Learning to Reconstruct Signals From Binary Measurements

Julián Tachella
CNRS, Physics Laboratory
ENS Lyon, France

Laurent Jacques
INMA/ICTEAM
UCLouvain, Belgium

March 5, 2024

Inverse problems (IP)

Measurements in $\mathbb{R}^m$

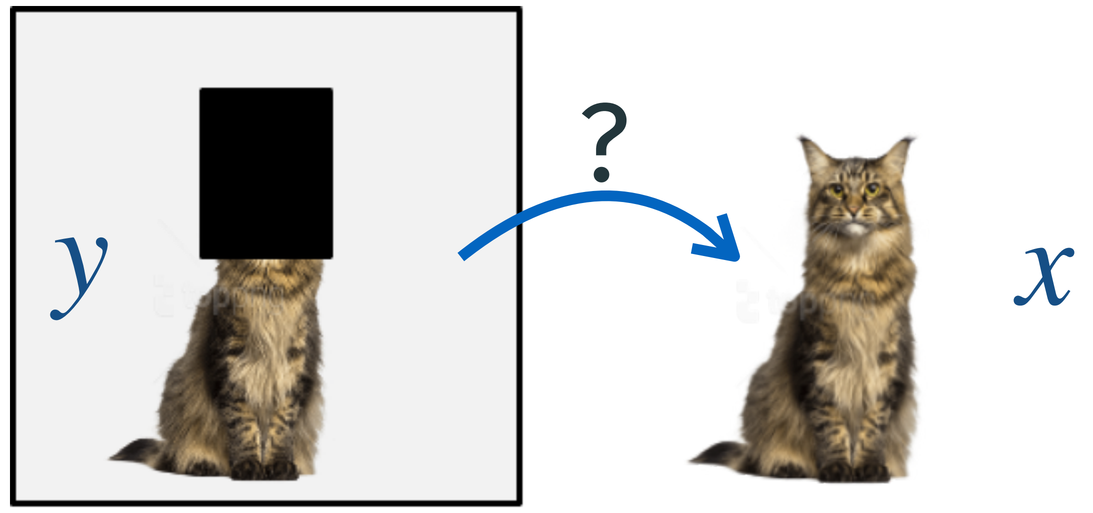
Measurements operator

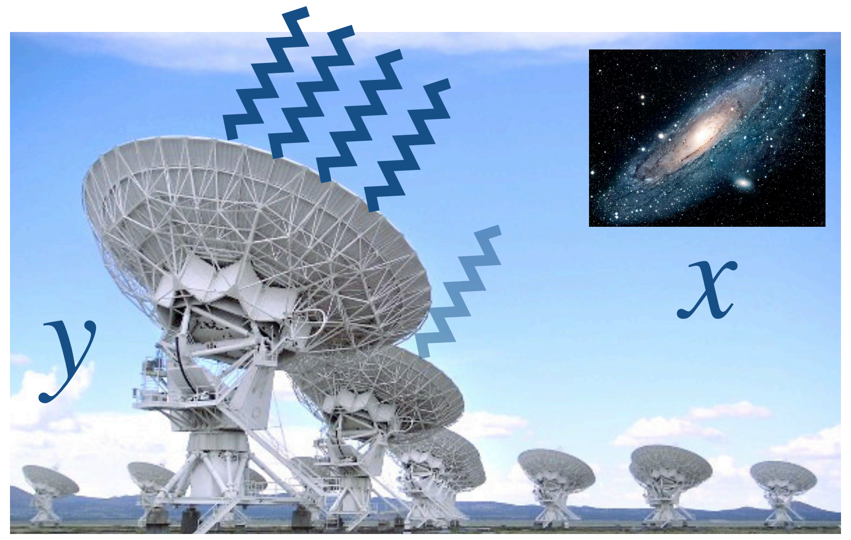
$$y = A(x) + \varepsilon$$

Signal in $\mathbb{R}^n$

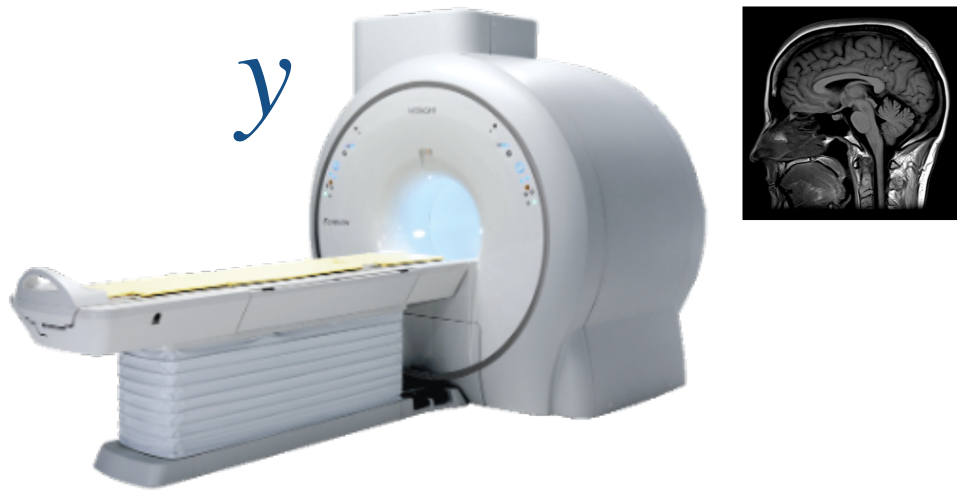
noise in $\mathbb{R}^m$

Image inpainting





Radio Astronomy



MRI

Recommender system

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

Inverse problems (IP)

Measurements in $\mathbb{R}^m$

Measurements operator

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

Signal in $\mathbb{R}^n$

noise in $\mathbb{R}^m$

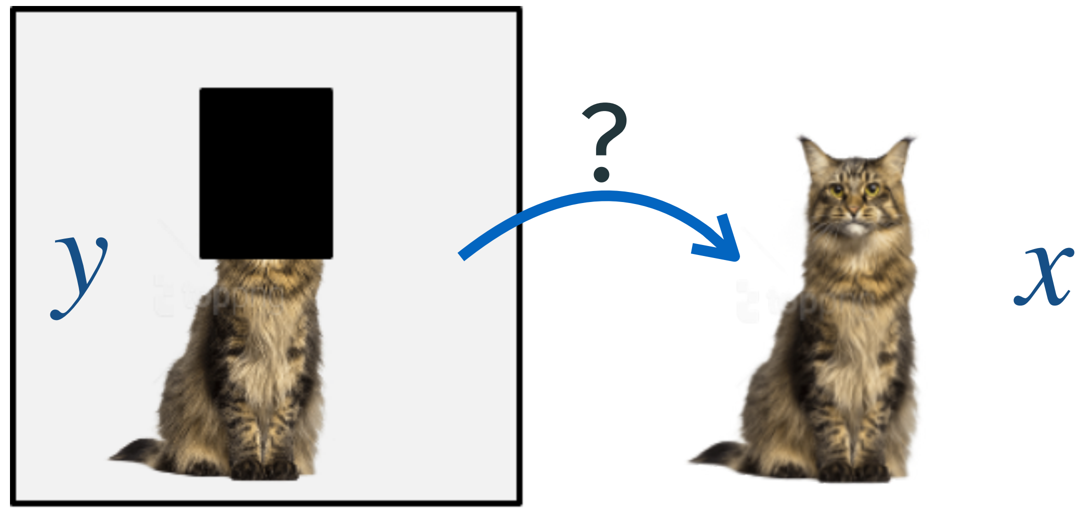
Ill-posed:

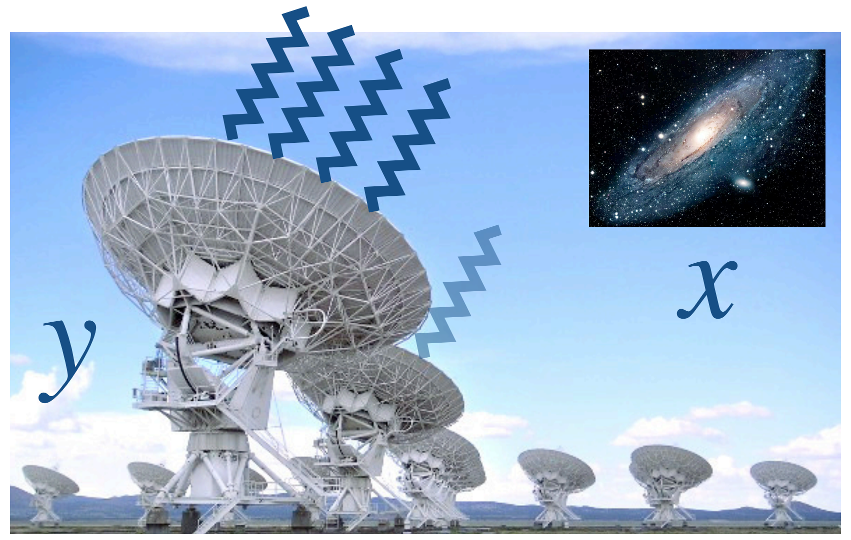
many x consistent with y (e.g., if $m < n$)

Solution:

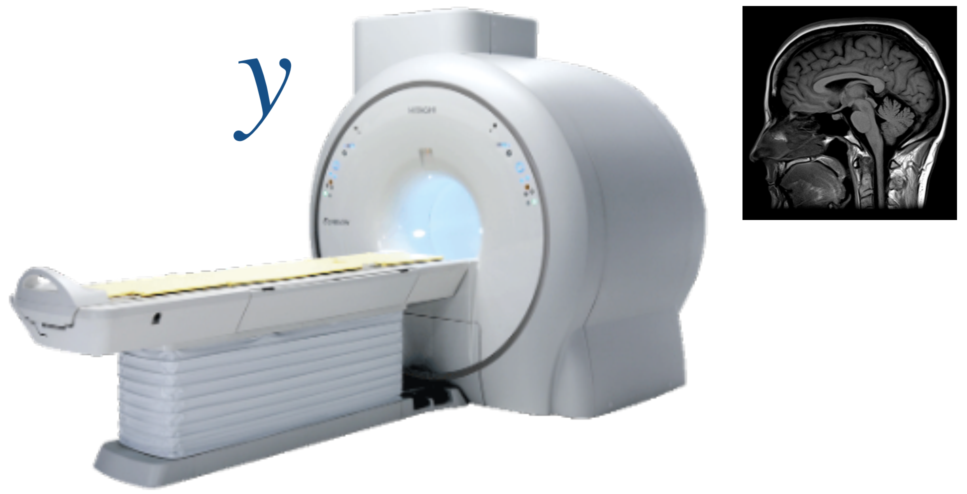
Restrict to a set of plausible signals $\underset{\text{Size?}}{\underline{X \ni}}\ x$

Image inpainting





Radio Astronomy



MRI

Recommender system

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

Solving IP: regularised reconstruction

Idea: use a *prior* ≡ *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, ...

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
- Hard to define a good $\rho(x)$ in real world problems

$$\mathscr{X} = \{$$

$$\}$$

Cat friendly presentation

Solving IP: learning approach

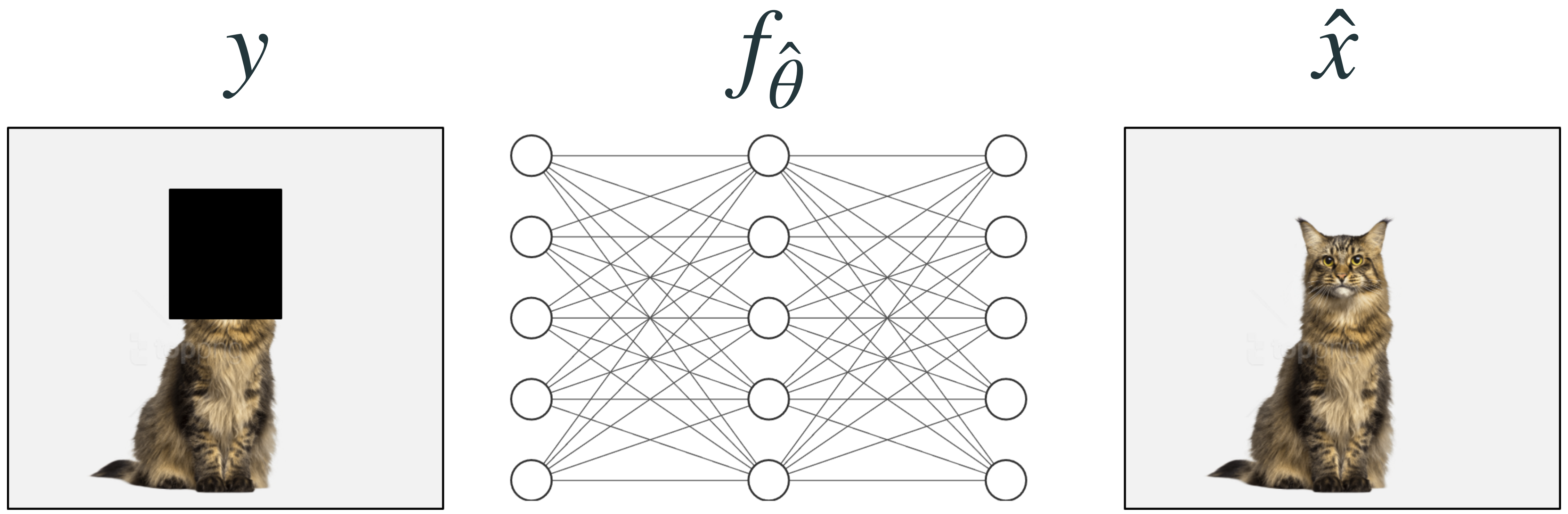
Idea:

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

Solving IP: learning approach

Idea:

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



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

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

Implicitly learn both the prior & distribution, and the reconstruction

Solving IP: learning approach

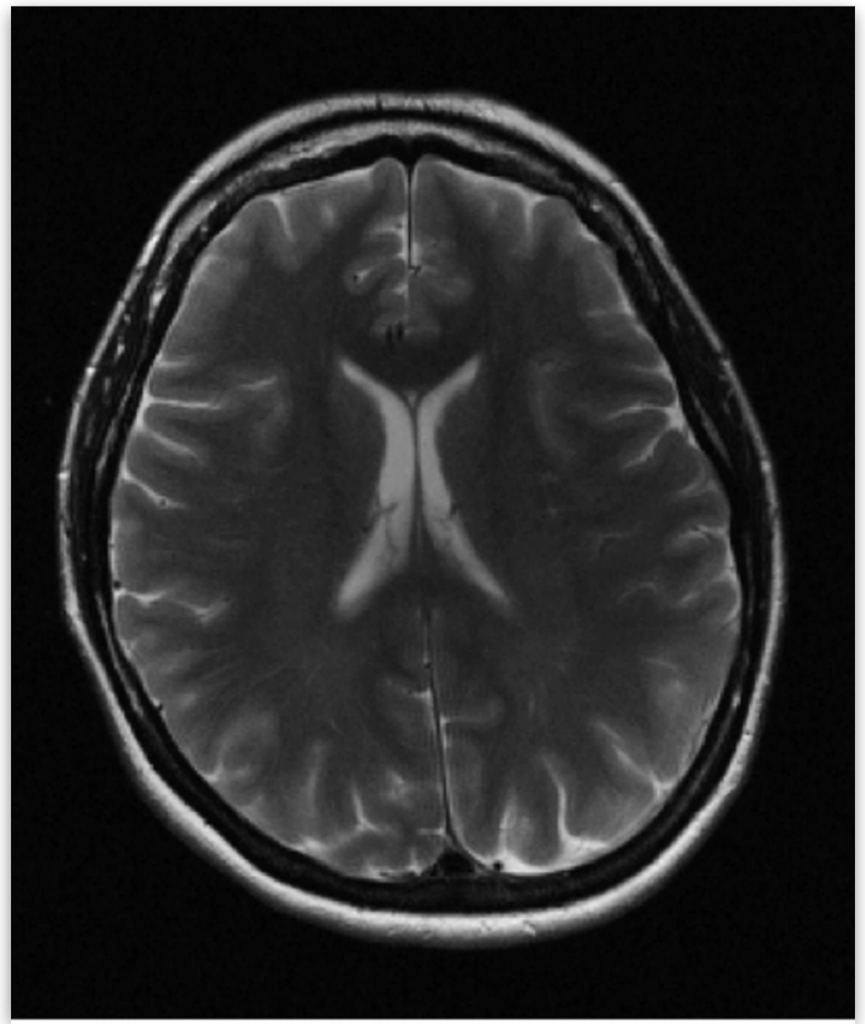
Advantages:

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

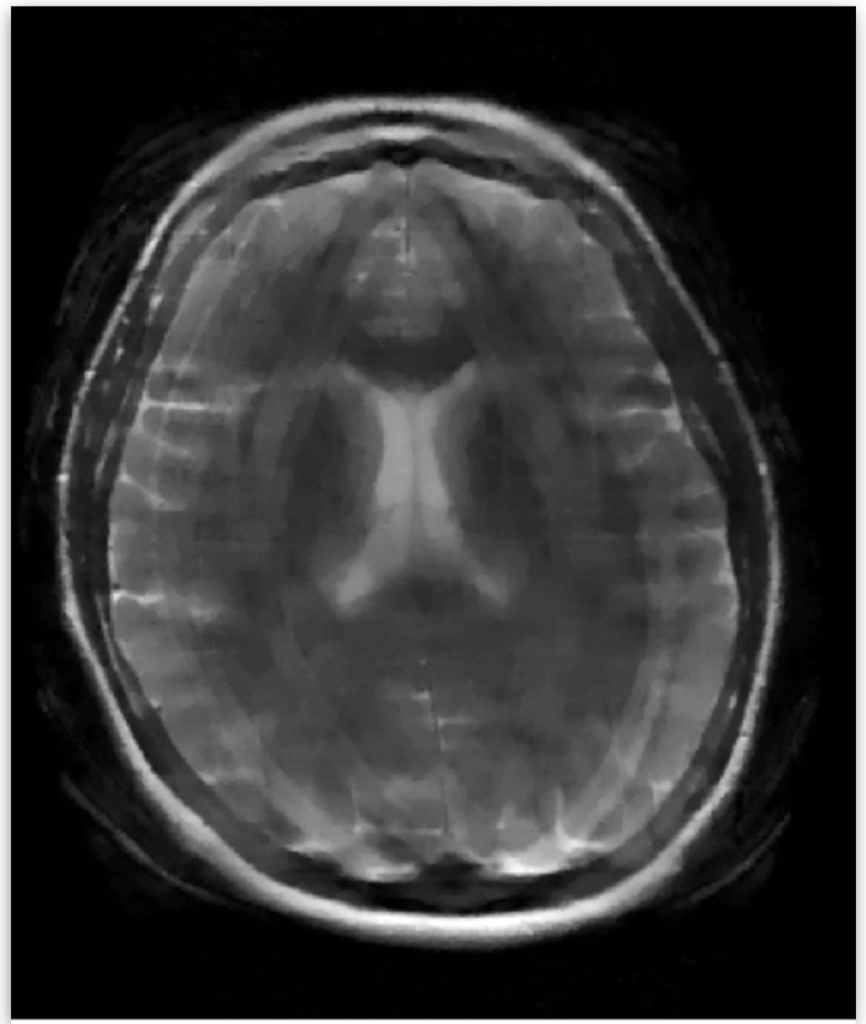
Example:

fastMRI

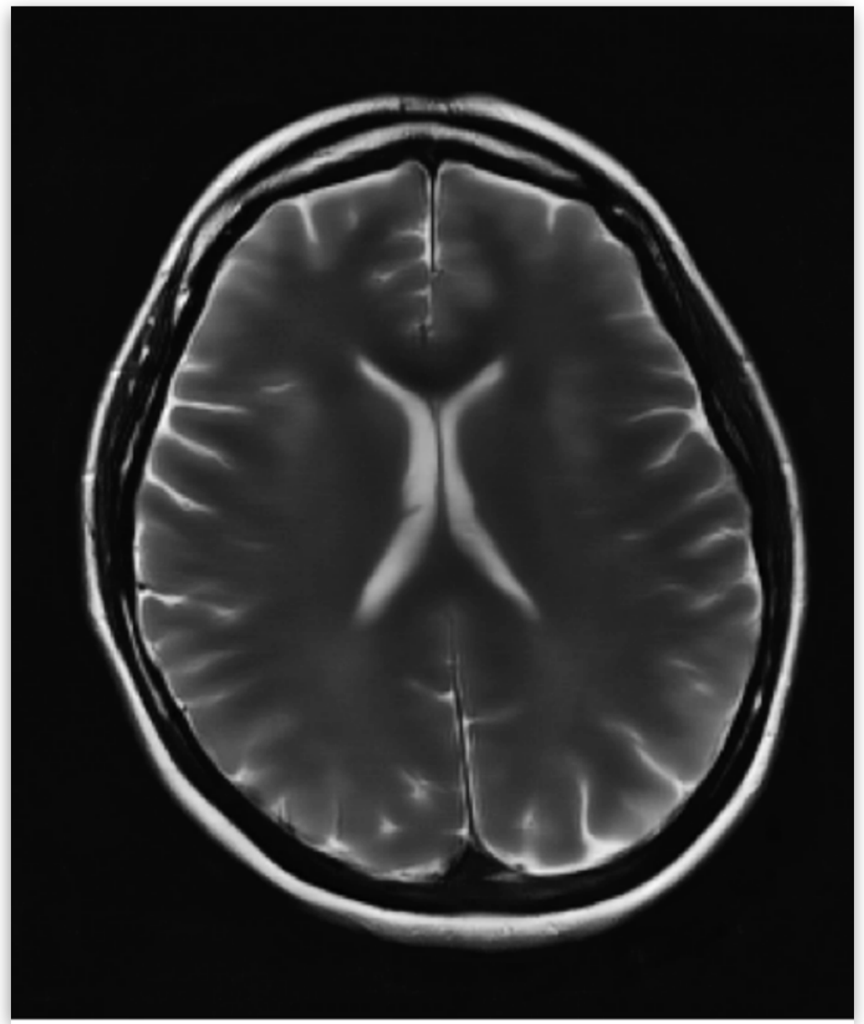
Accelerating MR Imaging with AI



Ground-truth



Total variation
(28.2 dB)



Deep network
(28.2 dB)

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

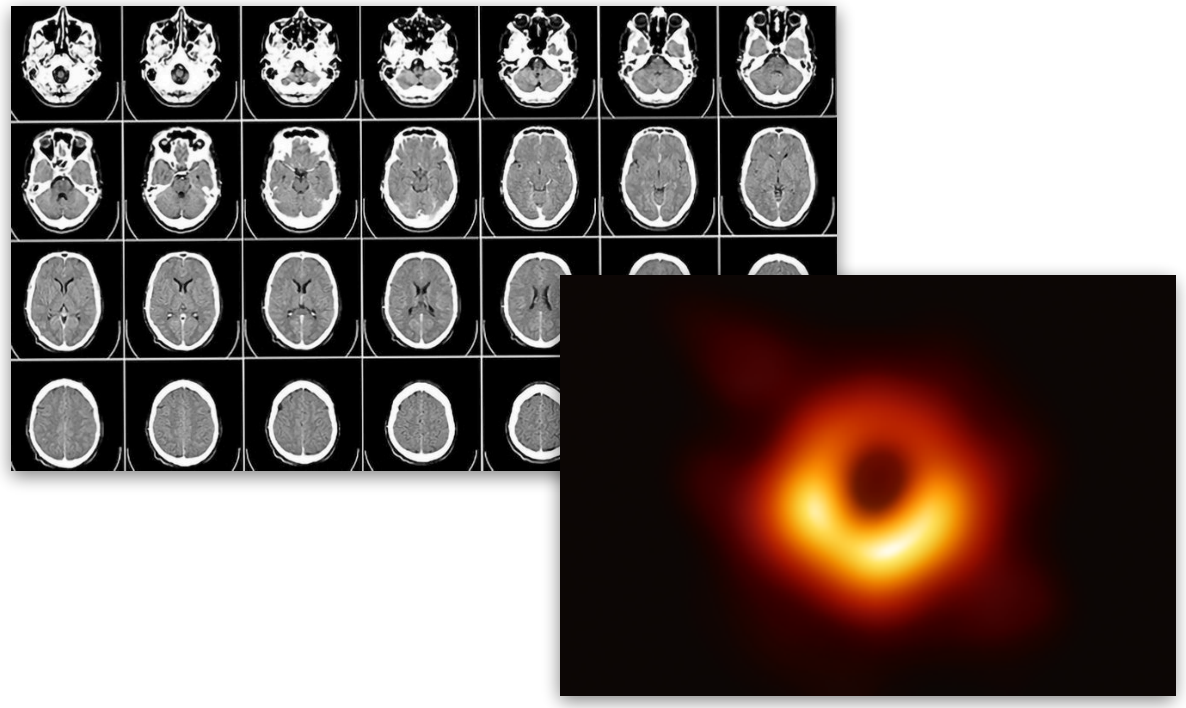
Solving IP: learning approach

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)



Consequence:

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

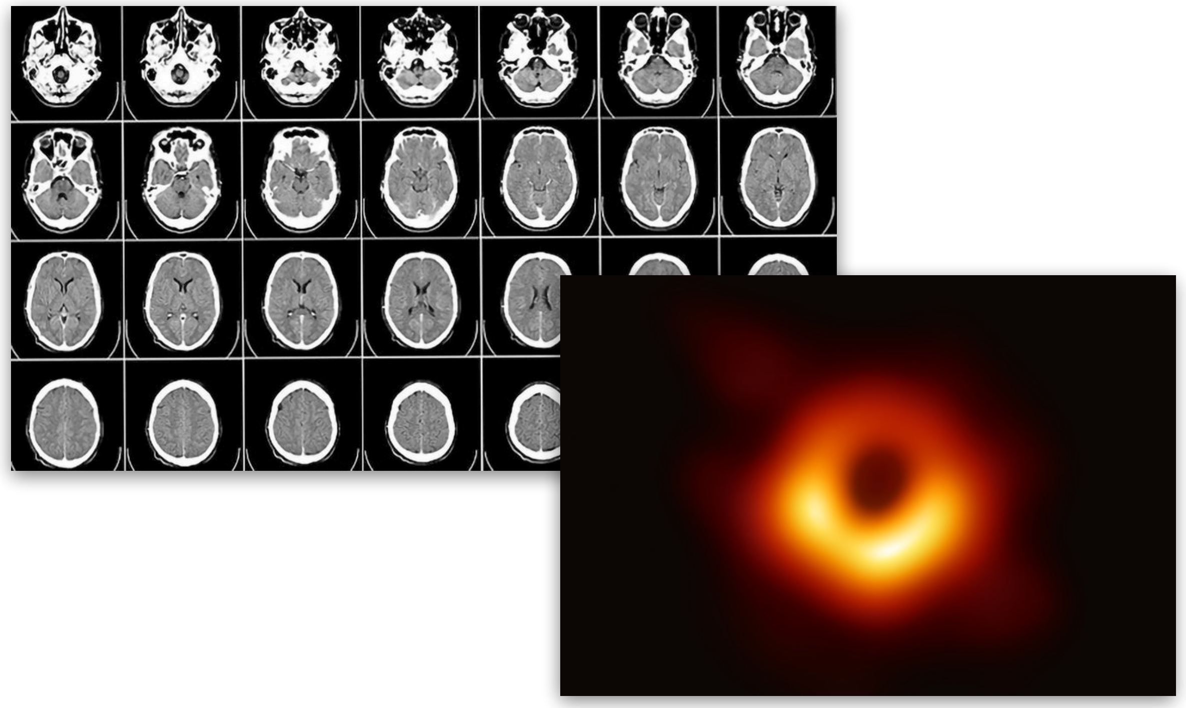
Solving IP: learning approach

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)

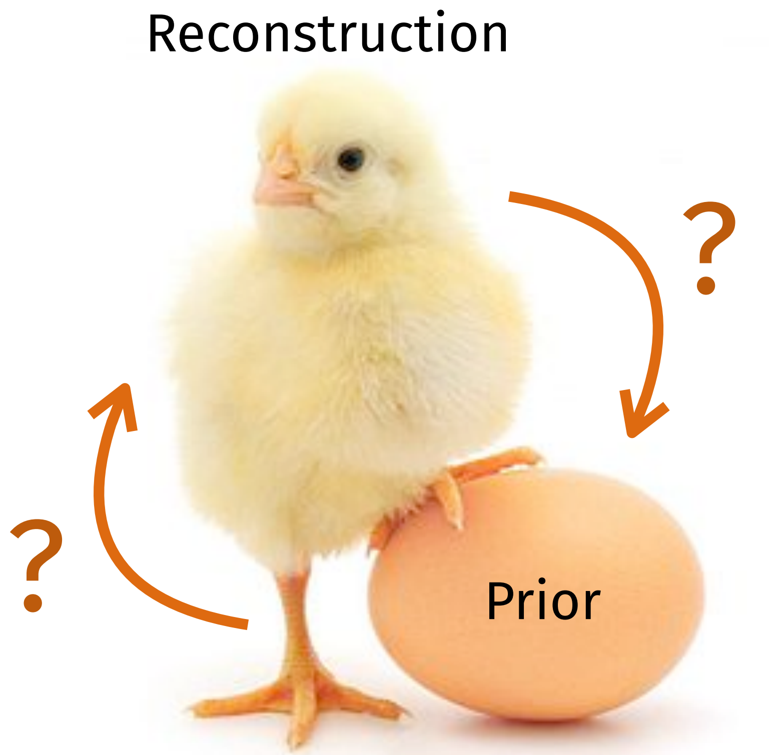


Consequence:

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

Conundrum:

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: $y = A(x) + \epsilon$ → **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]

Measurement-Driven Computational Imaging

Unsupervised context:

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

Linear inverse problems: $y = A(x) + \epsilon$ → **Yes**

If signal set $\mathscr{X}$ is low-dimensional, and, either multiple operators $\{A_i\}_{i=1}^G$,

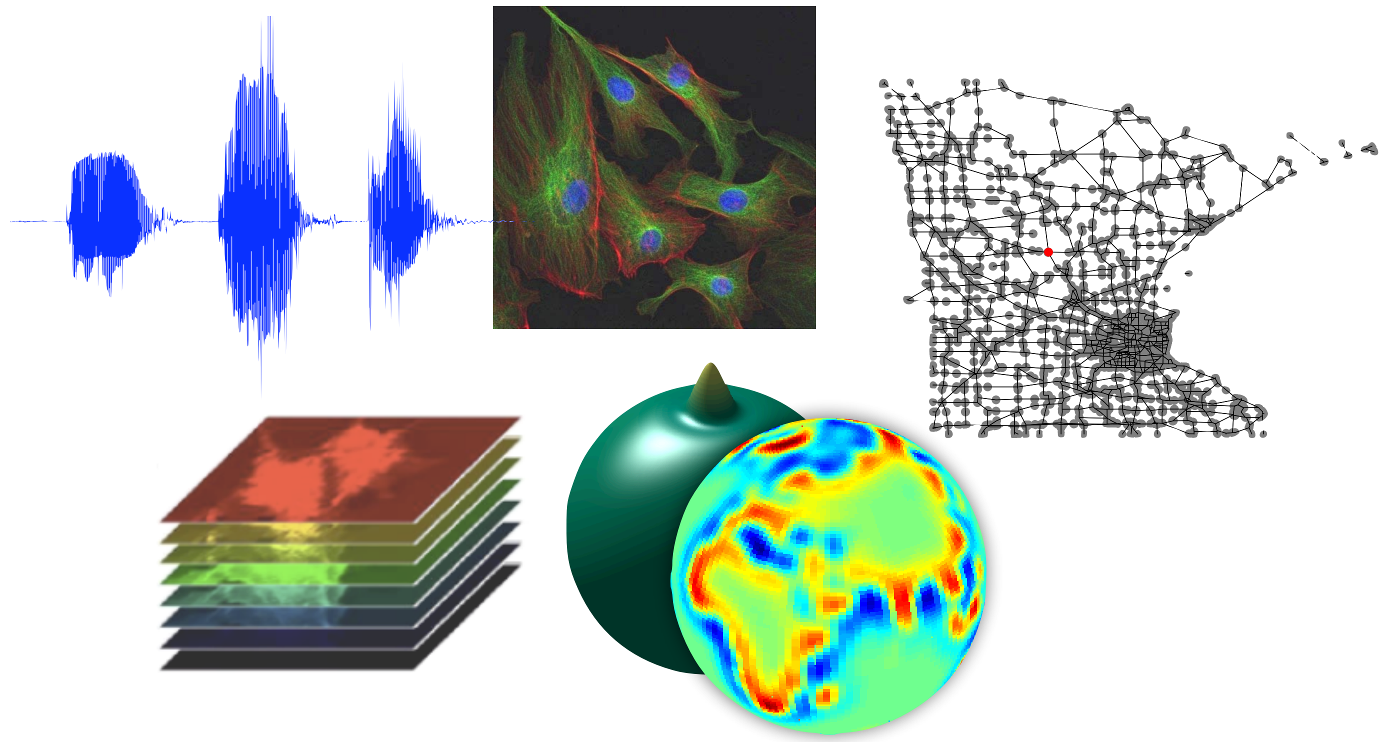
or $\mathscr{X}$ invariant to groups of transformations.

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

Non-linear inverse problems: $y = f \circ A(x) + \epsilon$ → *Today*

(with $f = \mathrm{sign}$, binary measurements)

Why binary measurements?

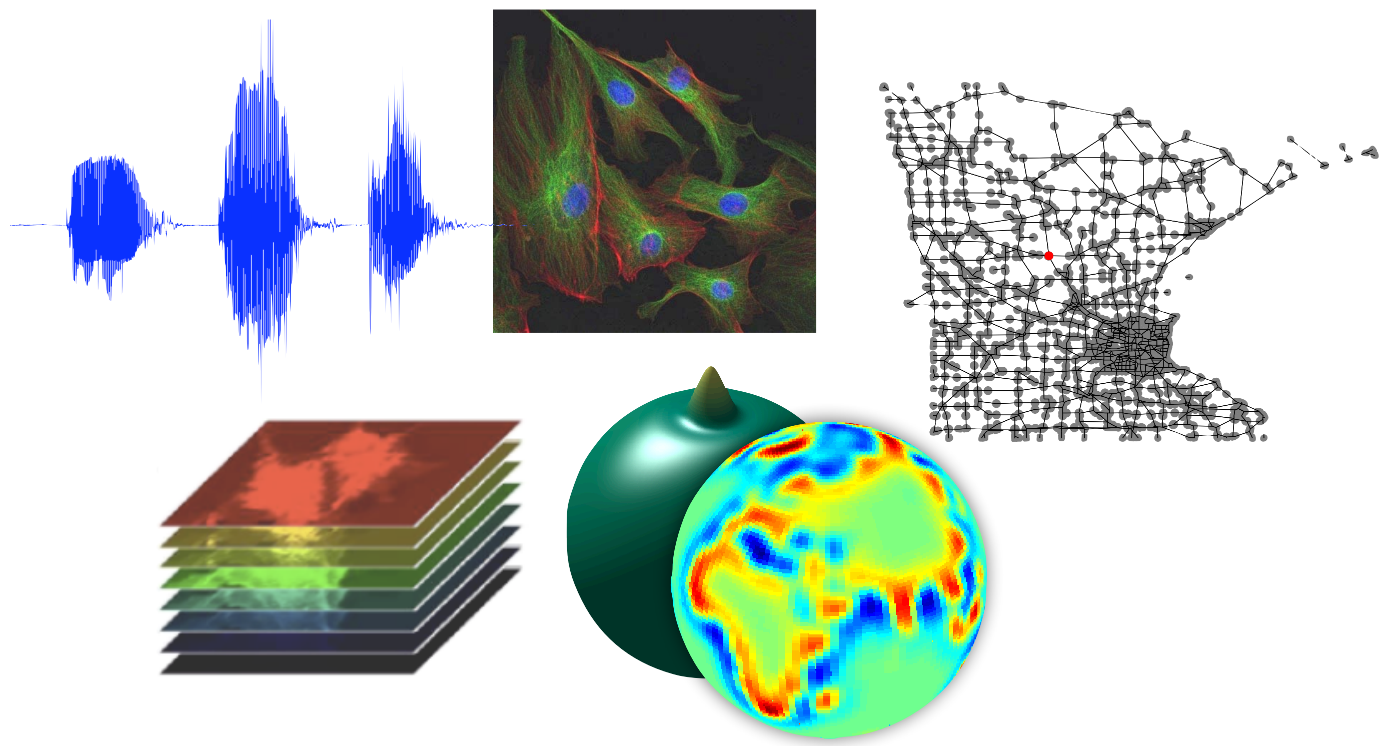


Sensing model Given $A = (a_1, \ldots, a_m)^\top, a_i \in \mathbb{R}^n$,

we observe a “signal” $x \in \mathbb{R}^n$ with m binary measurements :

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

Why binary measurements?



Sensing model Given $A = (a_1, \ldots, a_m)^\top, a_i \in \mathbb{R}^n$,

we observe a "signal" $x \in \mathbb{R}^n$ with m binary measurements :

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

Contexts

- Binary compressive sensing (1-bit CS): can we estimate x from y ?
- Binary/quantized dimensionality reduction: do $\mathrm{sign}(A\mathcal{S})$ capture the geometry of $\mathcal{S}$?
- Machine learning: 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

$$y_i = \mathrm{sign}(\underset{\substack{\downarrow \\ \text{known}}}{A_{g_i}} \underset{\substack{\downarrow \\ \text{unknown}}}{x_i}), \text{ with } 1 \leqslant i \leqslant N \text{ and } g_i \in \{1, \dots, G\}$$

Purpose of this talk

Learning to reconstruct from binary measurements?

Theoretical analysis: given N binary observations & G operators

$$y_i = \operatorname{sign}(\underset{\overset{\downarrow}{\text{known}}}{A_{g_i}}\,\underset{\overset{\downarrow}{\text{unknown}}}{x_i}), \text{ with } 1 \leqslant i \leqslant N \text{ and } g_i \in \{1, \dots, G\}$$

Estimate signal set $\mathcal{X} \supset \{x_i\}_{i=1}^{N}$? Error bounds (lower/upper)?
Sample Complexity (*i.e.*, N) bound?

Purpose of this talk

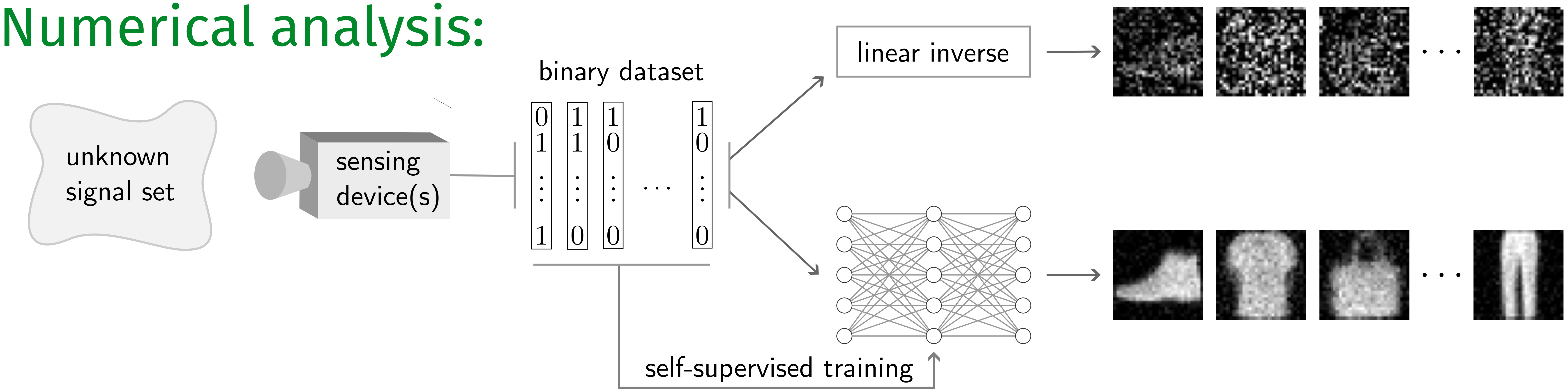
Learning to reconstruct from binary measurements?

Theoretical analysis: given N binary observations & G operators

$$y_i = \mathrm{sign}(\underset{\text{known}}{A_{g_i}}\,\underset{\text{unknown}}{x_i}), \text{ with } 1 \leqslant i \leqslant N \text{ and } g_i \in \{1, \dots, G\}$$

Estimate signal set $\mathcal{X} \supset \{x_i\}_{i=1}^{N}$? Error bounds (lower/upper)?
Sample Complexity (*i.e.*, N) bound?

Numerical analysis:



unknown signal set

sensing device(s)

binary dataset

$\begin{bmatrix}0\\1\\\vdots\\1\end{bmatrix} \begin{bmatrix}1\\1\\\vdots\\0\end{bmatrix} \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} \cdots \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}$

linear inverse

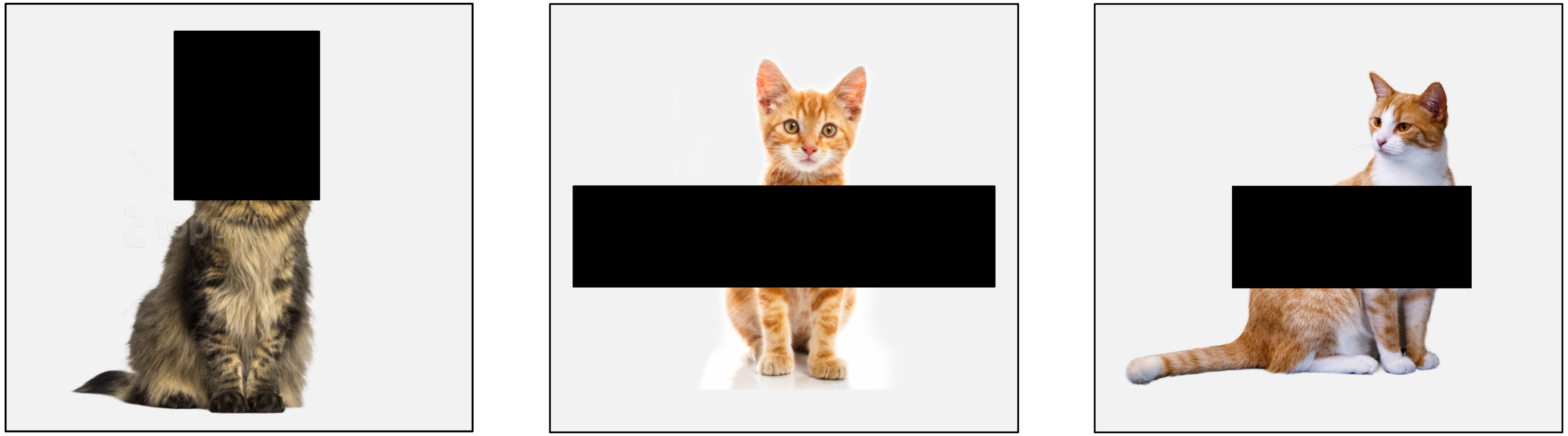
self-supervised training

Cost function? Network architecture? Comparison to linear case?

Sensing scenario 1: Multiple Operators

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

A_1x_1 A_2x_2 A_3x_3

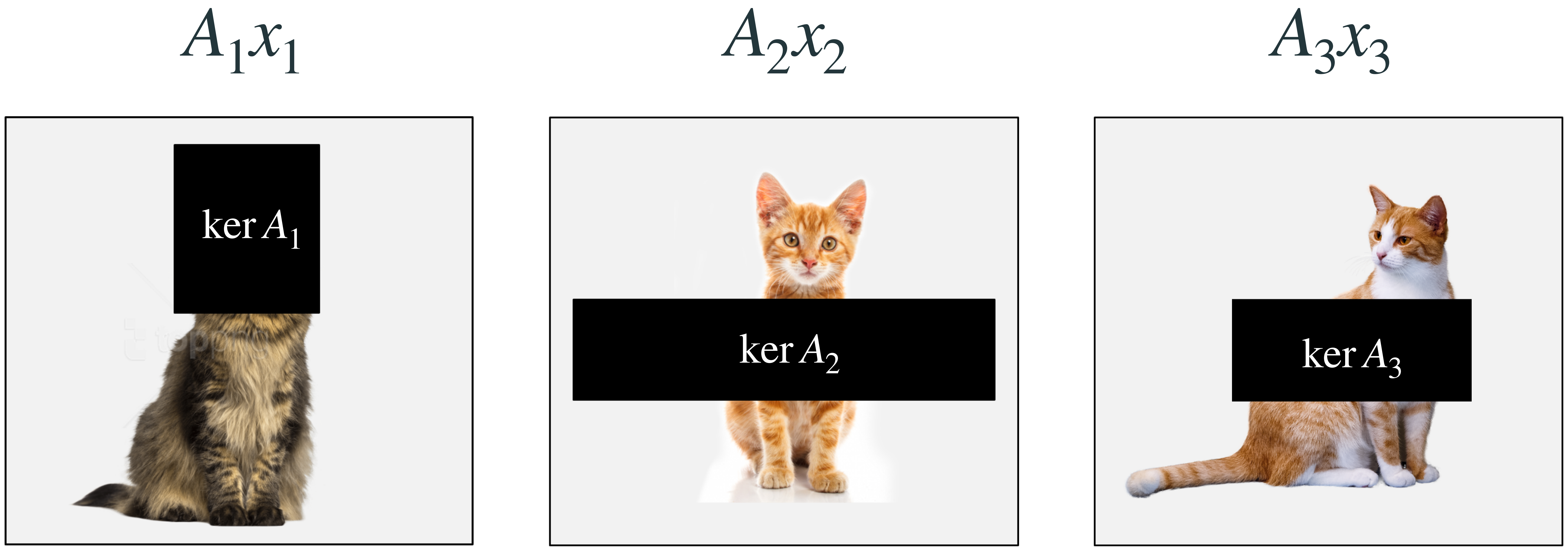


Examples:

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

Sensing scenario 1: Multiple Operators

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



Examples:

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

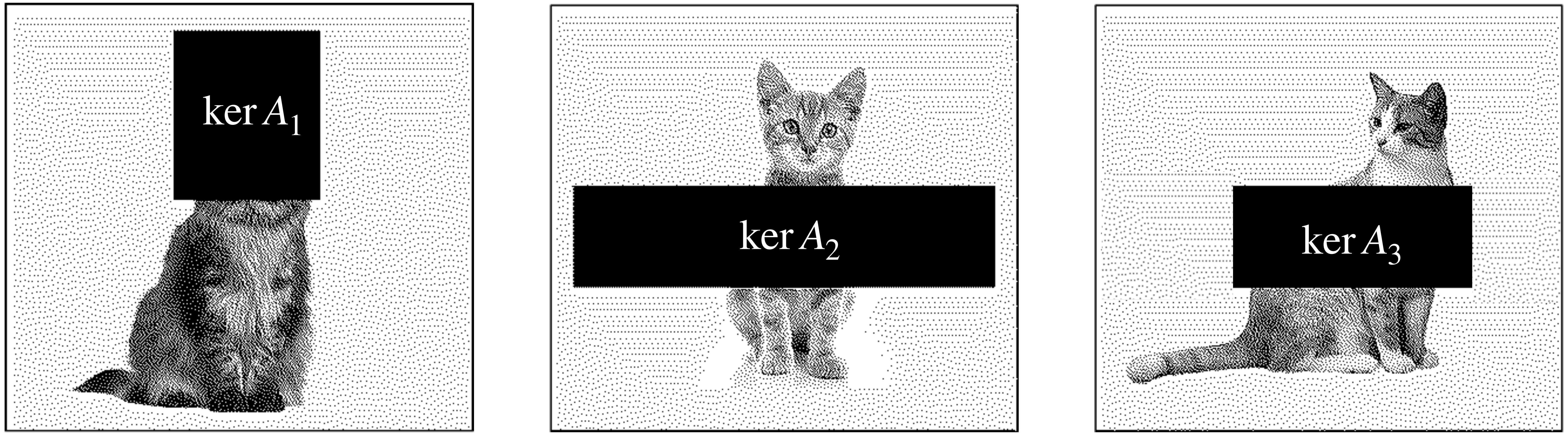
Principle (linear case):

Learning $\mathscr{X}$ 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

$\mathrm{sign}(A_1 x_1)$ $\mathrm{sign}(A_2 x_2)$ $\mathrm{sign}(A_3 x_3)$



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

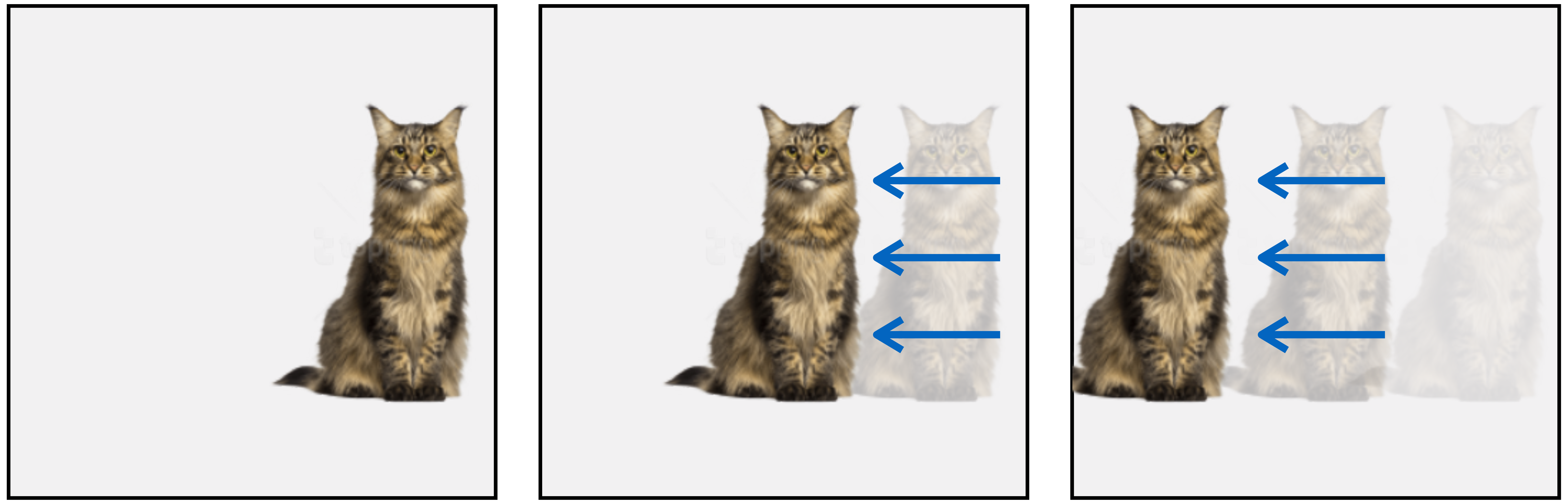
Most signals sets are invariant to groups of transformations:

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

(see S. Mataigne's seminar ;-))

(geometric prior)

Example: translation



Sensing scenario 2: Single operator & invariance

Most signals sets are invariant to groups of transformations:

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

(see S. Mataigne's seminar ;-))

(geometric prior)

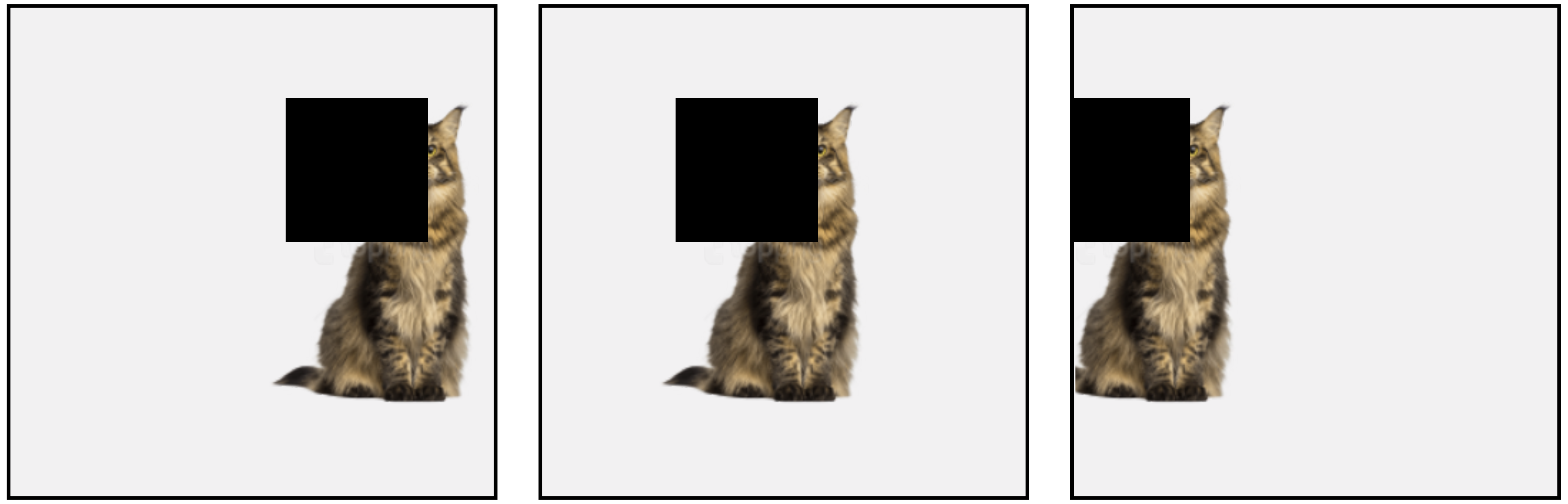
For all $g \in \{1, \dots, G\}$ we have

(linear case)

$$y = \quad Ax \quad = \quad \underset{A_g}{\underset{\downarrow}{AT_g}} \, \underset{x'}{\underset{\downarrow}{T_g^{-1} x}} \quad = \quad A_g x'$$

Implicit access to multiple operators
$A_g = AT_g$

AT_g for different g



Necessary condition

A is not equivariant:

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

for some $\tilde{T}_g$.

Do you see why?

Sensing scenario 2: Single operator & invariance

Most signals sets are invariant to groups of transformations:

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

(see S. Mataigne's seminar ;-))

(geometric prior)

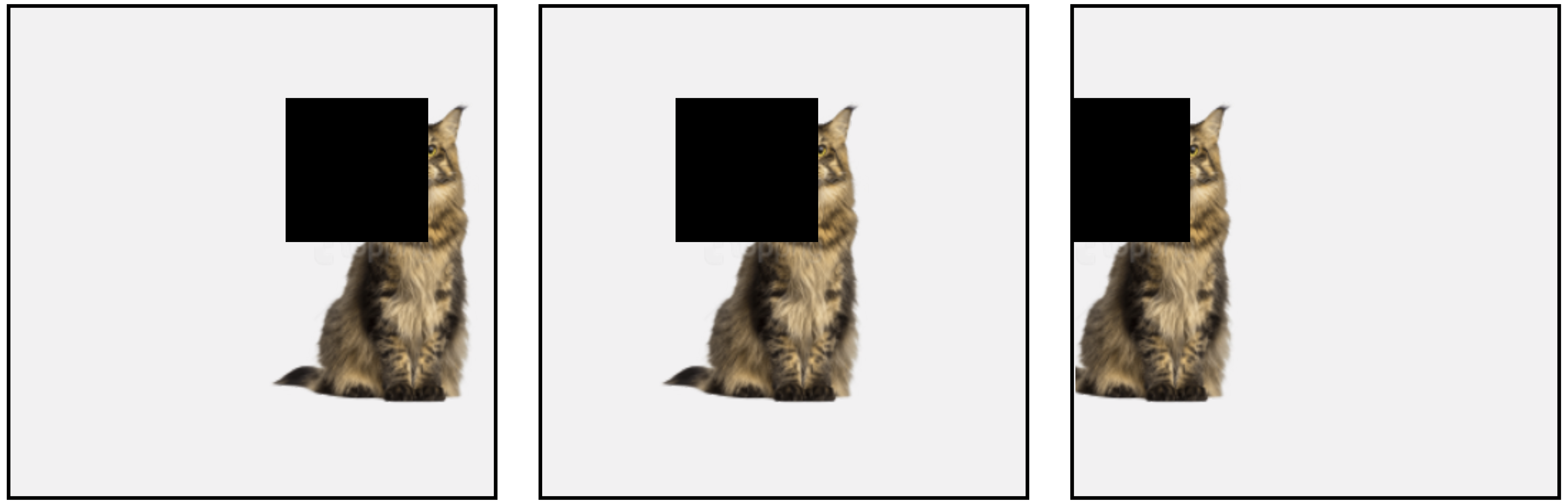
For all $g \in \{1, \ldots, G\}$ we have

(binary case)

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

Implicit access to multiple operators $A_g = AT_g$

AT_g for different g



Necessary condition

A is not equivariant:

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

for some $\tilde{T}_g$.

Do you see why?

Model identification: the problem

Assumption: enough points of $\mathscr{X}$ have been observed for all operators.

∞ for now

(More on this later)

Model identification: the problem

Assumption: enough points of $\mathscr{X}$ have been observed for all operators.

∞ for now

(More on this later)

Question: Given the observed sets

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

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

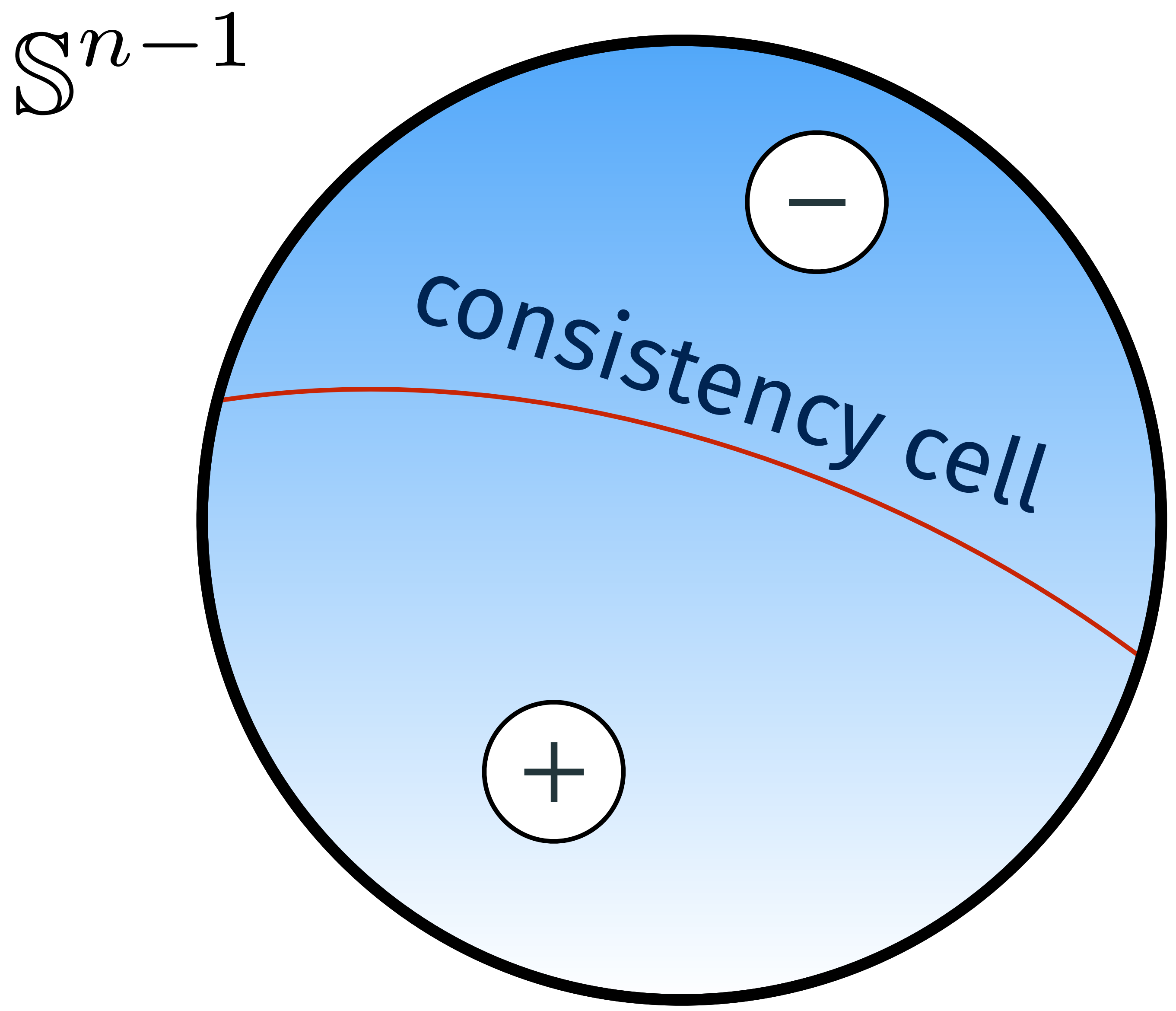
meaning?

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

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

Model identification: geometric intuition

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



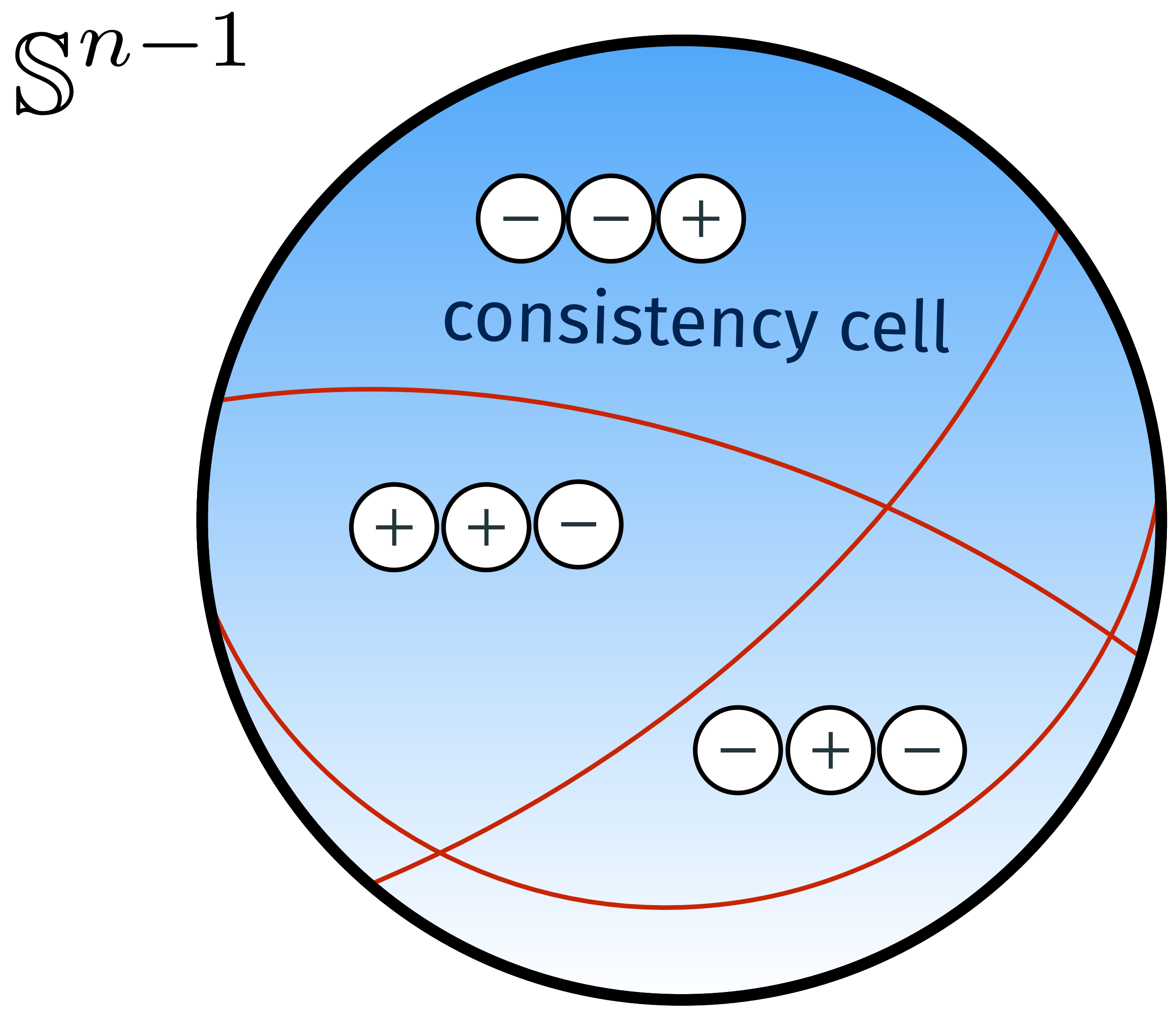
$m = 1$

$\operatorname{sign}(A_g\,\cdot)$ 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 : $(A_g)_{ij} \sim_{\text{iid}} \mathcal{N}(0,1)$



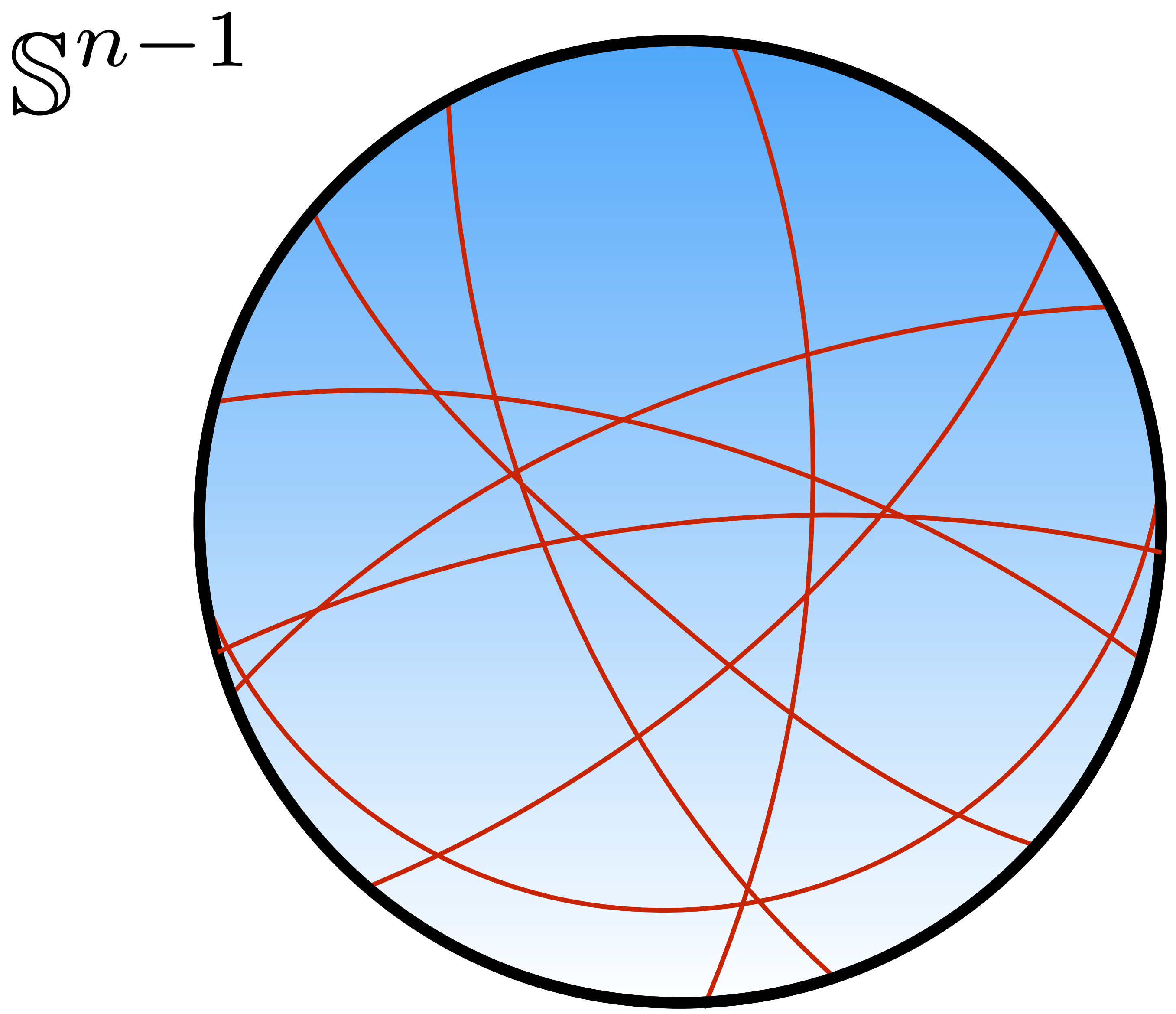
$m = 3$

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

Growing number of consistency cells as m ↑
(bounded by 2^m)

Model identification: geometric intuition

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



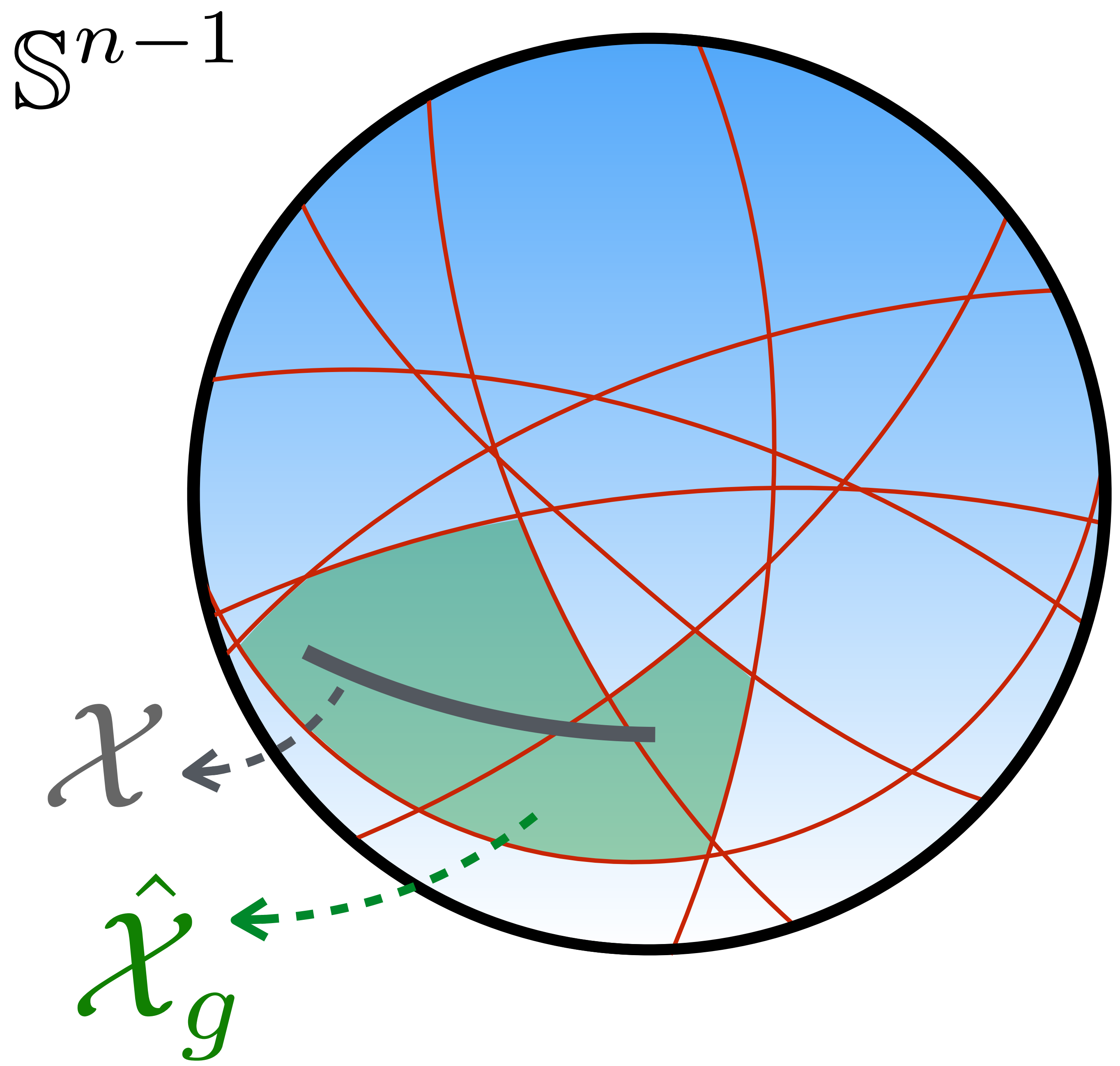
$m = 8$

$\mathrm{sign}(A_g \,\cdot)$ 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 : $(A_g)_{ij} \sim_{\text{iid}} \mathcal{N}(0,1)$



$\mathrm{sign}(A_g\,\cdot)$ tessellates $\mathbb{S}^{n-1}$

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

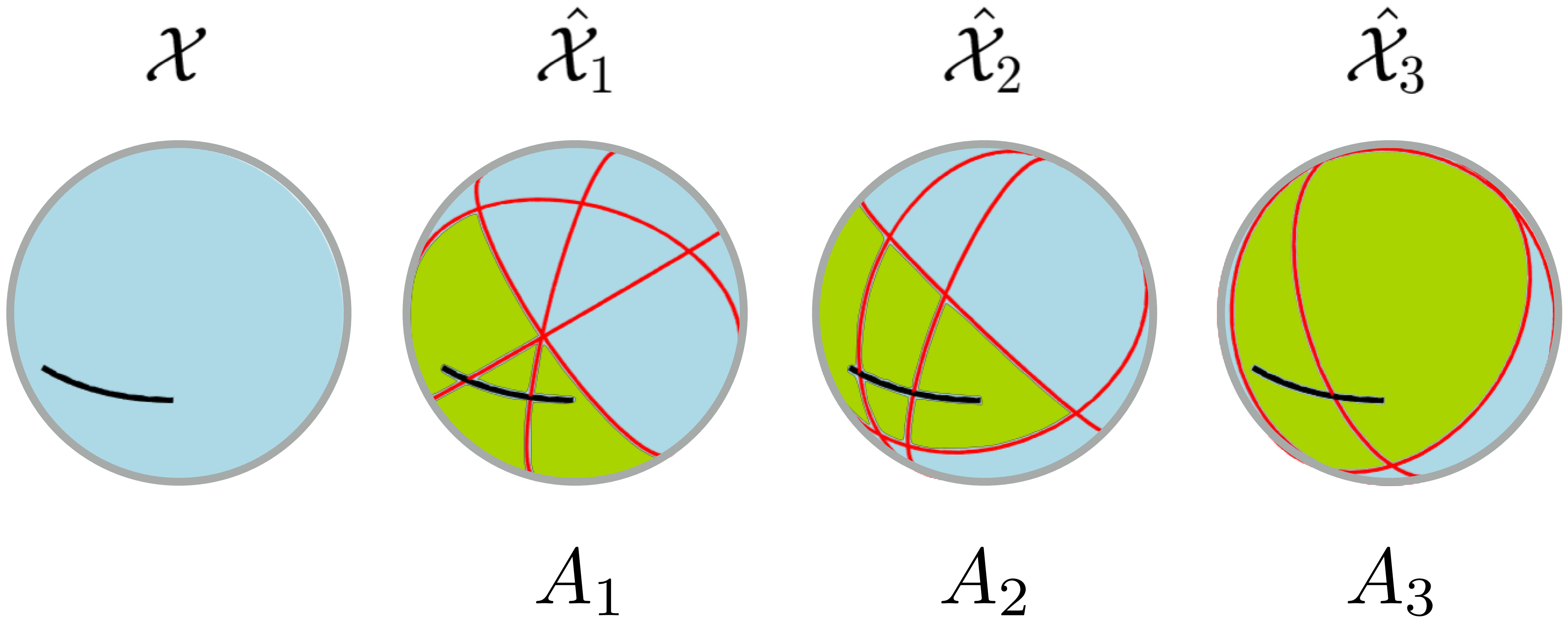
Let us define the biggest set binary consistent (wrt A_g) with $\mathcal{X}$:

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

→ dilation of $\mathcal{X}$ by the "uncertainty" of $\mathrm{sign} \circ A_g$

Model identification: geometric intuition

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

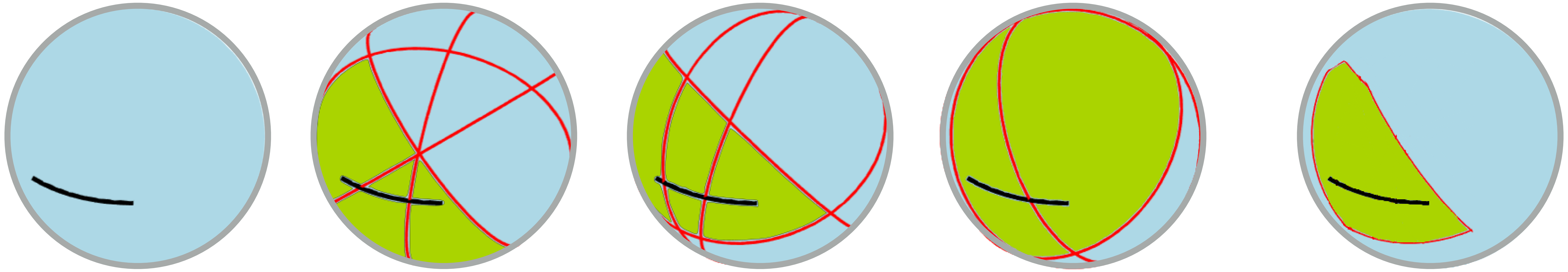


Different dilations of $\mathscr{X}$

Model identification: geometric intuition

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

$\mathcal{X}$ $\hat{\mathcal{X}}_1$ $\hat{\mathcal{X}}_2$ $\hat{\mathcal{X}}_3$ $\hat{\mathcal{X}} := \cap_{g=1}^{G} \hat{\mathcal{X}}_g$



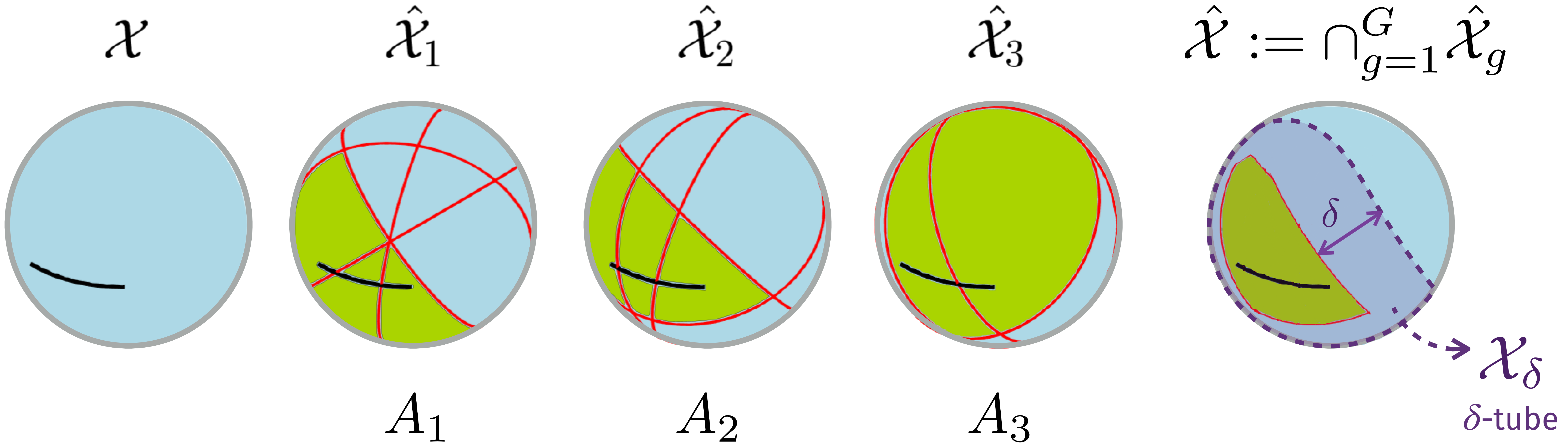
A_1 A_2 A_3 all As

Different dilations of $\mathscr{X}$

Intersected dilations

Model identification: geometric intuition

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



***Identification error* (definition)**

Identify signal set up to global error δ ➔ $\hat{\mathscr{X}}$ is in a δ-tube $\mathscr{X}_\delta$

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

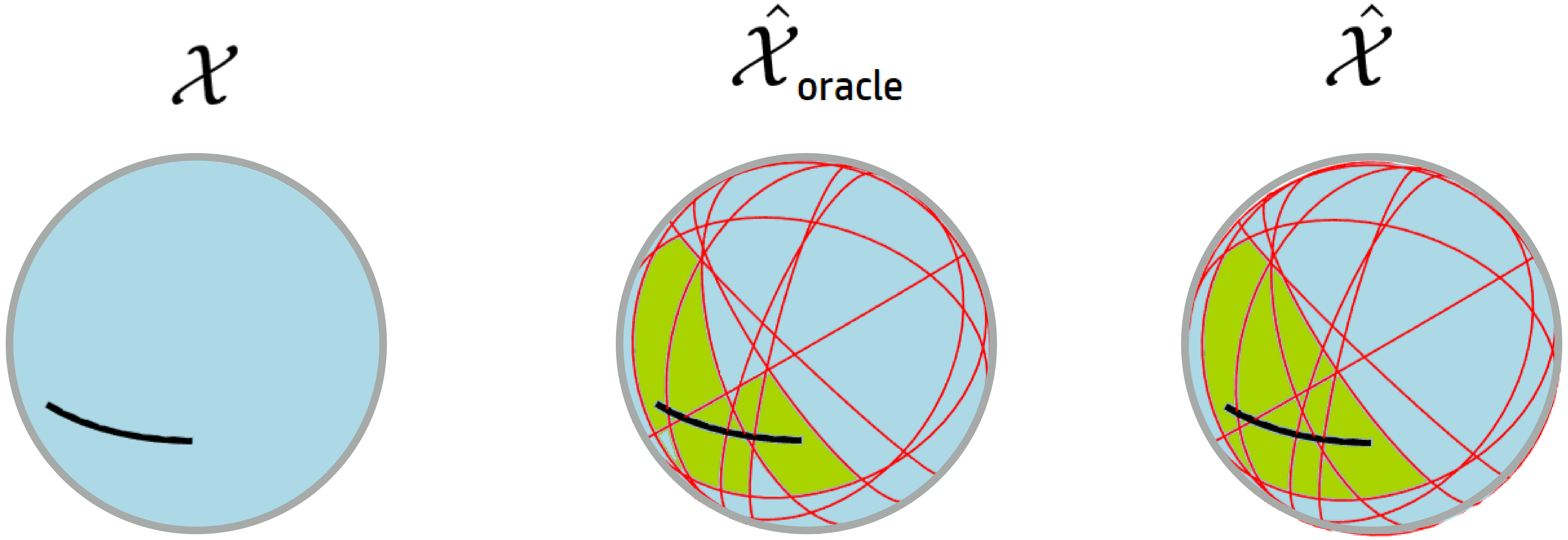
Upper/Lower bound on δ ?   Sample complexity?

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

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

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

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



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

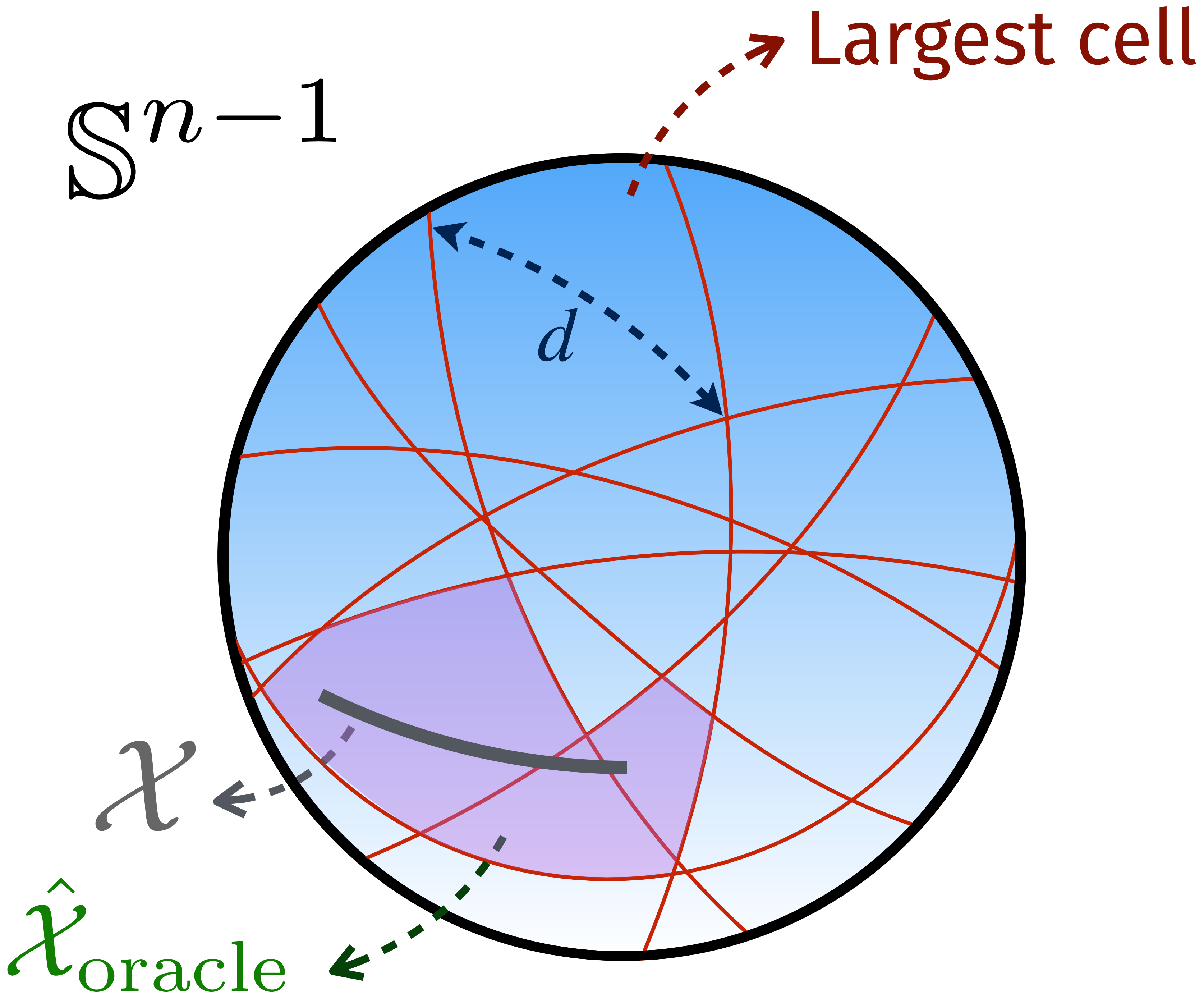
Question: smallest δ such that $\mathscr{X}_{\text{oracle}} \subset \mathscr{X}_{\delta}$? *(whatever $\mathscr{X}$'s orientation)*

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

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

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

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

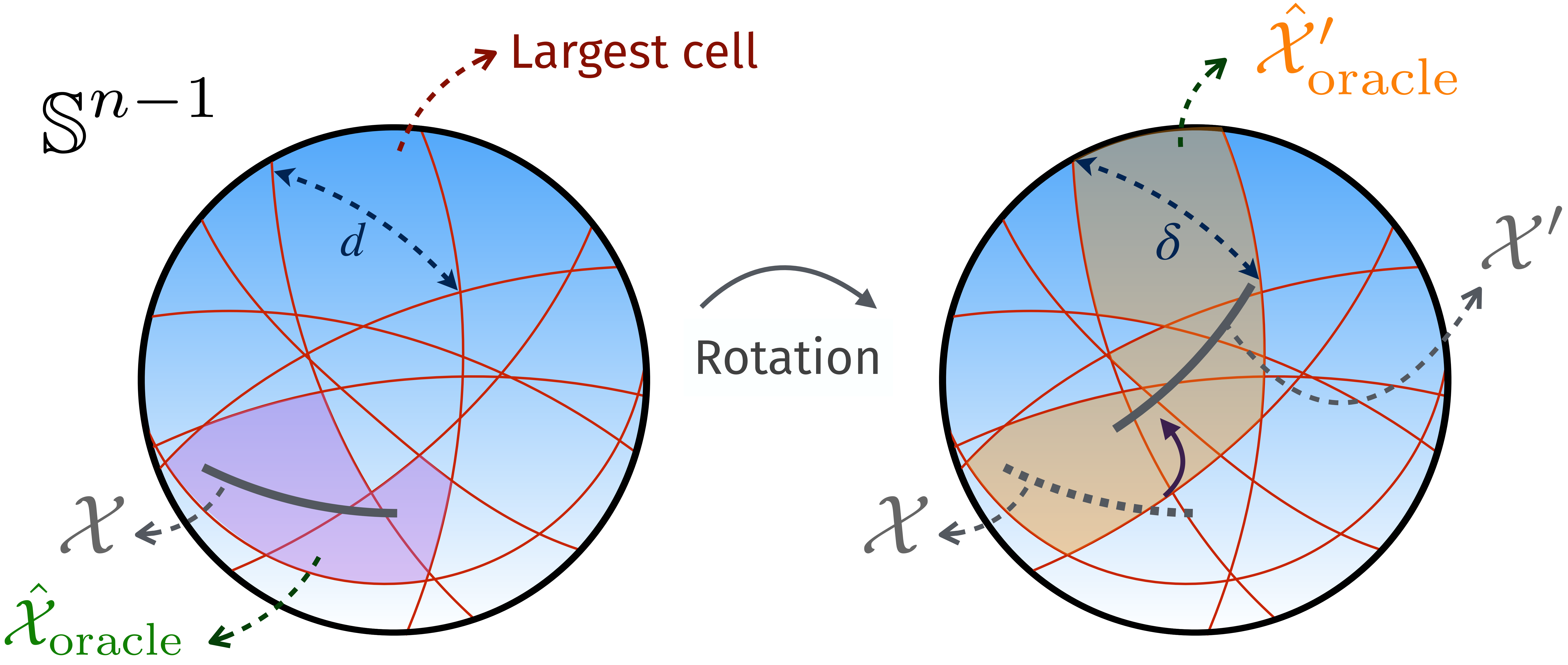


Lower bound on δ (via an *oracle* standpoint)

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

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

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



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.*

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

Consequences: we can show the following

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

→ Model identification error is trivially large

> *Proof*: Given $x = \bar{A}^\top u$, for some u, define $x_\pm := \frac{x \pm v}{\|x \pm v\|}$ with unit vector $v \in \ker \bar{A} (\neq 0)$,
>
> We have:
>
> - $\|x_\pm\|^2 = \|x\|^2 + \|v\|^2 = 1 + \|x\|^2$
> - $\operatorname{sign} \bar{A}x_+ = \operatorname{sign} \bar{A}x_- = \operatorname{sign} \bar{A}x$ → x_+, x_- are unit vectors in the same cell
>
> Moreover,
>
> $$\operatorname{dist}(x_+, x_-) = \frac{2\|v\|}{\sqrt{\|x\| + \|v\|}} = \frac{2}{\sqrt{1 + \|x\|}} \quad (\to 2, \text{ if } \|x\| \to 0)$$

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

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

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

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

Consequences: we can show the following

1. If $\operatorname{rank}(\bar{A}) < n$, ∃ 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 $\mathscr{X}$ with $G = 1$ (w/o invariance)

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

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

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

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

Consequences: we can show the following

1. If $\mathrm{rank}(\bar{A}) < n$, ∃ 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 $\mathscr{X}$ with $G = 1$ (w/o invariance)

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

   → $\boxed{\delta \text{ cannot decrease faster than } \propto m^{-1}G^{-1}}$

Thao & Vetterli (1996, Theorem A.7)

$$|\{\mathrm{sign}\,(\bar{A}\mathcal{X})\}| \leqslant \binom{mG}{n} 2^n$$

Upper bound on δ (with the help of randomness)

Definition: $$\mathrm{boxdim}(S) = \lim\sup_{\epsilon \to 0^+} \frac{\log \mathfrak{N}(S, \epsilon)}{\log 1/\epsilon}$$

Assumption: The signal set $\mathscr{X}$ is low-dimensional

$\leftrightarrow \mathscr{X}$ has box-counting dimension $k \ll n$

Examples: sparse dictionaries, manifold models, etc.

Upper bound on δ (with the help of randomness)

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

Assumption: The signal set $\mathscr{X}$ is low-dimensional

$\leftrightarrow$ $\mathscr{X}$ has box-counting dimension $k \ll n$

Examples: sparse dictionaries, manifold models, etc.

Theorem. *If* $\operatorname{boxdim}(\mathcal{X}) < k$ *and* $\overbrace{A_1, \ldots, A_G}^{\text{distinct operators}} \in \mathbb{R}^{m \times n}$ *have i.i.d. Gaussian entries with*

$$m > \frac{4}{\delta}\left(k + \frac{n}{G}\right) \log\left(\frac{5\sqrt{n}}{m}\right) + \frac{1}{G} \log \frac{1}{\xi} + \frac{n}{G} \log 3,$$

$\leftrightarrow$ $m \simeq (k + \frac{n}{G})$

then

$$\mathbb{P}(\hat{\mathcal{X}} \subseteq \mathcal{X}_\delta) \geq 1 - \xi.$$

Upper bound on δ (with the help of randomness)

Consequences of this theorem:

- → The identification error of $\mathscr{X}$ decreases as

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

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

Upper bound on δ (with the help of randomness)

Consequences of this theorem:

→ The identification error of $\mathcal{X}$ decreases as

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

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

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

$\leftrightarrow$ Any consistent decoder f s.t.

$$y \in \{\pm 1\}^m \mapsto f(y) \in \{x \in \mathbb{S}^{n-1} | \operatorname{sign}(Ax) = y \text{ and } x \in \mathcal{X}\}$$

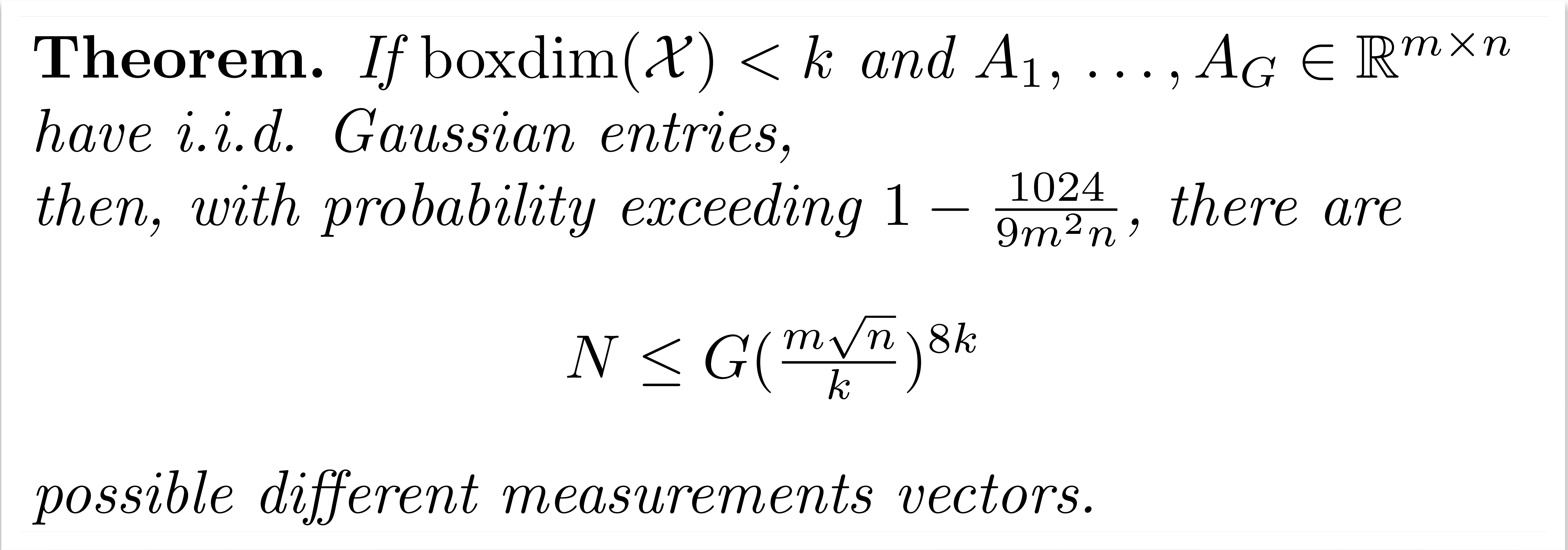
has an error $\max_{x \in \mathcal{X}} \|x - f\big(\operatorname{sign}(Ax)\big)\| = O\big(\frac{k}{m}\big)$

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

Sample Complexity bound (again with randomness)

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

→ *an upper bound on* N *is* $\bigcup_{g=1}^{G} |\operatorname{sign}(A_g \mathscr{X})|$



Theorem. *If* $\operatorname{boxdim}(\mathcal{X}) < k$ *and* $A_1, \ldots, A_G \in \mathbb{R}^{m \times n}$ *have i.i.d. Gaussian entries, then, with probability exceeding* $1 - \frac{1024}{9m^2 n}$, *there are*

$$N \leq G(\tfrac{m\sqrt{n}}{k})^{8k}$$

possible different measurements vectors.

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

Algorithms

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 $\mathscr{L}$ which uses $\{(y_i, A_{g_i})\}_{i=1}^N$

Warning:

No clear link with the theory (yet)

Multi-operator case

Self-supervised training loss: given a reconstruction model f_θ

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

with:

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

(Logistic loss)

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

Problem: the function $f_\theta(y, A_g) = A_g^\dagger y := A_g^\top (A_g A_g^\top)^{-1} y$ is a consistent reconstruction

→ 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

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

with:

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

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

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

→ 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

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

with:

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

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

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

→ 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)

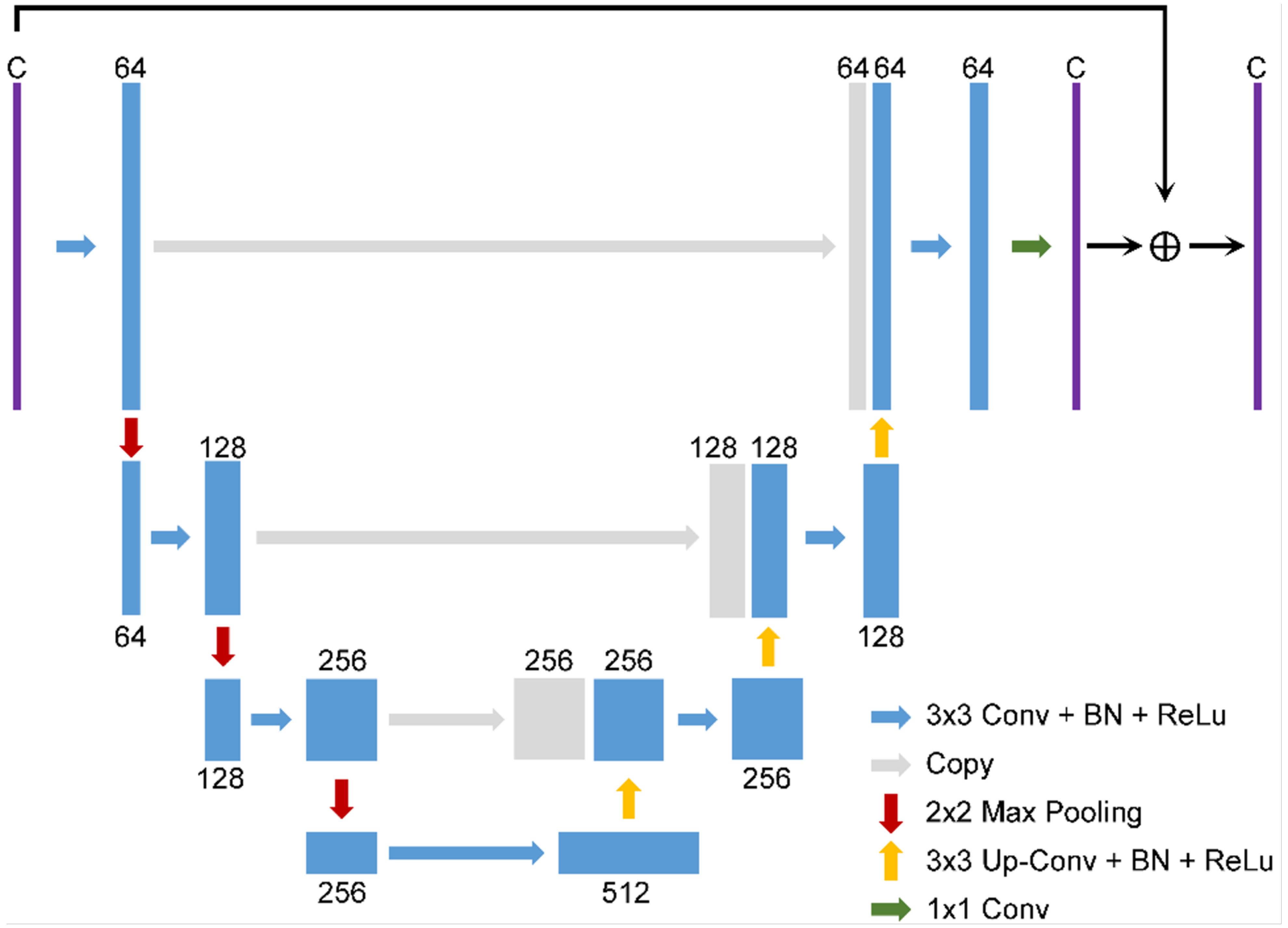
Experiments

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

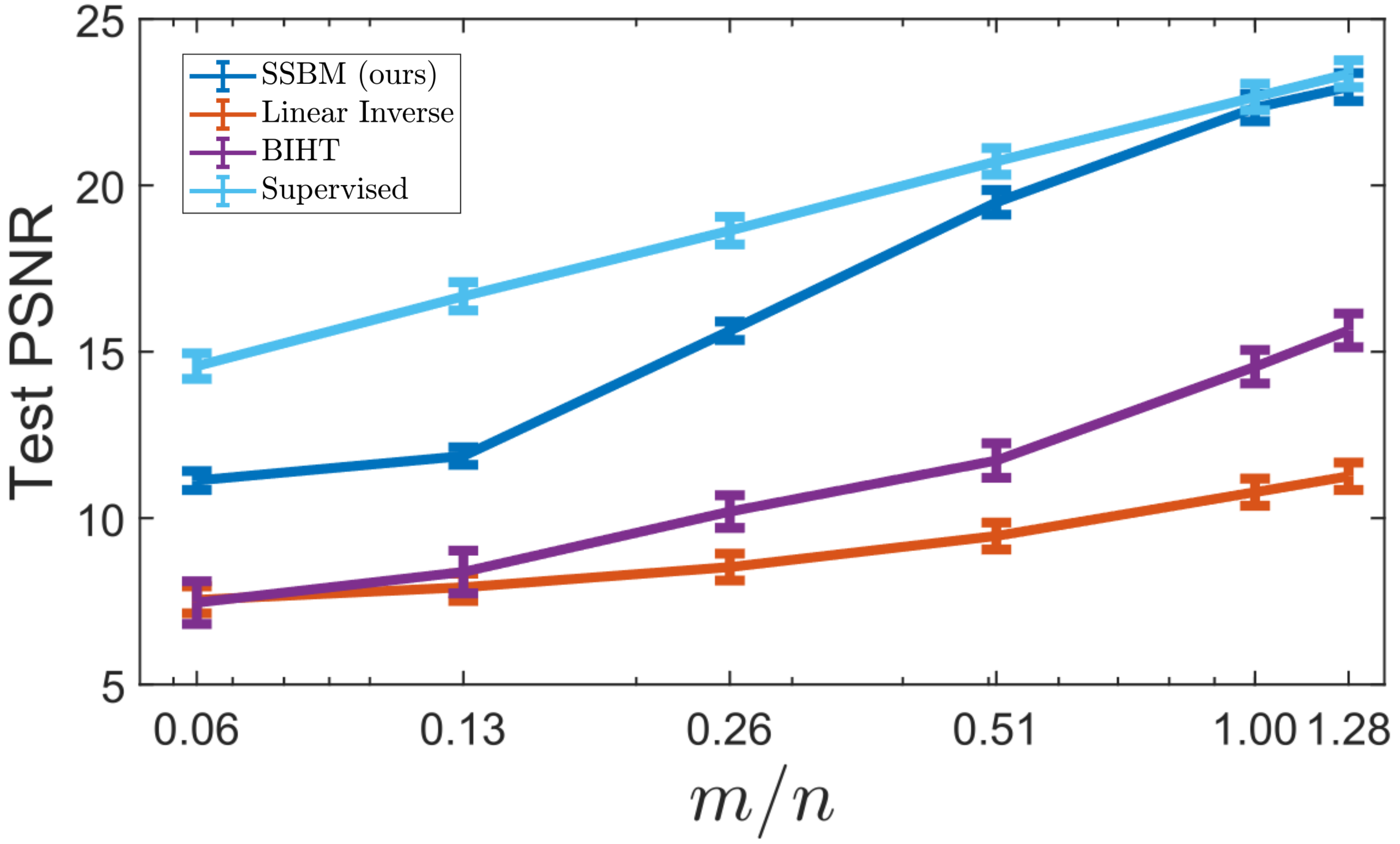


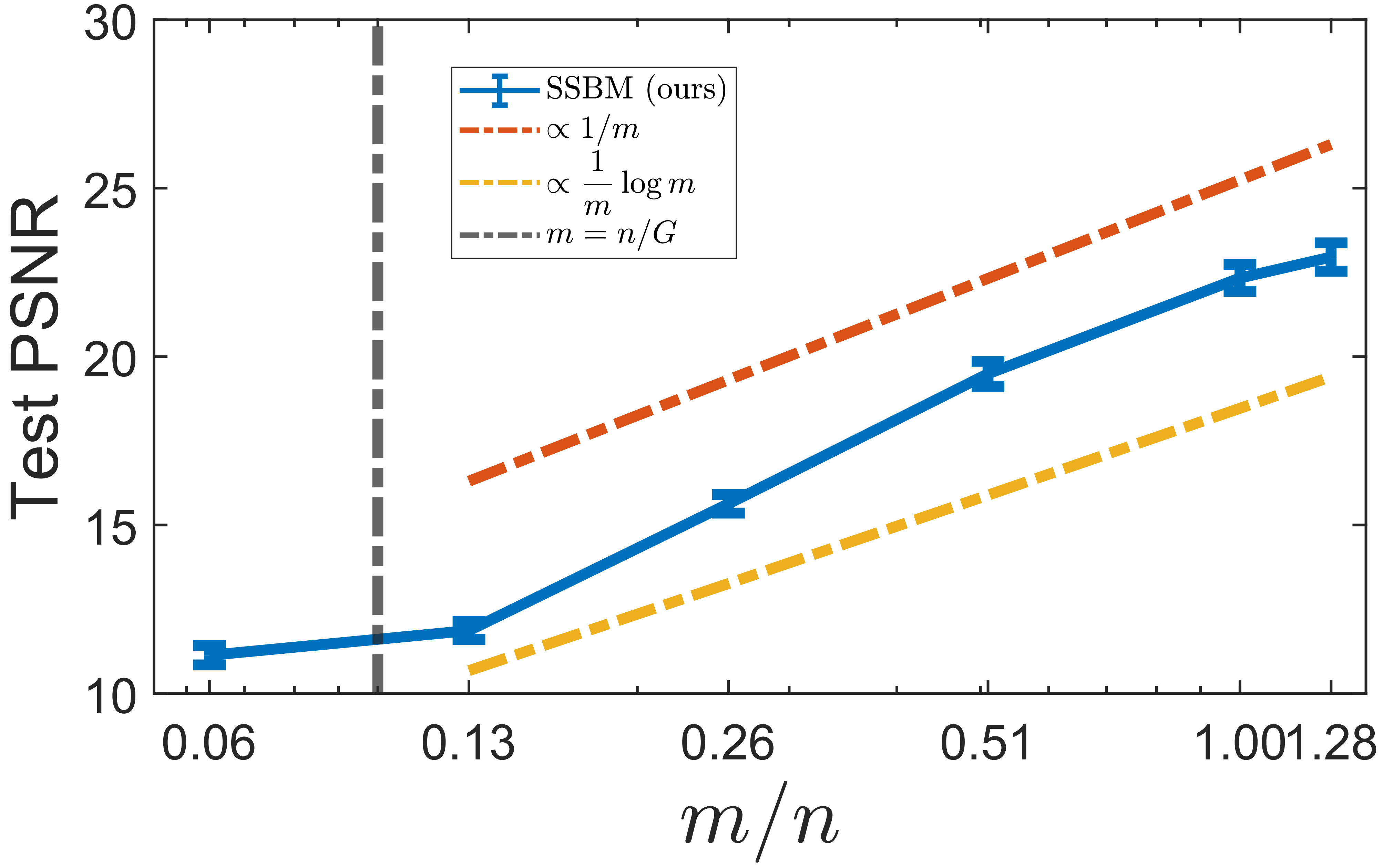
Comparison with

- 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.

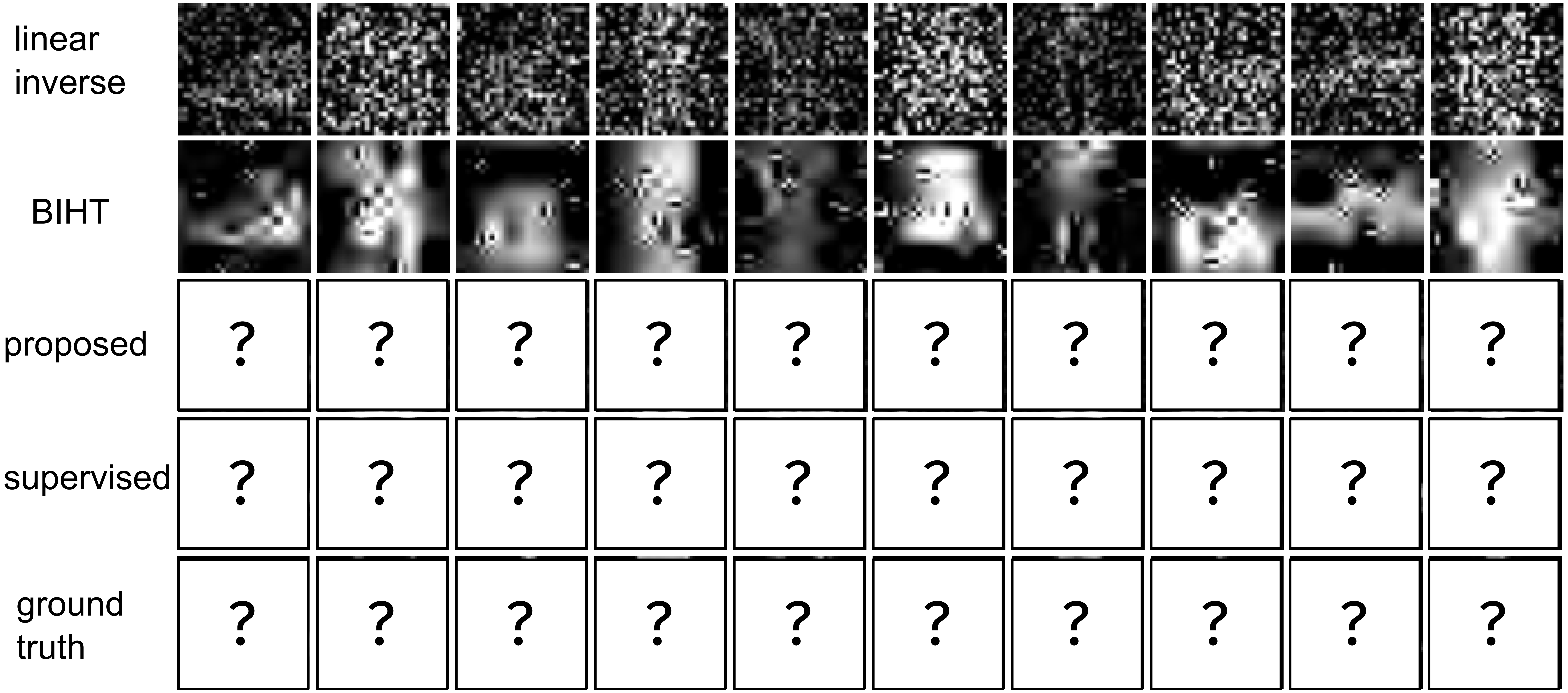




$$\text{Test PSNR} := \frac{1}{N'} \sum_{i=1}^{N'} \text{PSNR}\left(x_i', f_\theta\left(\text{sign}\left(A_{g_i} x_i'\right), A_{g_i}\right)\right),\ x_i' \in \text{“test set”}$$

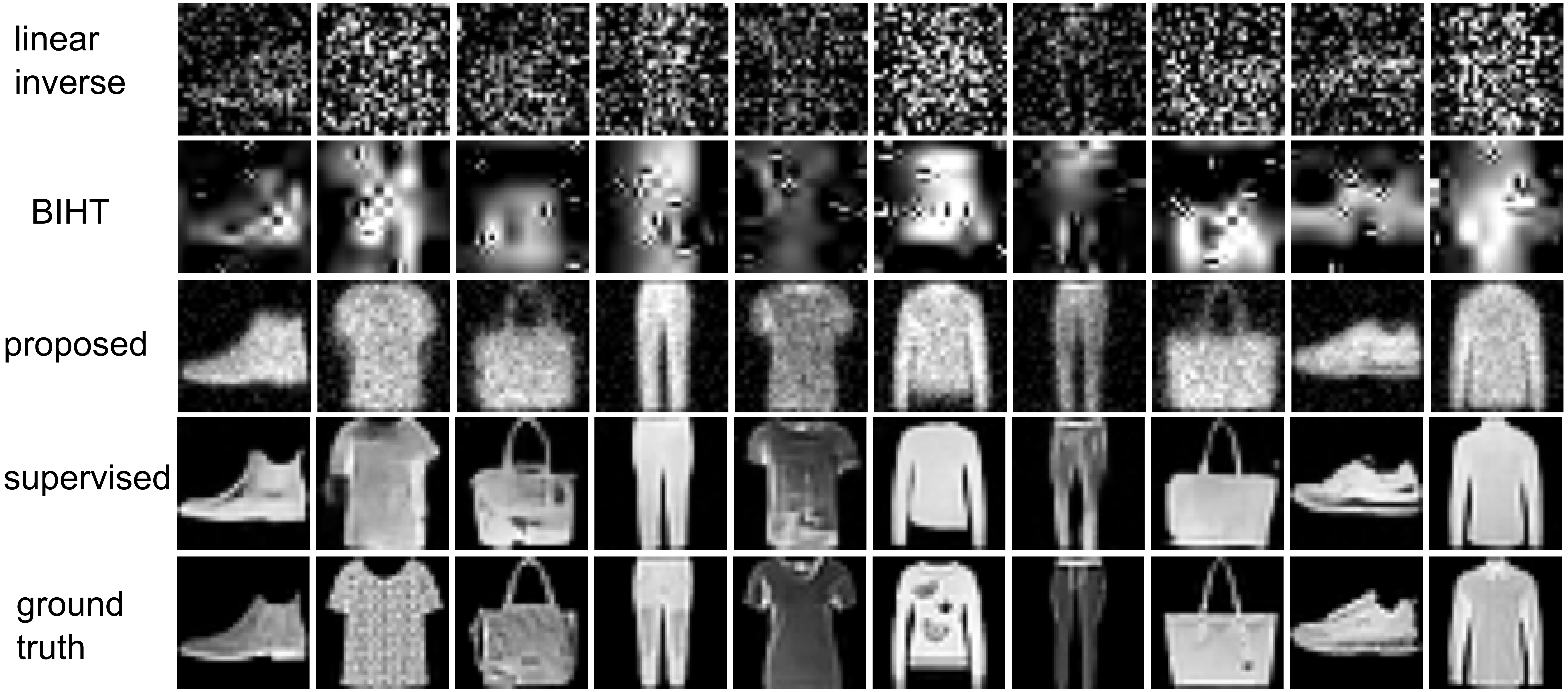
Fashion MNIST dataset

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



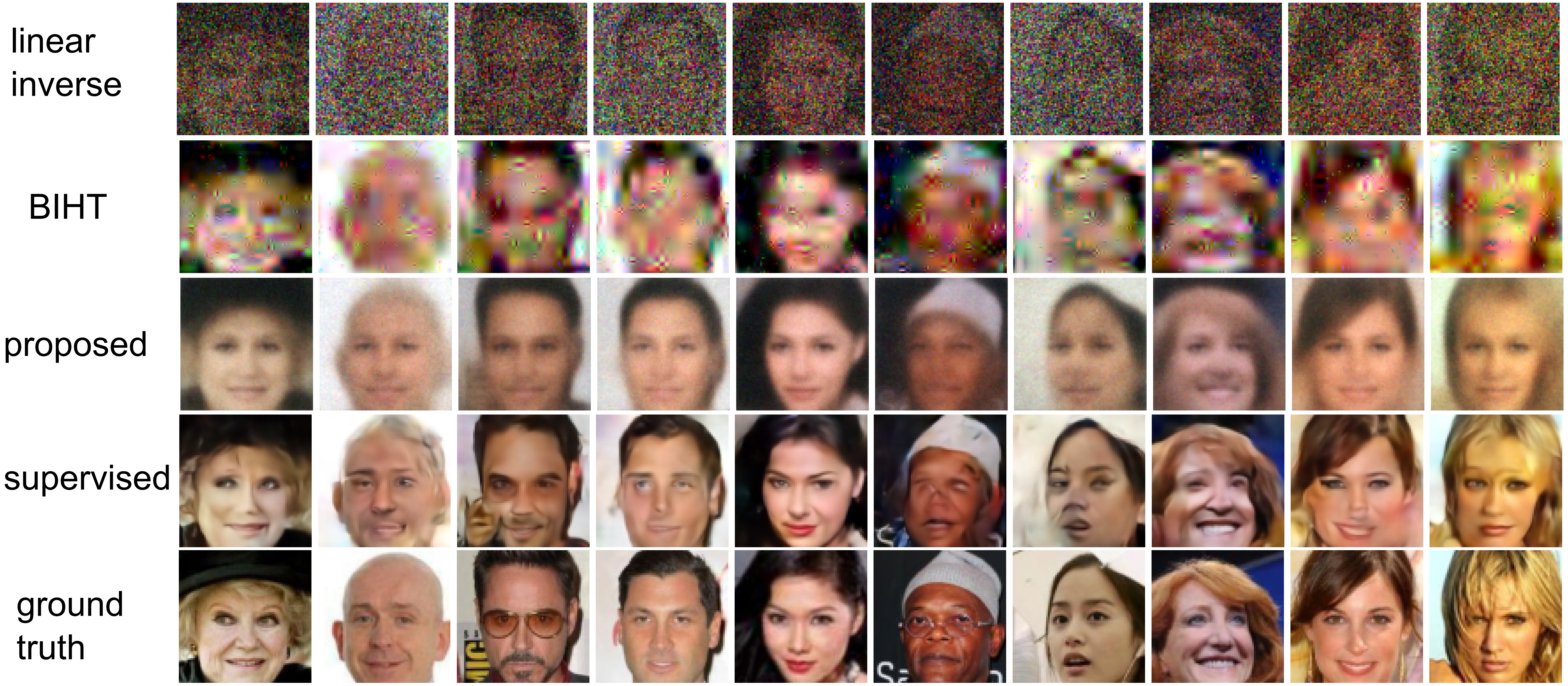
Fashion MNIST dataset

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 $\mathscr{X}$, 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