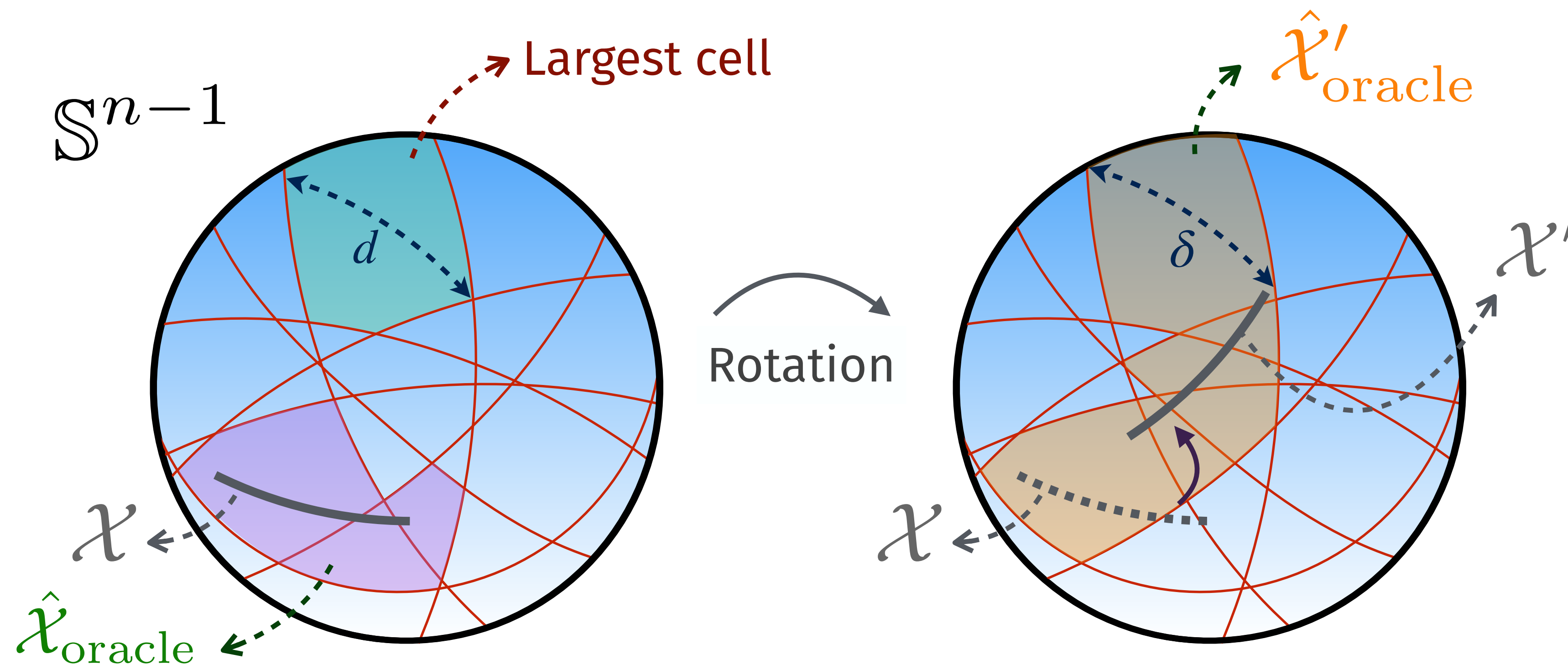


Lower bound on δ (via an oracle standpoint)

Question: smallest δ such that $\mathcal{X}_{\text{oracle}} \subset \mathcal{X}_\delta$? (whatever \mathcal{X} 's orientation)

Theorem: for any set $\mathcal{X} \subset \mathbb{S}^{n-1}$, there exists a rotated set \mathcal{X}' s.t.

$$\delta \geq d := \text{diameter largest consistency cell of } \text{sign}(\bar{A} \cdot)$$



Lower bound on δ (via an *oracle* standpoint)

Question: smallest δ such that $\mathcal{X}_{\text{oracle}} \subset \mathcal{X}_\delta$? (whatever \mathcal{X} 's orientation)

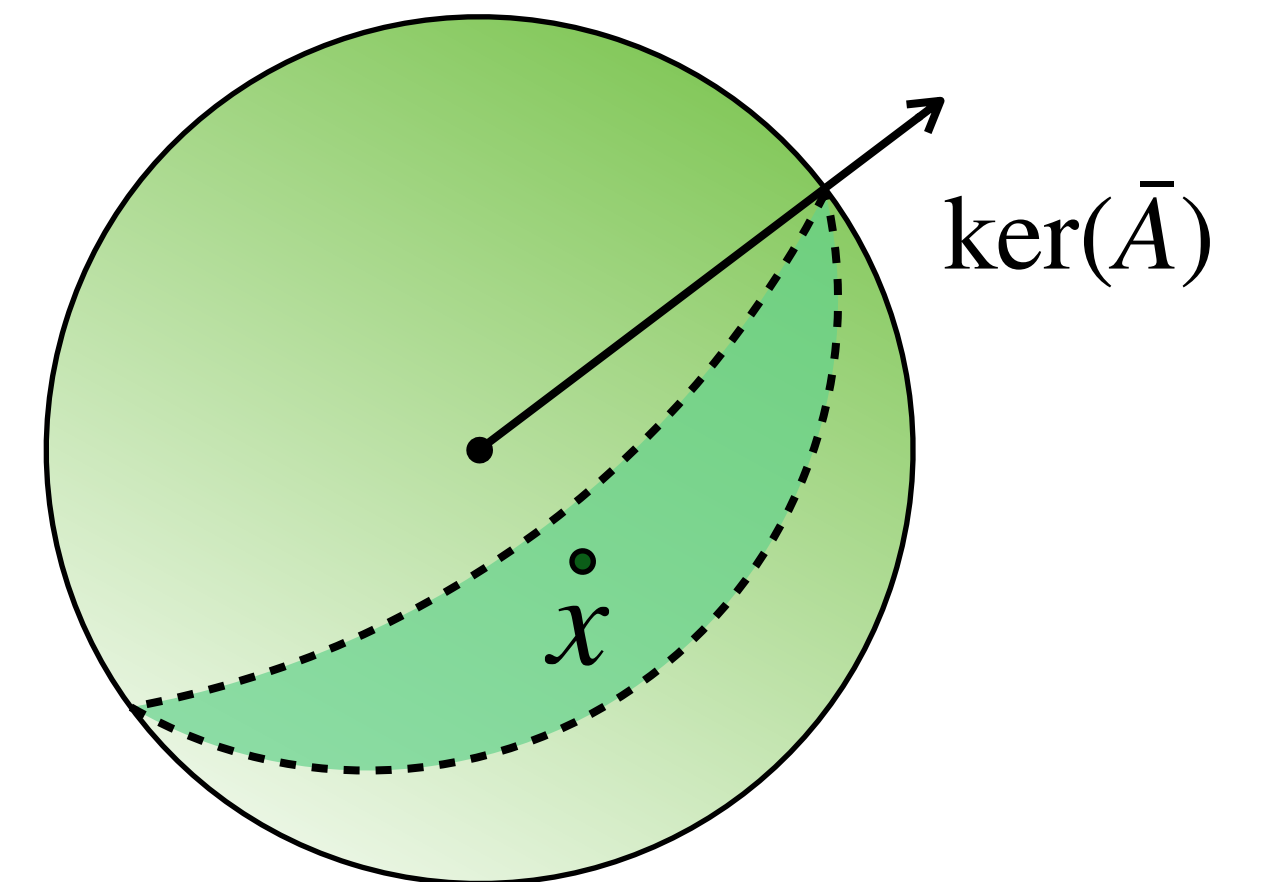
Theorem: for any set $\mathcal{X} \subset \mathbb{S}^{n-1}$, there exists a rotated set \mathcal{X}' s.t.

$$\delta \geq d := \text{diameter largest consistency cell of } \text{sign}(\bar{A} \cdot)$$

Consequences: we can show the following

1. If $\text{rank}(\bar{A}) < n$, \exists consistency cells with diameter 2

→ Model identification error is trivially large



Lower bound on δ (via an *oracle* standpoint)

Question: smallest δ such that $\mathcal{X}_{\text{oracle}} \subset \mathcal{X}_\delta$? (whatever \mathcal{X} 's orientation)

Theorem: for any set $\mathcal{X} \subset \mathbb{S}^{n-1}$, there exists a rotated set \mathcal{X}' s.t.

$$\delta \geq d := \text{diameter largest consistency cell of } \text{sign}(\bar{A} \cdot)$$

Consequences: we can show the following

1. If $\text{rank}(\bar{A}) < n$, \exists consistency cells with diameter 2

→ Model identification error is trivially large

2. We need at least $m > n/G$ measurements per operator

→ No learning of \mathcal{X} with $G = 1$ (w/o invariance)

3. The maximum cell radius $\geq \frac{2n}{3mG}$ (counting argument)

→ δ cannot decrease faster than $\propto m^{-1}G^{-1}$

Thao & Vetterli (1996, Theorem A.7)

$$|\{\text{sign}(\bar{A}\mathcal{X})\}| \leq \binom{mG}{n} 2^n$$

Upper bound on δ (with the help of randomness)



Definition: $\text{boxdim}(S) = \lim_{\epsilon \rightarrow 0^+} \sup \frac{\log \mathfrak{N}(S, \epsilon)}{\log 1/\epsilon}$

Assumption: The signal set \mathcal{X} is **low-dimensional**
 $\Leftrightarrow \mathcal{X}$ has **box-counting dimension** $k \ll n$

Examples: sparse dictionaries, manifold models, etc.

PAC-like result:

If $\text{boxdim}(\mathcal{X}) < k$, given $\delta > 0$, and A_1, \dots, A_G iid $m \times n$ Gaussian,

Which conditions on $m \gtrsim ?$ to get $\mathbb{P}[\hat{\mathcal{X}} \subset \mathcal{X}_\delta] \simeq 1$?



Upper bound on δ (with the help of randomness)



Definition: $\text{boxdim}(S) = \lim_{\epsilon \rightarrow 0^+} \sup \frac{\log \mathfrak{N}(S, \epsilon)}{\log 1/\epsilon}$

Assumption: The signal set \mathcal{X} is **low-dimensional**

$\Leftrightarrow \mathcal{X}$ has **box-counting dimension** $k \ll n$

Examples: sparse dictionaries, manifold models, etc.



PAC-like result:

If $\text{boxdim}(\mathcal{X}) < k$, given $\delta > 0$, and A_1, \dots, A_G iid $m \times n$ Gaussian,

and if $m \geq C \frac{1}{\delta} (k + \frac{n}{G}) \text{plog}(m, n, G, k, \xi)$

then $\mathbb{P}[\hat{\mathcal{X}} \subset \mathcal{X}_\delta] \geq 1 - \xi$

Upper bound on δ (with the help of randomness)



Consequences of this theorem:

1. The identification error of \mathcal{X} decreases as

$$\delta = \frac{(k + \frac{n}{G})}{m} \log\left(\frac{nm}{k + \frac{n}{G}}\right)$$

2. We require at least $m \geq k + \frac{n}{G}$ measurements *per operator*

3. For $G > \frac{n}{k}$, error $\delta \sim$ signal recovery error in one-bit compressive sensing

If $\text{boxdim}(\mathcal{X}) < k$, and $x \in \mathcal{X}$ measured as $y = \text{sign}(Ax)$, then, w.h.p,

$$\min_{x' \in \mathcal{X}} \|\text{sign}(Ax') - y\| = O\left(\frac{k}{m}\right)$$

[LJ, J. Laska, P. Boufounos, R. Baraniuk, 2013] [J. Tachella, LJ, 2023]

Sample Complexity bound (again with randomness)



How many distinct binary observations N to hope to estimate $\hat{\mathcal{X}}$?

→ an upper bound on N is $S := |\bigcup_{g=1}^G \text{sign}(A_g \mathcal{X})|$

If $\text{boxdim}(\mathcal{X}) < k$, and A_1, \dots, A_G are $m \times n$ matrices,

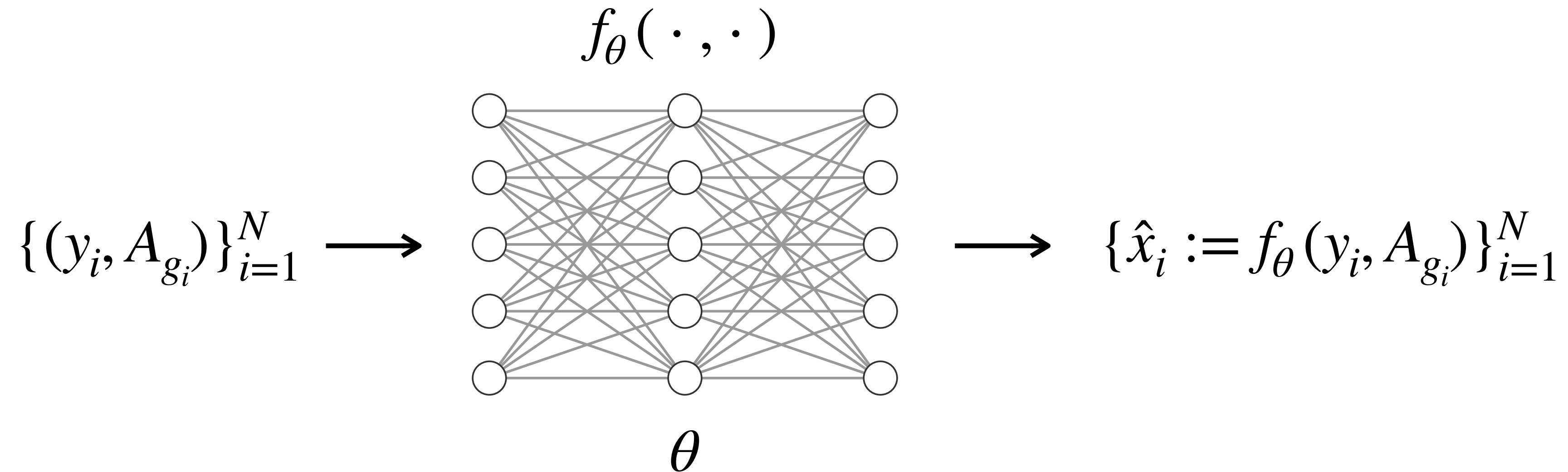
If \mathcal{X} is a subspace, then $S \leq G 2^k \binom{m}{k} \leq G \left(\frac{2em}{k}\right)^k$ [Thao, Vetterli, '96]

If \mathcal{X} is a union of L subspaces, then $S \leq GL \left(\frac{2em}{k}\right)^k$

If \mathcal{X} general, & iid Gaussian A_1, \dots, A_G , then $S \leq G \left(\frac{m\sqrt{n}}{k}\right)^{5k}$ w.h.p.

→ Exponential in the model dim k but not the ambient dimension n !

Learning to reconstruct from binary measurements **in practice?**



Goal:

Learning a reconstruction network $\hat{x} = f_{\theta}(y, A_g)$

with a **self-supervised loss** \mathcal{L} which uses only $\{(y_i, A_{g_i})\}_{i=1}^N$

Warning: No clear link with the theory (yet)

Multi-operator case

Self-supervised training loss: given a reconstruction model f_θ

$$\arg \min_{\theta} \mathcal{L}_{\text{MC}}(\theta)$$

with:

$$\mathcal{L}_{\text{MC}}(\theta) := \sum_{i=1}^N \log [1 + \exp (-y_i A_{g_i} f_\theta(y_i, A_{g_i}))] \quad (\text{Logistic loss})$$

→ promotes measurement consistency : $y_i \approx \text{sign}(A_{g_i} f_\theta(y_i, A_{g_i}))$

Problem: $f_\theta(y, A_g) = A_g^\dagger y := A_g^\top (A_g A_g^\top)^{-1} y$ is solution

→ but f_θ acts independently for each A_g

→ no gain in increasing G !

Multi-operator case

Self-supervised training loss: given a reconstruction model f_θ

$$\arg \min_{\theta} \mathcal{L}_{\text{MC}}(\theta) + \mathcal{L}_{\text{CC}}(\theta)$$

with:

$$\mathcal{L}_{\text{MC}}(\theta) := \sum_{i=1}^N \log [1 + \exp (-y_i A_{g_i} f_\theta(y_i, A_{g_i}))] \quad (\text{Logistic loss})$$

→ promotes **measurement consistency**: $y_i \approx \text{sign}(A_{g_i} f_\theta(y_i, A_{g_i}))$

$$\mathcal{L}_{\text{CC}}(\theta) := \sum_{i=1}^N \sum_{s=1}^G \left\| f_\theta(A_s f_\theta(y_i, A_{g_i}), A_s) - f_\theta(y_i, A_{g_i}) \right\|^2$$

→ promotes **cross-operator consistency**, e.g., prevents MC sol $f_\theta(y, A_g) = A_g^\dagger y$

Remarks:

- Network-agnostic scheme (applicable to any existing deep model)
- We called this “*Self-Supervised learning loss for training reconstruction networks from Binary Measurement data alone*” (SSBM)

Single operator with equivariance

Self-supervised training loss: given a reconstruction model f_θ

$$\arg \min_{\theta} \mathcal{L}_{\text{MC}}(\theta) + \mathcal{L}_{\text{Eq}}(\theta)$$

with:

$$\mathcal{L}_{\text{MC}}(\theta) := \sum_{i=1}^N \log \left[1 + \exp \left(- y_i A_{g_i} f_\theta(y_i, A_{g_i}) \right) \right] \quad (\text{Logistic loss})$$

→ promotes **measurement consistency**: $y_i \approx \text{sign}(A_{g_i} f_\theta(y_i, A_{g_i}))$

$$\mathcal{L}_{\text{Eq}}(\theta) := \sum_{i=1}^N \sum_{g=1}^G \left\| f_\theta \left(A T_g f_\theta(y_i, A), A T_g \right) - T_g f_\theta(y_i, A) \right\|^2$$

→ promotes **equivariance** of $f_\theta \circ A$: $(f \circ A)(T_g \cdot) = T_g(f \circ A)(\cdot)$

Remarks:

- Network-agnostic scheme (applicable to any existing deep model)
- We called this “*Self-Supervised learning loss for training reconstruction networks from Binary Measurement data alone*” (SSBM)

Experiments

Operators

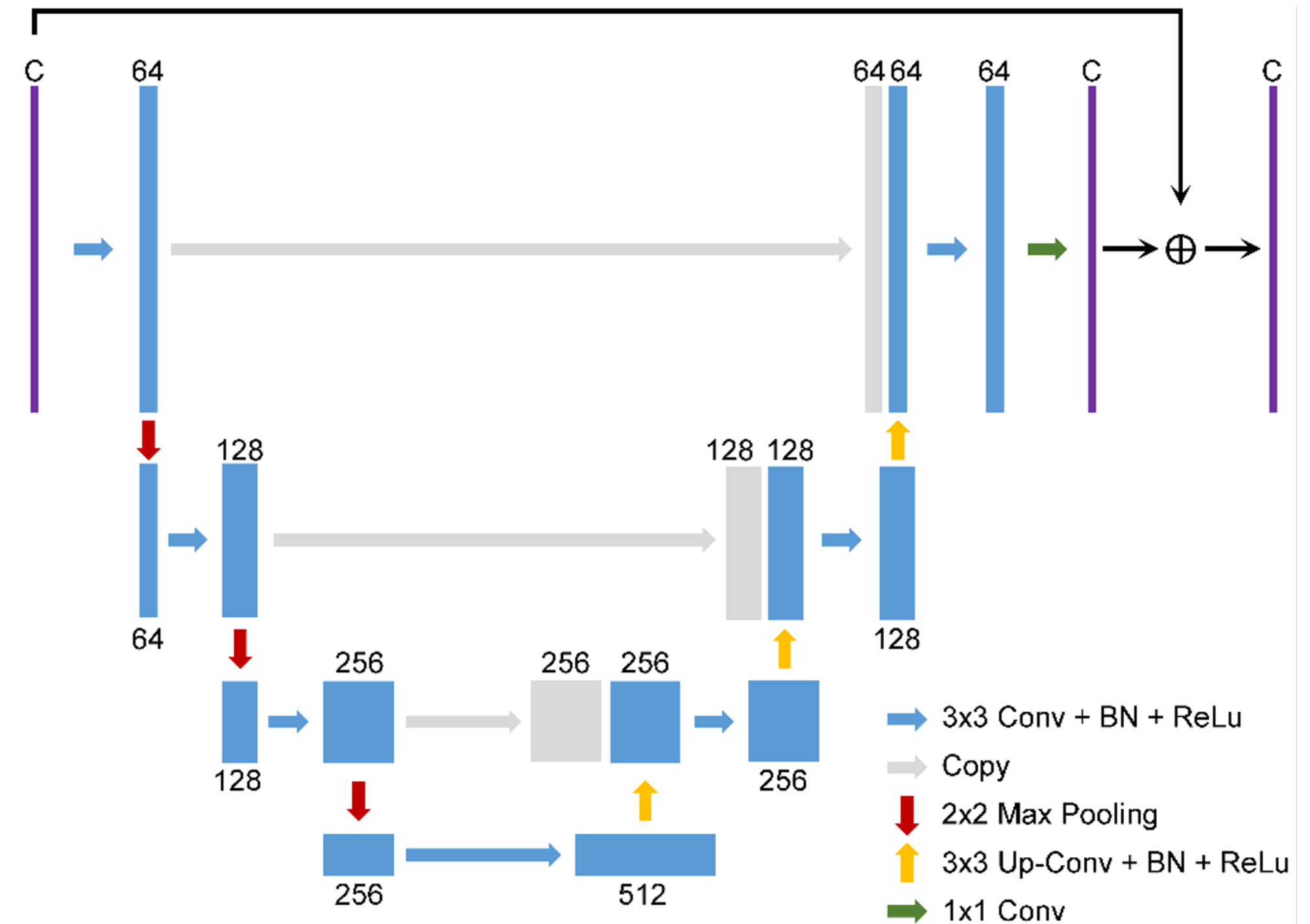
- $\{A_g\}_{g=1}^G$ with Gaussian iid entries

Network

- $f_{\theta}(y, A) = g_{\theta} \circ A^{\top}(y)$
where g_{θ} is a U-net CNN

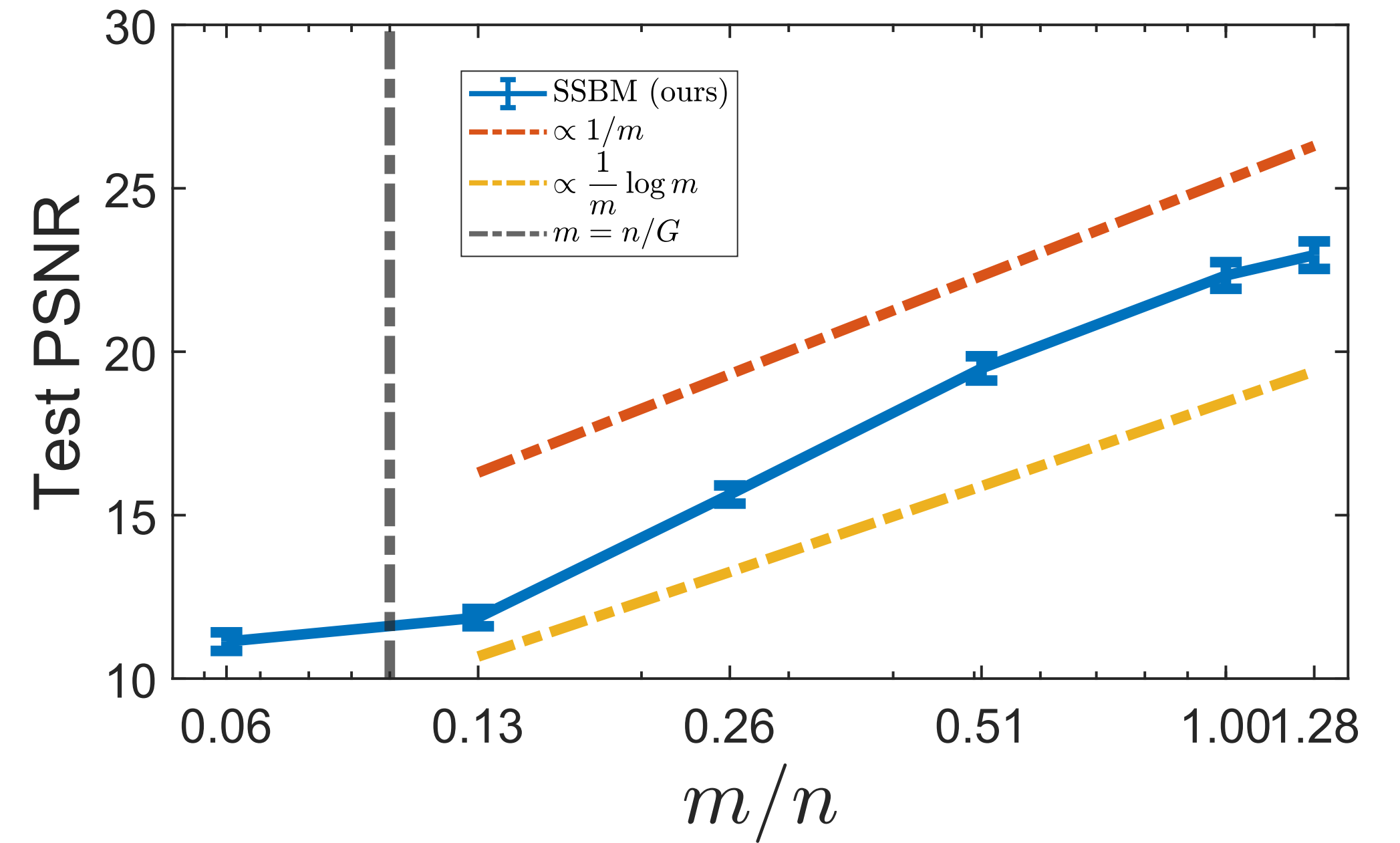
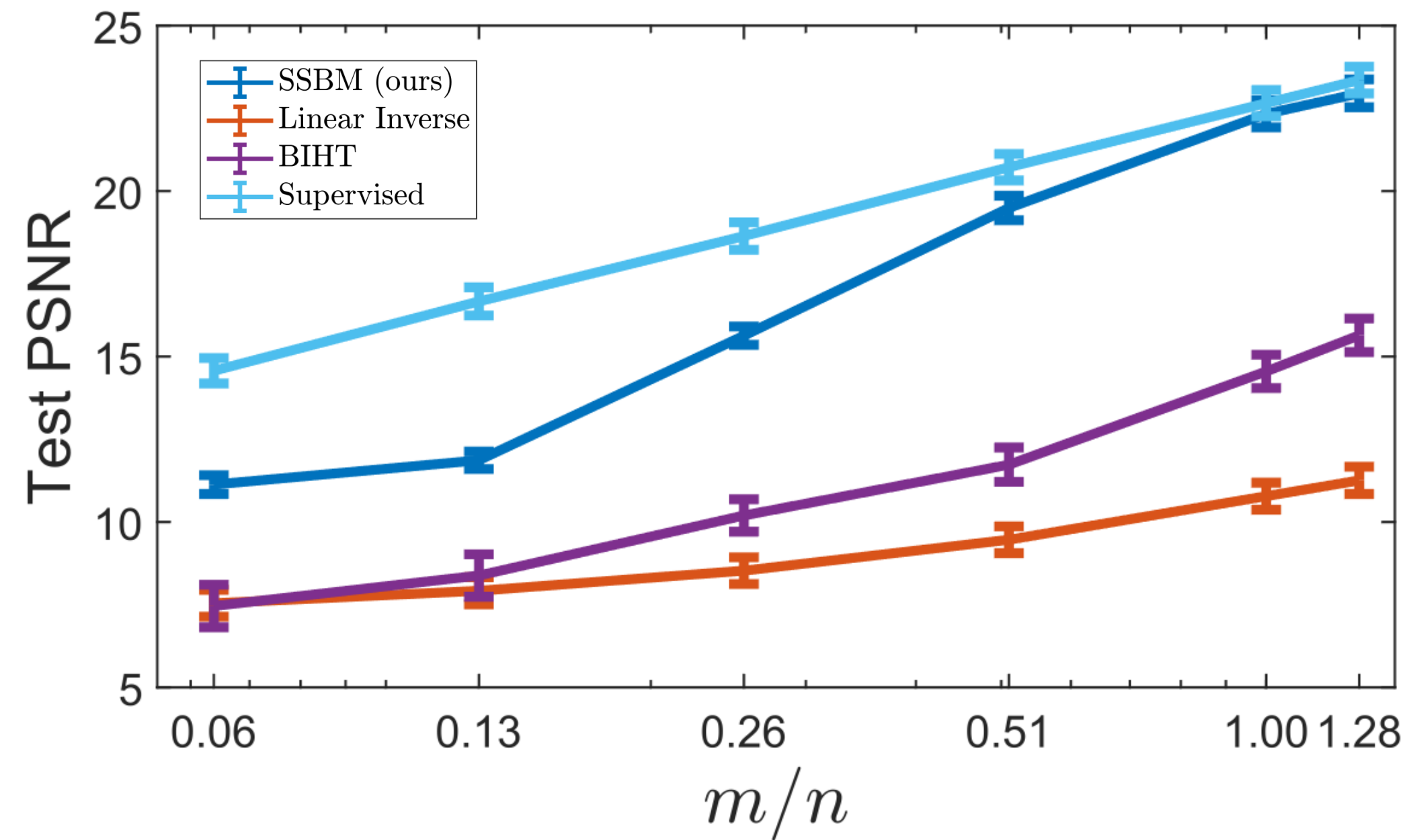
Comparison with

- ▶ Linear inverse $A^\top y_i$ (no training)
- ▶ Binary IHT (BIHT) with wavelets (no training)
- ▶ Fully supervised loss
- ▶ **SSBM (proposed)**



MNIST dataset

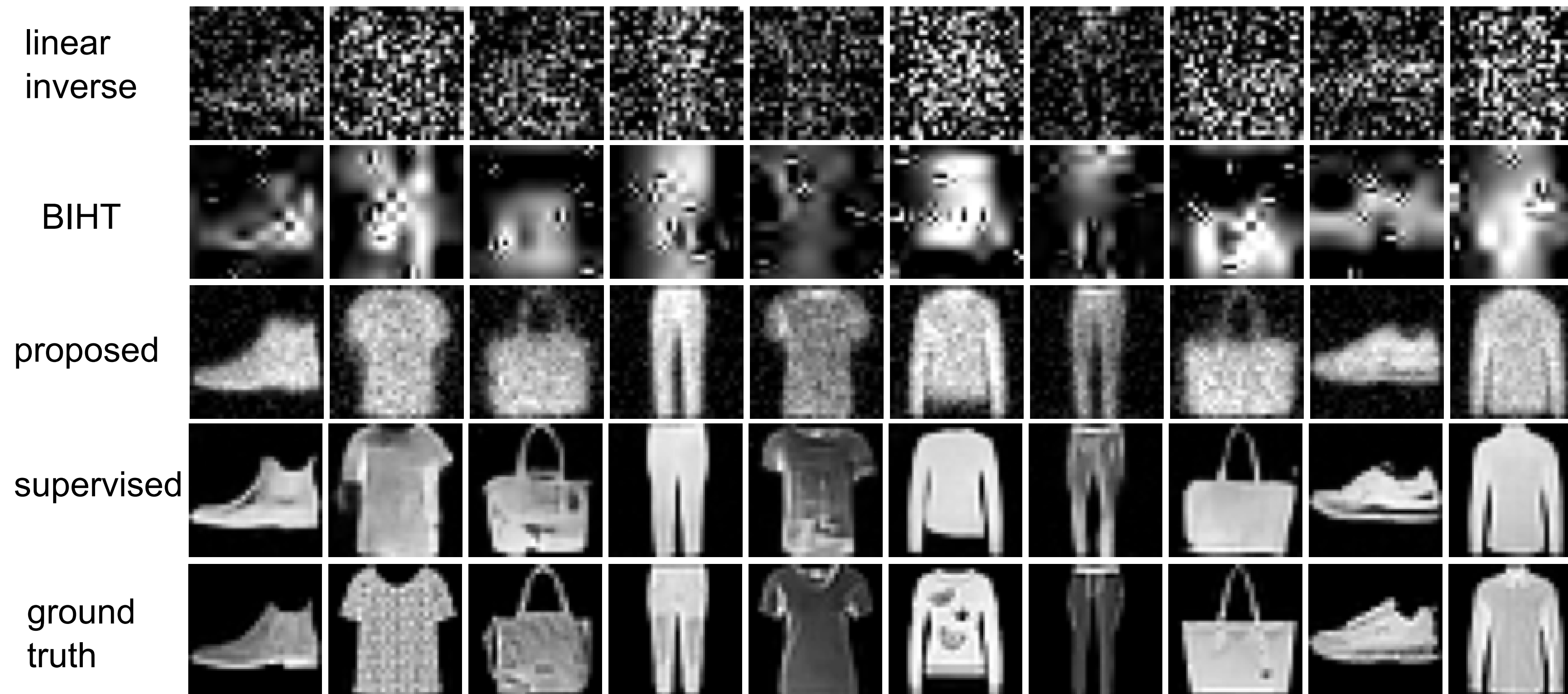
Multiple operators ($G = 10$), images have $n = 784$ pixels.



$$\text{Test PSNR} := \frac{1}{N'} \sum_{i=1}^{N'} \text{PSNR} (x'_i, f_{\theta} (\text{sign} (A_{g_i} x'_i), A_{g_i})), x'_i \in \text{“test set”}$$

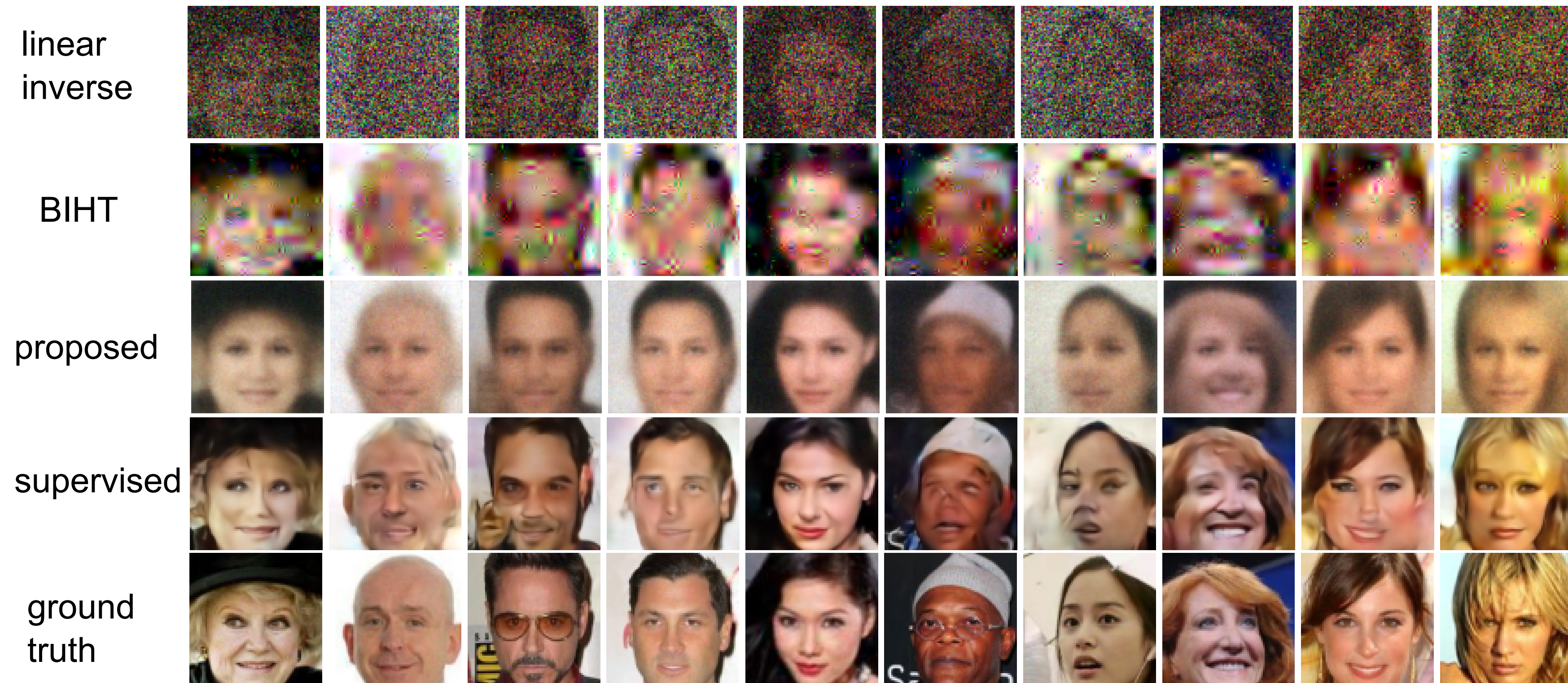
Fashion MNIST dataset

Multiple operators ($G=10$), with $m=300$, images have $n=784$ pixels.



CelebA dataset

Multiple operators ($G=10$) with $m=9830$, images have $n=49152$ pixels.



Conclusion and take-away messages

New unsupervised learning framework for binary data

Theory: we have studied several conditions for learning \mathcal{X} , e.g.,

- ▶ Lower and upper bounds on its identification error δ
- ▶ Required number of measurements N

Practice: Deep learning approach

- ▶ Self-supervised loss which can be applied to any model

Ongoing/future work

- ▶ Other non-linear inverse problems (such as *saturation*, or *phase retrieval* $|Ax|^2$)
- ▶ Upper bounds for the invariant case?
- ▶ Noise/dither ? $y = \text{sign}(Ax + \epsilon)$

Thank you!

- 📄 Chen, Tachella and Davies, “[Robust Equivariant Imaging: a fully unsupervised framework for learning to image from noisy and partial measurements](#)”, CVPR 2022 (Oral)
- 📄 Chen, Tachella and Davies, “[Equivariant Imaging: Learning Beyond the Range Space](#)”, ICCV 2021 (Oral)
- 📄 Tachella, Chen and Davies, “[Unsupervised Learning From Incomplete Measurements for Inverse Problems](#)”, NeurIPS 2022.
- 📄 Tachella, Chen and Davies, “[Sensing Theorems for Unsupervised Learning in Inverse Problems](#)”, JMLR 2023.
- 📄 Chen, Davies, Eerhardt, Schonlieb, Ferry and Tachella, “[Imaging with Equivariant Deep Learning](#)”, IEEE SPM 2023.
- 📄 **Tachella and Jacques, “[Learning to Reconstruct Signals from Binary Measurements](#)”, TMLR+ICLR’24, 2023**

... and others on demand