Learning to Reconstruct Signals From Binary Measurements

Julián Tachella
CNRS, Physics Laboratory
ENS Lyon, France

Laurent Jacques
INMA/ICTEAM
UCLouvain, Belgium

SAMPTA 23

Inverse problems (IP)

Measurements in $\mathbb{R}^m$

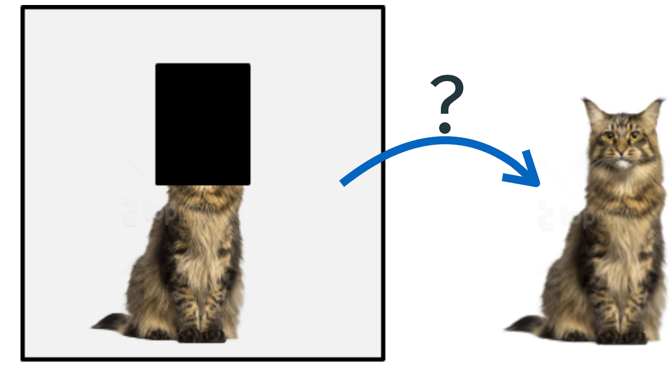
Measurements operator

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

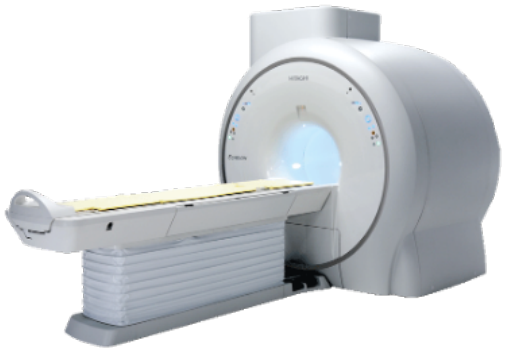
Signal in $\mathbb{R}^n$

noise in $\mathbb{R}^m$

Image inpainting



Radio Astronomy



MRI

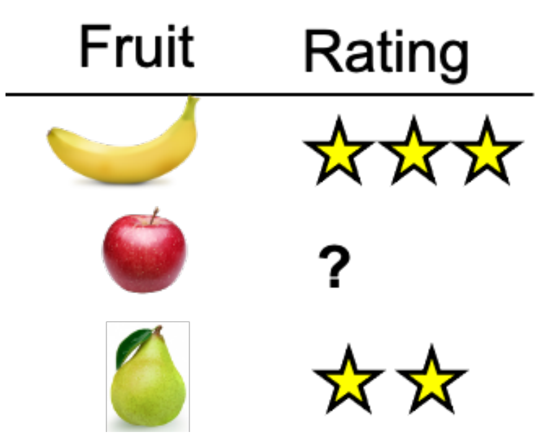
Ill-posed:

many x consistent with y

Solution:

Restrict to set of plausible signals $x \in \mathscr{X}$

Recommender system



Solving IP: regularised reconstruction

Idea: use a *prior* ≡ *loss* $\rho(x)$ for plausible reconstructions

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

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

Disadvantages:

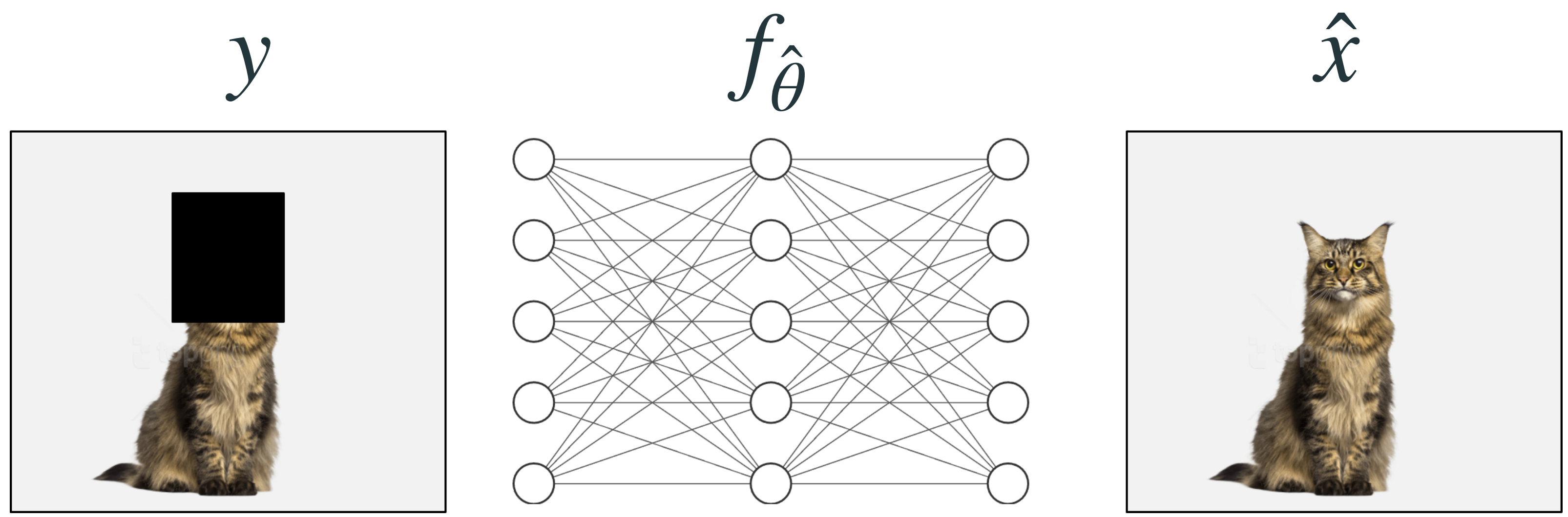
- Hard to define a good $\rho(x)$ in real world problems
- Loose with respect to the true signal distribution

Cat friendly presentation

Solving IP: learning approach

Idea:

- use training pairs of signals and measurements $\{(x_i, y_i)\}_{i=1}^N$
- learn the 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.

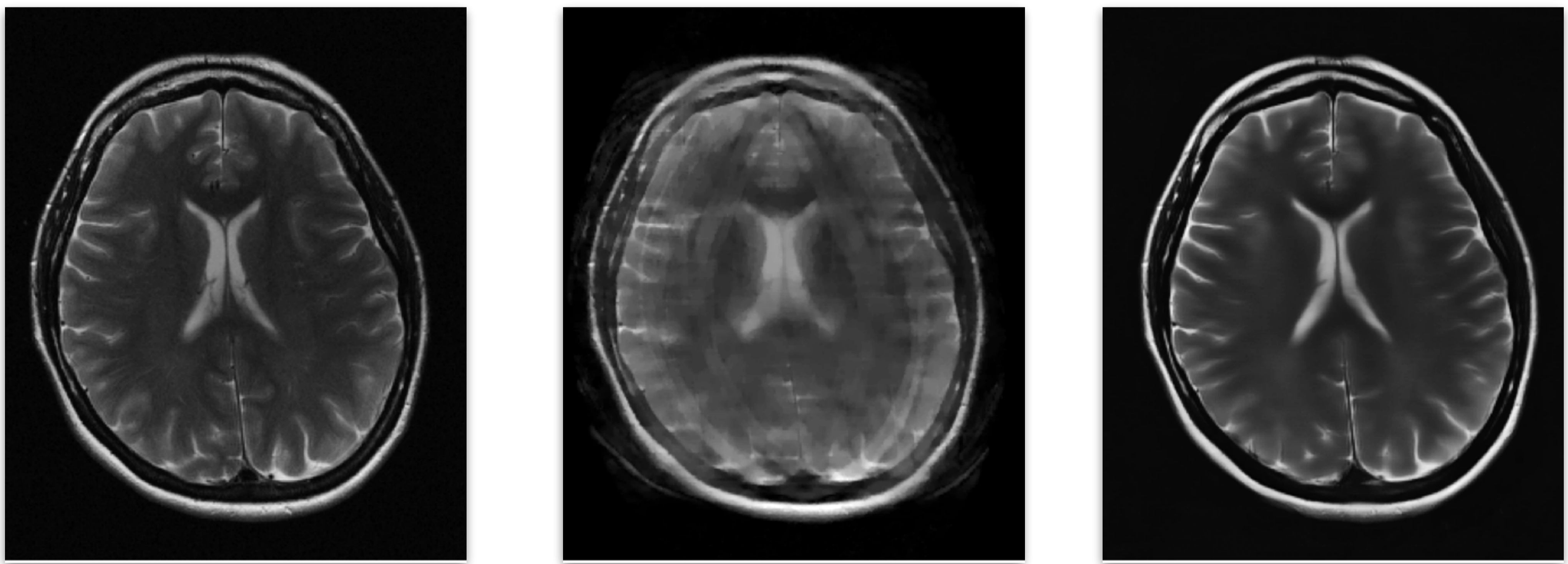
Solving IP: learning approach

Advantages:

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

fastMRI

Accelerating MR Imaging with AI



Ground-truth | Total variation (28.2 dB) | Deep network (28.2 dB)

→ x8 accelerated MRI [Zbontar et al., 2019]

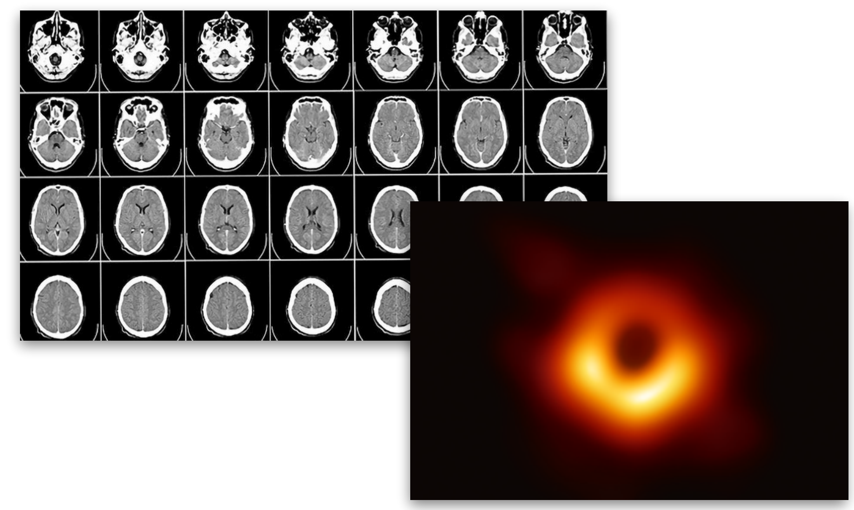
Solving IP: learning approach

Main disadvantage:

Obtaining training signals $\{x_i\}_i$ can be expensive/impossible.

For instance:

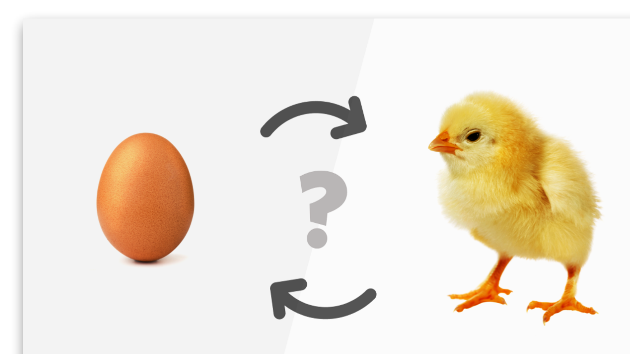
- Biomedical sciences (e.g., CT, MRI)
- Astronomical imaging (e.g., EHT)



Consequence:

- Risk to solve expected solution (off-distrib. prob)

Prior or *reconstruction*, which comes first?



Measurement-Driven Computational Imaging

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
& 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 $f = \mathrm{sign}$)

Purpose of this talk

Learning to reconstruct from binary measurements?

Theoretical analysis: given N binary observations & G operators

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

Estimate signal set $\mathcal{X} \supset \{x_i\}_{i=1}^N$? Error (LB/UB)? Sample Complexity?

Purpose of this talk

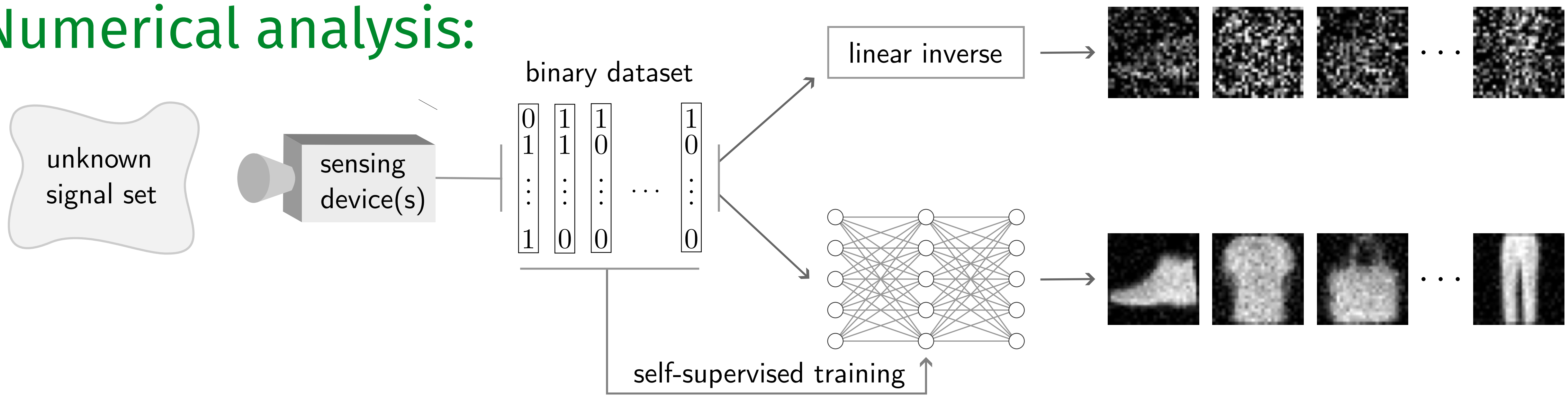
Learning to reconstruct from binary measurements?

Theoretical analysis: given N binary observations & G operators

$$y_i = \mathrm{sign}(\underbrace{A_{g_i}}_{\text{known}} \underbrace{x_i}_{\text{unknown}}) \text{ with } 1 \leq i \leq N \text{ and } g_i \in \{1, \ldots, G\}$$

Estimate signal set $\mathcal{X} \supset \{x_i\}_{i=1}^N$? Error (LB/UB)? Sample Complexity?

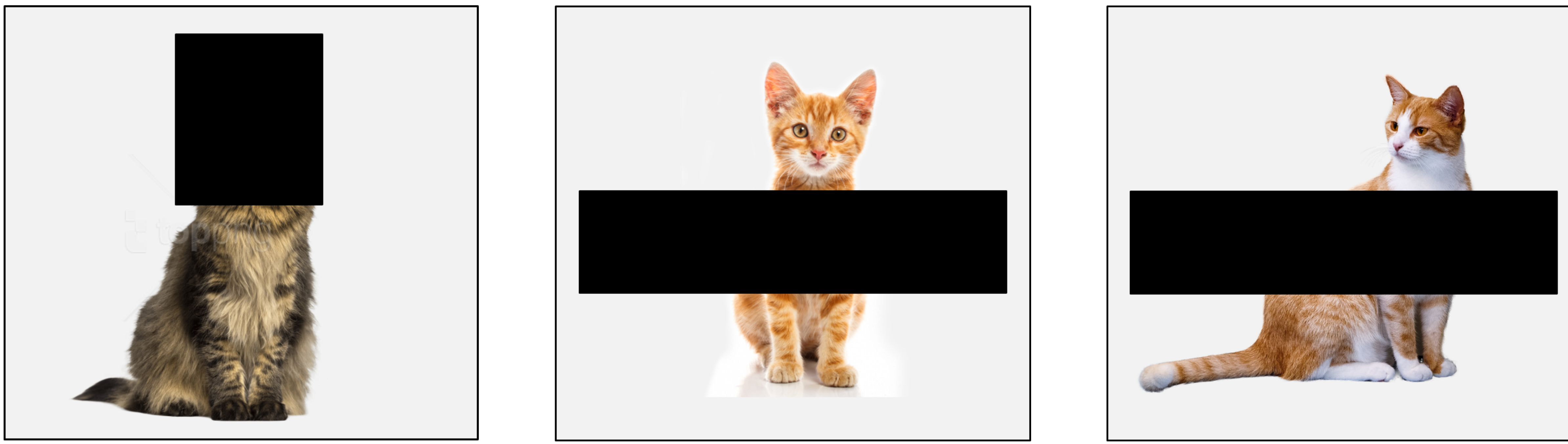
Numerical analysis:



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:

- $\neq$ access ratings for recommendation systems with $\neq$ users
- dynamic sensors: $\{A_t : t = n\Delta_T\}$
- multi-coil MRI

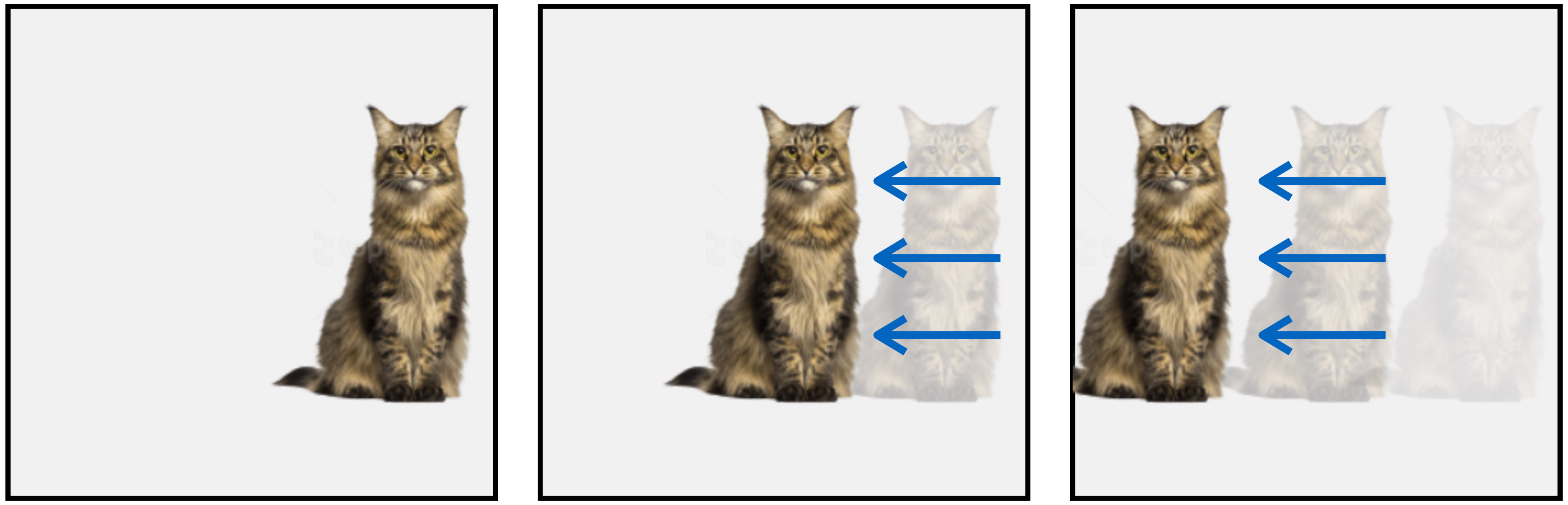
Sensing scenario 2: One operator & invariance

Most signals sets are invariant to groups of transformations:

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

(geometric prior)

Example: translation



Sensing scenario 2: One operator & invariance

Most signals sets are invariant to groups of transformations:

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

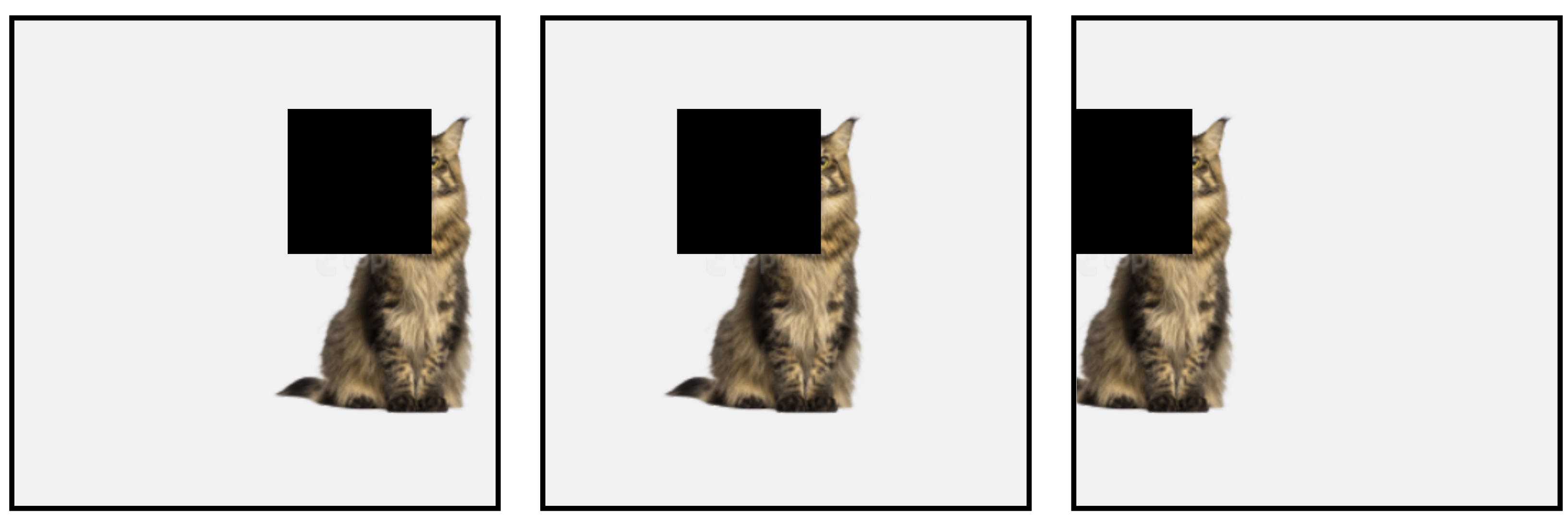
(geometric prior)

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

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

Implicit access to multiple operators
$A_g = AT_g$

AT_g for different g



Model identification: the problem

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

(More on this later)

Model identification: the problem

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

(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 $\mathscr{X}$?

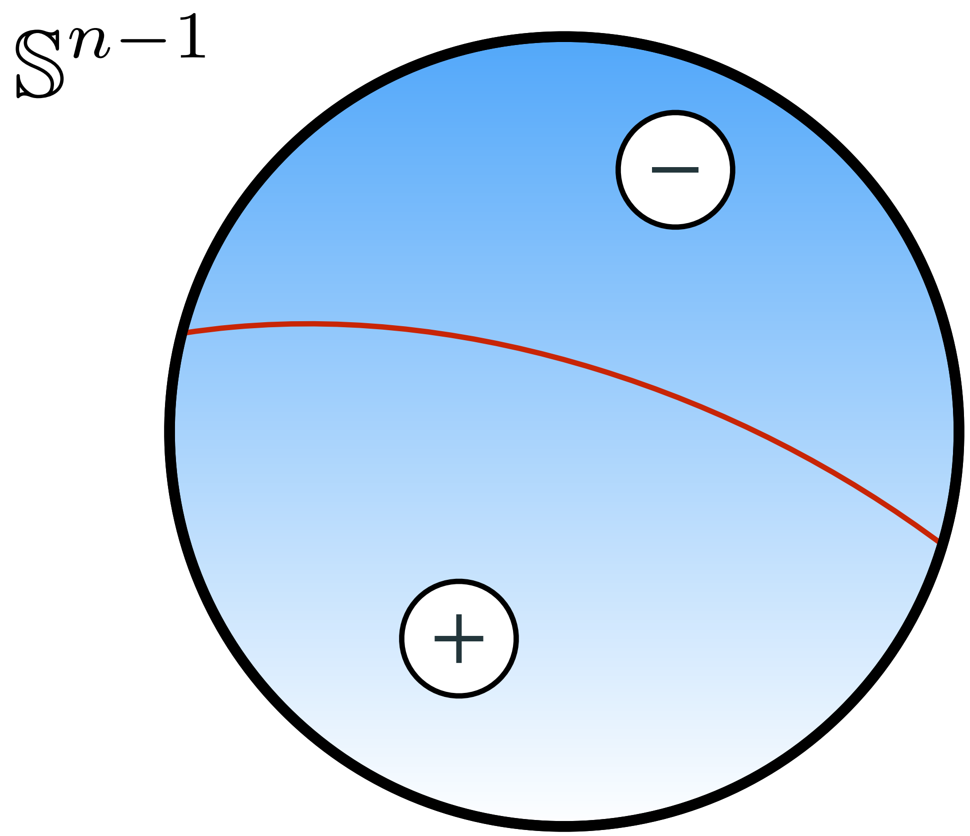
meaning?

→ From this set, a consistent decoder reads:

$$f(y) \in \{x \in \mathbb{S}^{n-1} | \, \mathrm{sign}(Ax) = y \text{ and } x \in \hat{\mathcal{X}}\}$$

Model identification: geometric intuition

Toy example: $n = 3$, increasing m, A_g has Gaussian iid entries



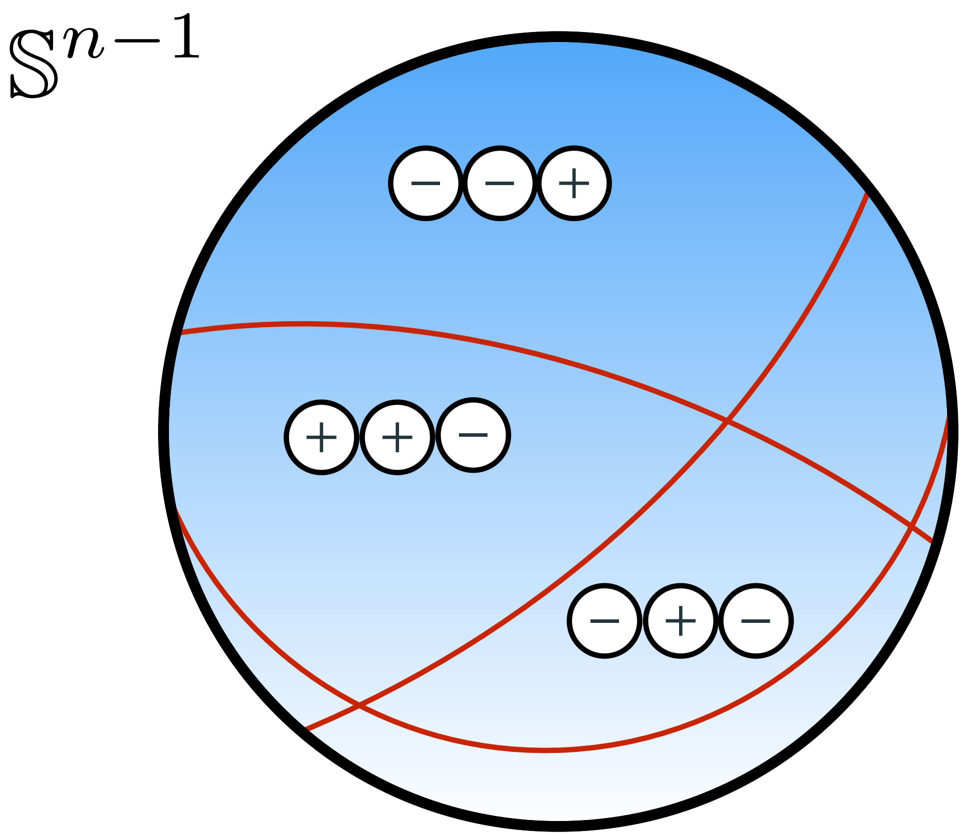
$m = 1$

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

Growing number of consistency cells as $m \uparrow$

Model identification: geometric intuition

Toy example: $n = 3$, increasing m, A_g has Gaussian iid entries



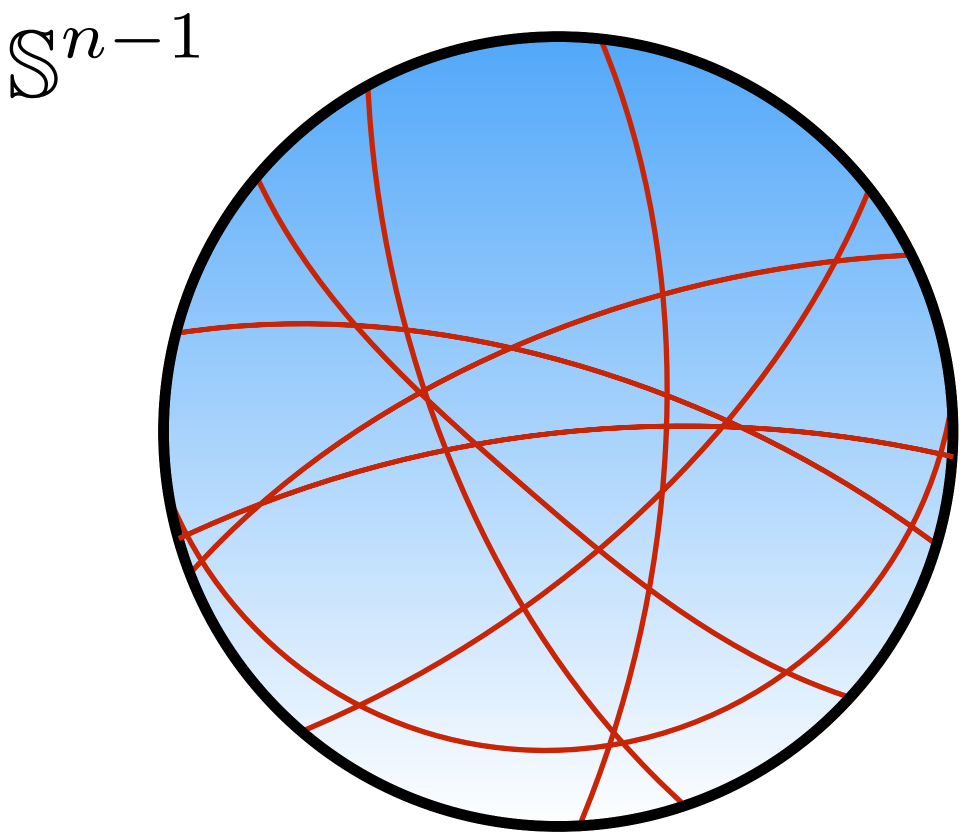
$m = 3$

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

Growing number of consistency cells as $m \uparrow$

Model identification: geometric intuition

Toy example: $n = 3$, increasing m, A_g has Gaussian iid entries



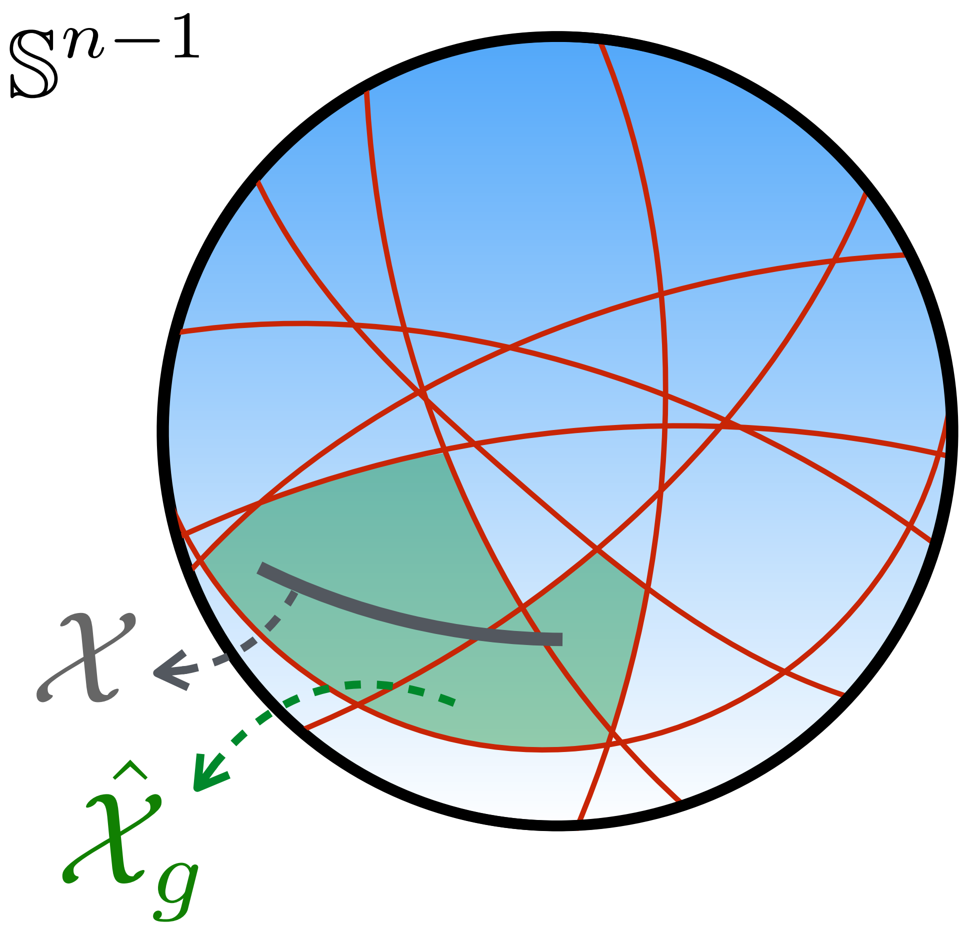
$m = 8$

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

Growing number of
consistency cells as $m \uparrow$

Model identification: geometric intuition

Toy example: $n = 3$, increasing m, A_g has Gaussian iid entries



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

Growing number of consistency cells as $m \uparrow$

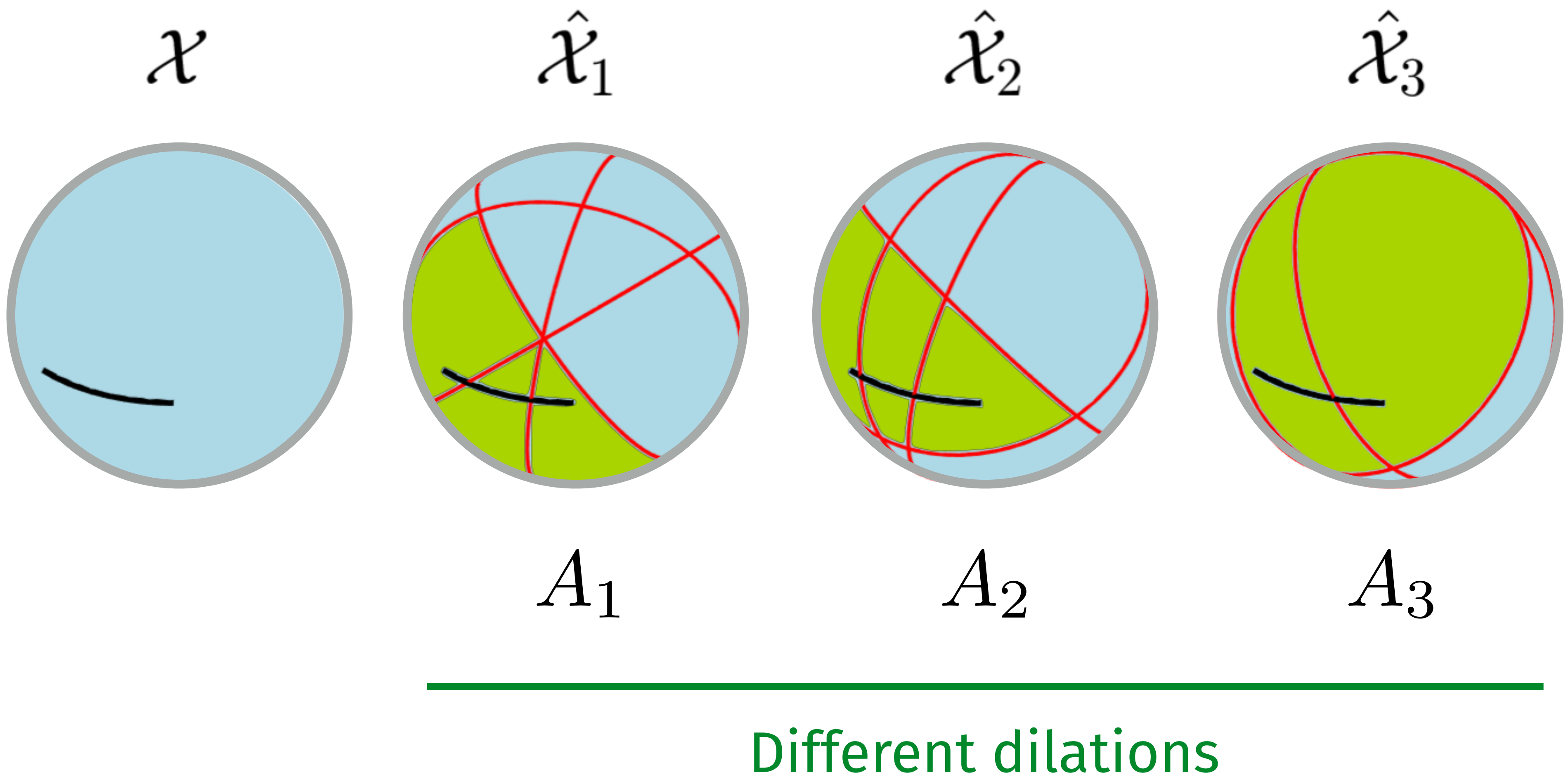
Let us define:

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

→ dilation of $\mathscr{X}$ by the "uncertainty" of $\operatorname{sign} \circ A_g$

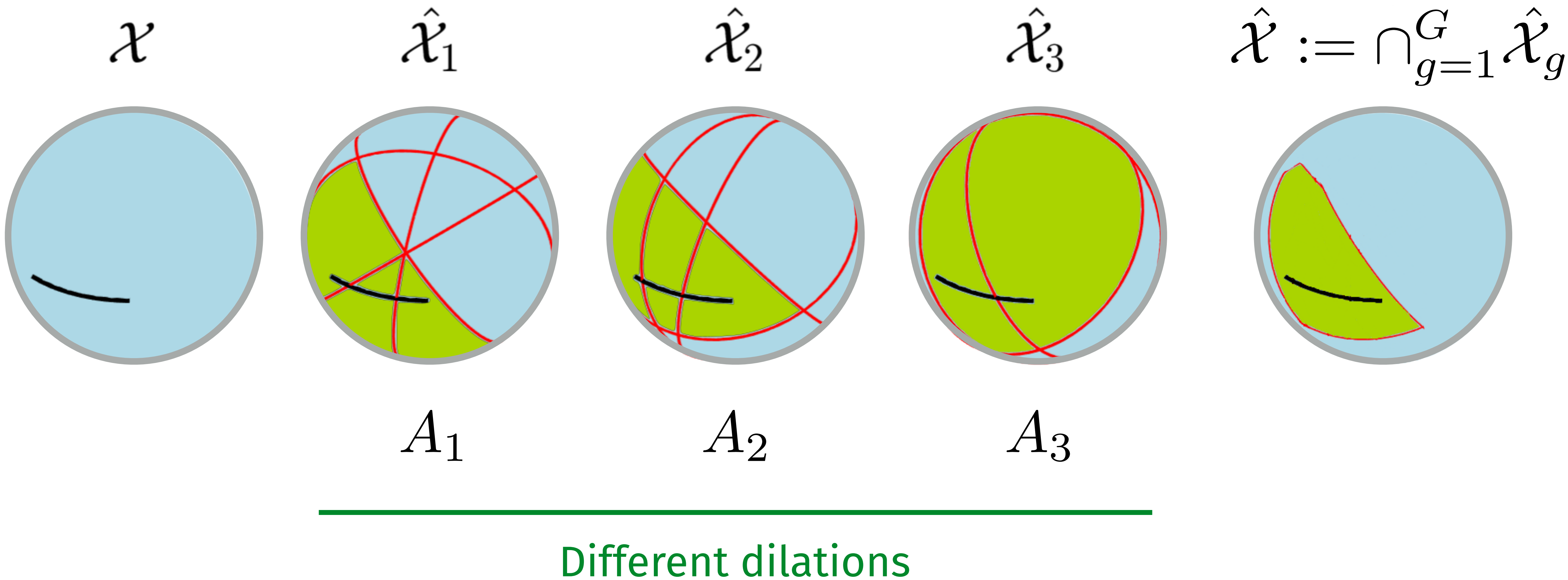
Model identification: geometric intuition

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



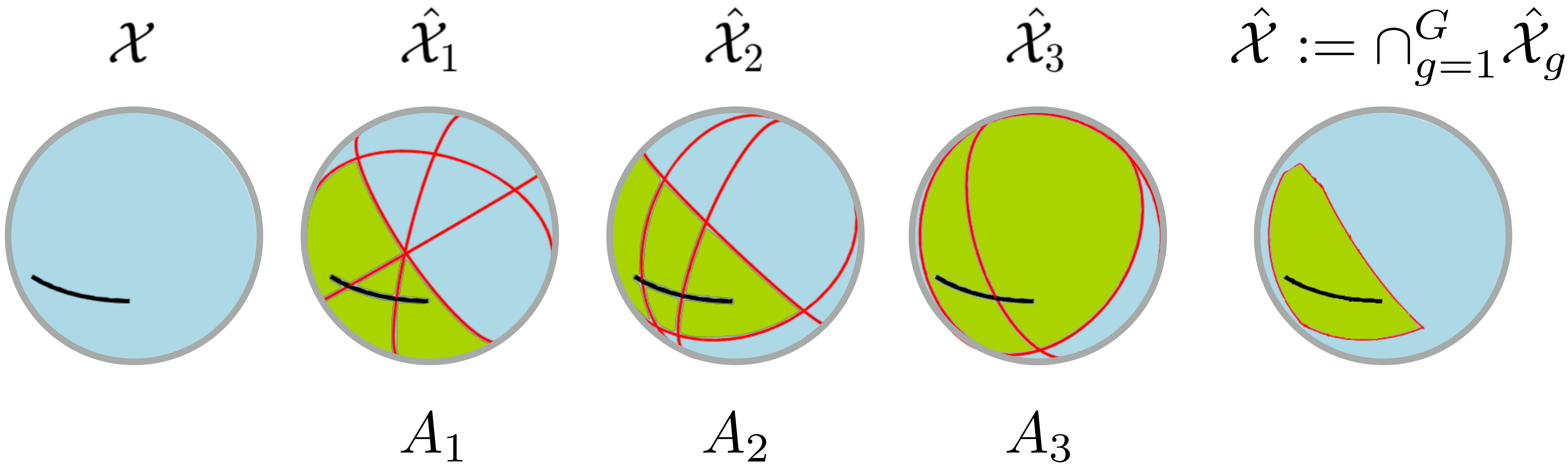
Model identification: geometric intuition

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



Model identification: geometric intuition

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



Identification error definition: (adapted to one-bit sensing)

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

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

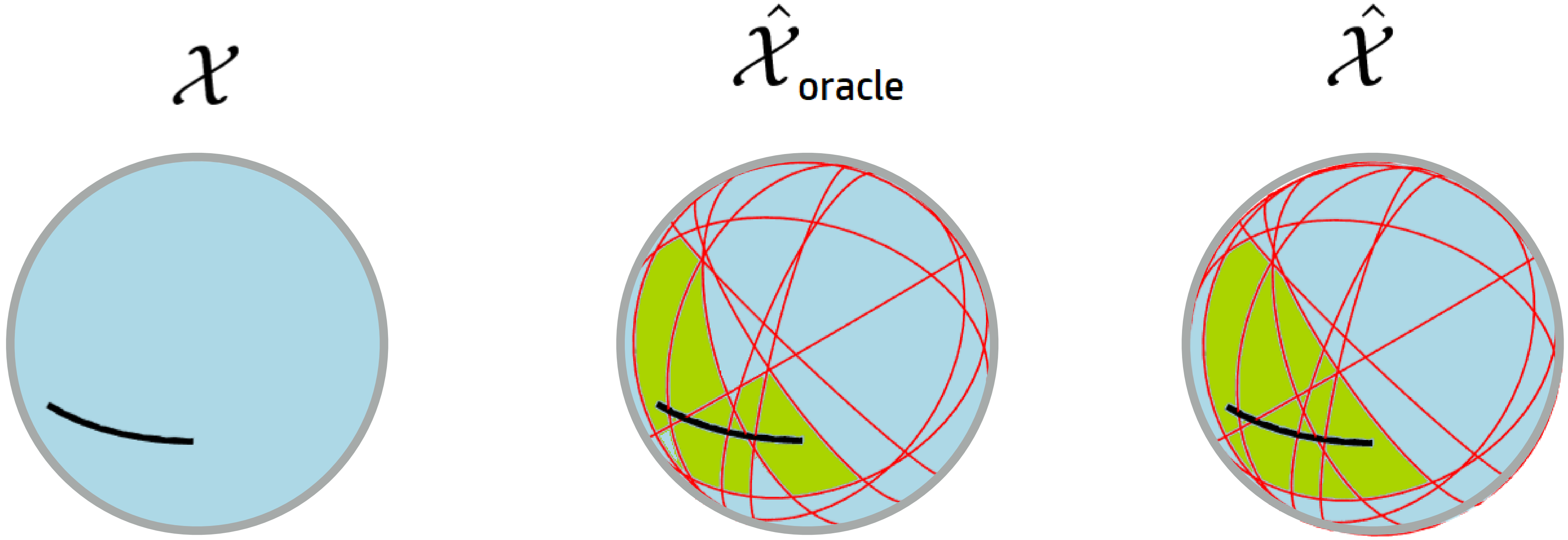
Upper bound? Lower bound? Sample complexity?

Lower bound (via an oracle standpoint)

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

$$\bar{A} = \begin{bmatrix} A_1 \\ \vdots \\ A_G \end{bmatrix} \in \mathbb{R}^{mG \times n}$$

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



Lower bound (via an oracle standpoint)

Theorem:

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

Proof sketch: Given $\mathcal{X}$ s.t. $\mathcal{X}_\delta \subsetneq \mathcal{X}_{\delta_0}$ if $\delta < \delta_0$, if $\delta < \min\{d, \delta_0\}$, then, $\exists R \in SO(n)$ such that, for $\mathcal{X}' = R\mathcal{X}$, $\hat{\mathcal{X}}'_{\text{oracle}} \not\subset \mathcal{X}'_\delta$.

Lower bound (via an oracle standpoint)

Theorem:

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

Proof sketch: Given $\mathcal{X}$ s.t. $\mathcal{X}_\delta \subsetneq \mathcal{X}_{\delta_0}$ if $\delta < \delta_0$, if $\delta < \min\{d, \delta_0\}$, then, $\exists R \in SO(n)$ such that, for $\mathcal{X}' = R\mathcal{X}$, $\hat{\mathcal{X}}'_{\text{oracle}} \not\subset \mathcal{X}'_\delta$.

Consequences: we can show the following

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

$\rightarrow$ Model identification error is trivially large

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

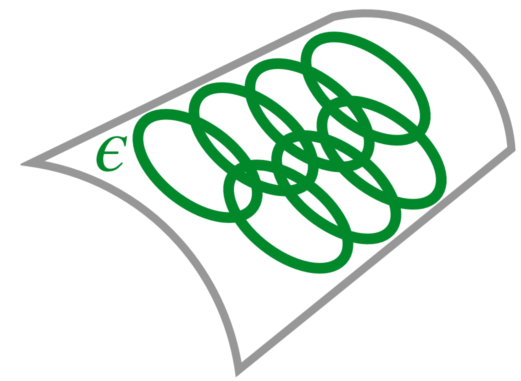
$\rightarrow$ No learning with $G = 1$ for incomplete operator (w/o invariance)

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

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

Upper bound (with 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.

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

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

then

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

Consequences of the theorem:

→ The identification error 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 $\sim$ signal recovery errors from one-bit CS

Sample Complexity (with randomness)

How many binary observations to obtain $\hat{\mathscr{X}}$?

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

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*

$$N \leq G\left(\frac{m\sqrt{n}}{k}\right)^{8k}$$

possible different measurements vectors.

→ Exponential on the model dim k and not the ambient dim n !

Algorithms

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

Goal:

Learn 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:

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

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

Problem: this is a consistent reconstruction $f_\theta(y, A_g) = A_g^\dagger y$

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:

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

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

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

→ 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 "***S****elf-****S****upervised learning loss for training reconstruction networks from* ***B****inary* ***M****easurement 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:

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

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

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

→ promotes equivariance of $f_\theta \circ A$, i.e., $(f \circ A)(T_g \cdot) = T_g (f \circ A)(\,\cdot\,)$

Remarks:

- Network-agnostic scheme (applicable to any existing deep model)
- We called this "***S****elf-****S****upervised learning loss for training reconstruction networks from* ***B****inary* ***M****easurement data alone*" (SSBM)

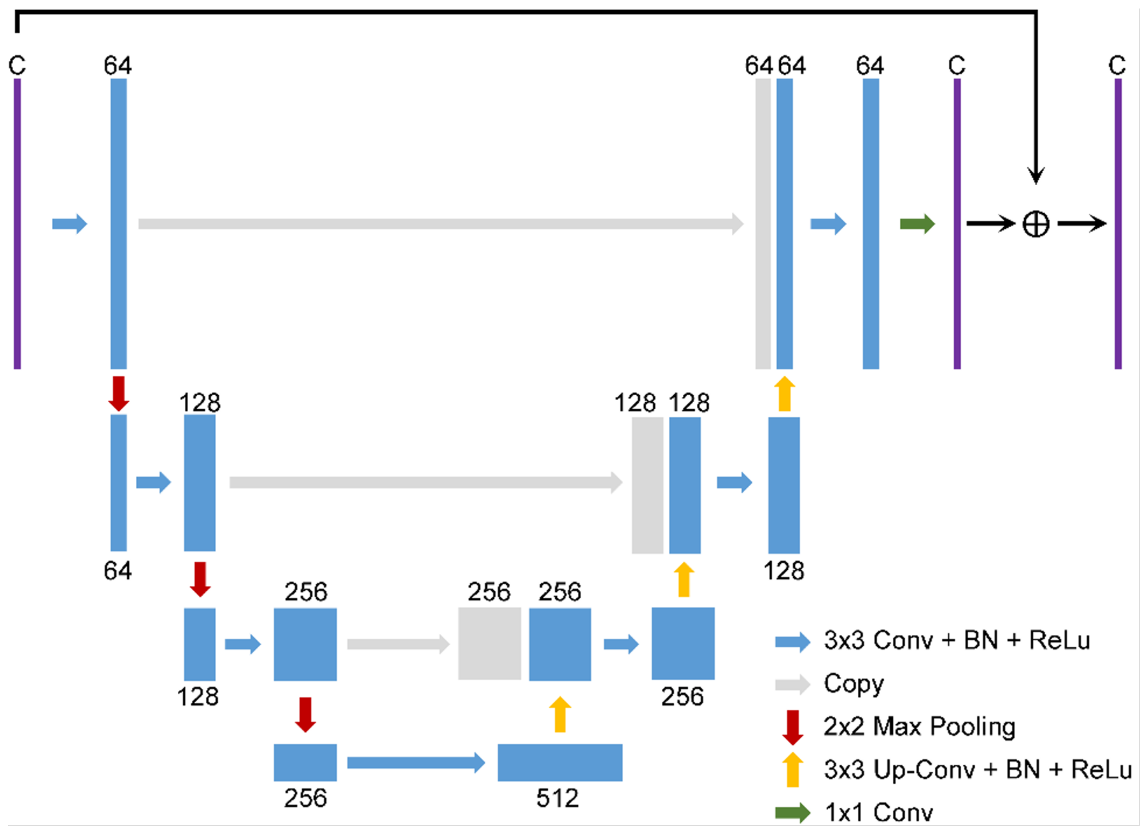
Experiments

Operators:

- A_g with Gaussian iid entries

Network:

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

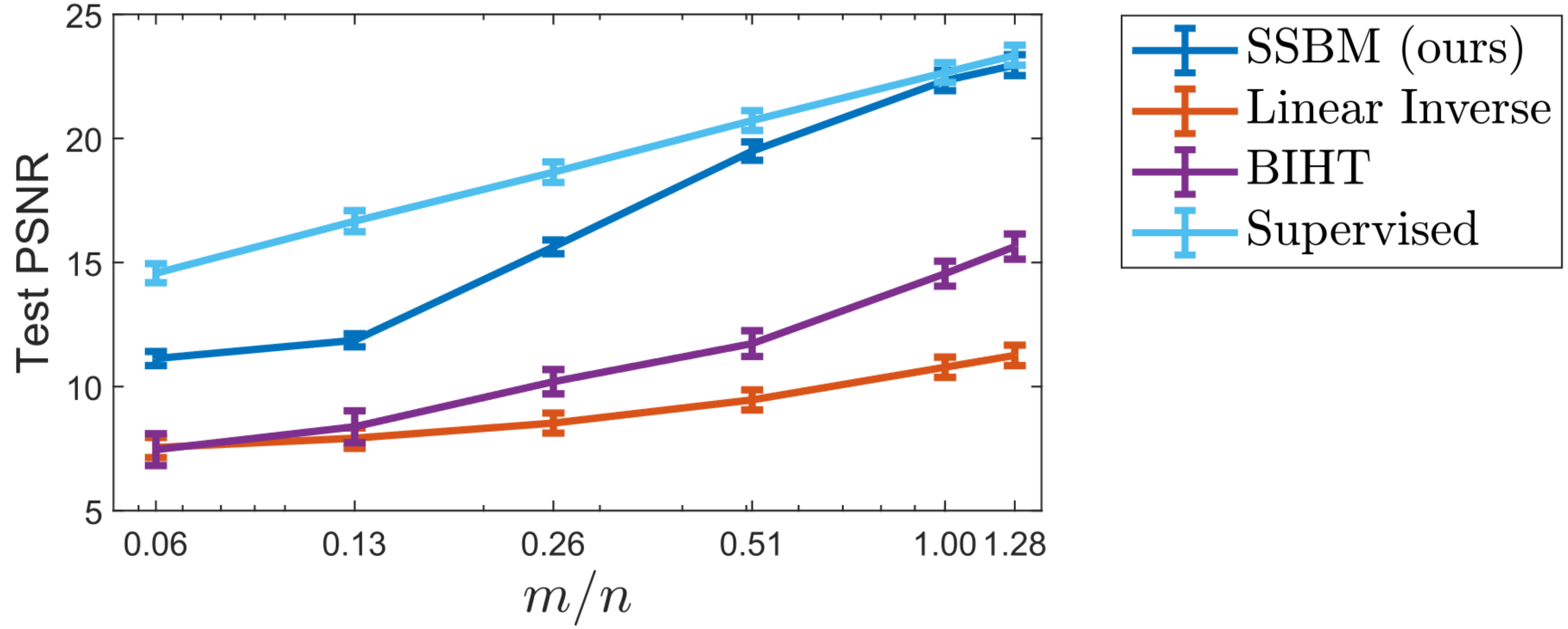


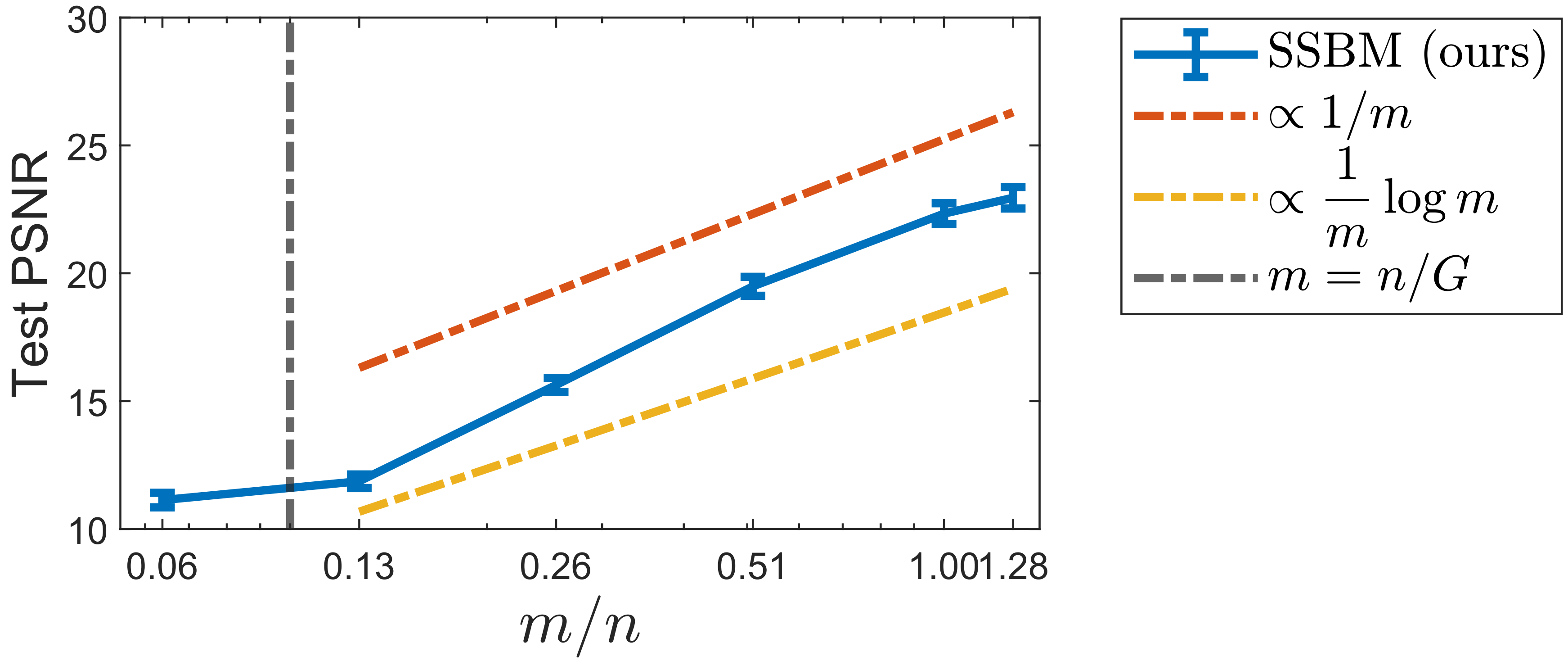
Comparison:

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

MNIST

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

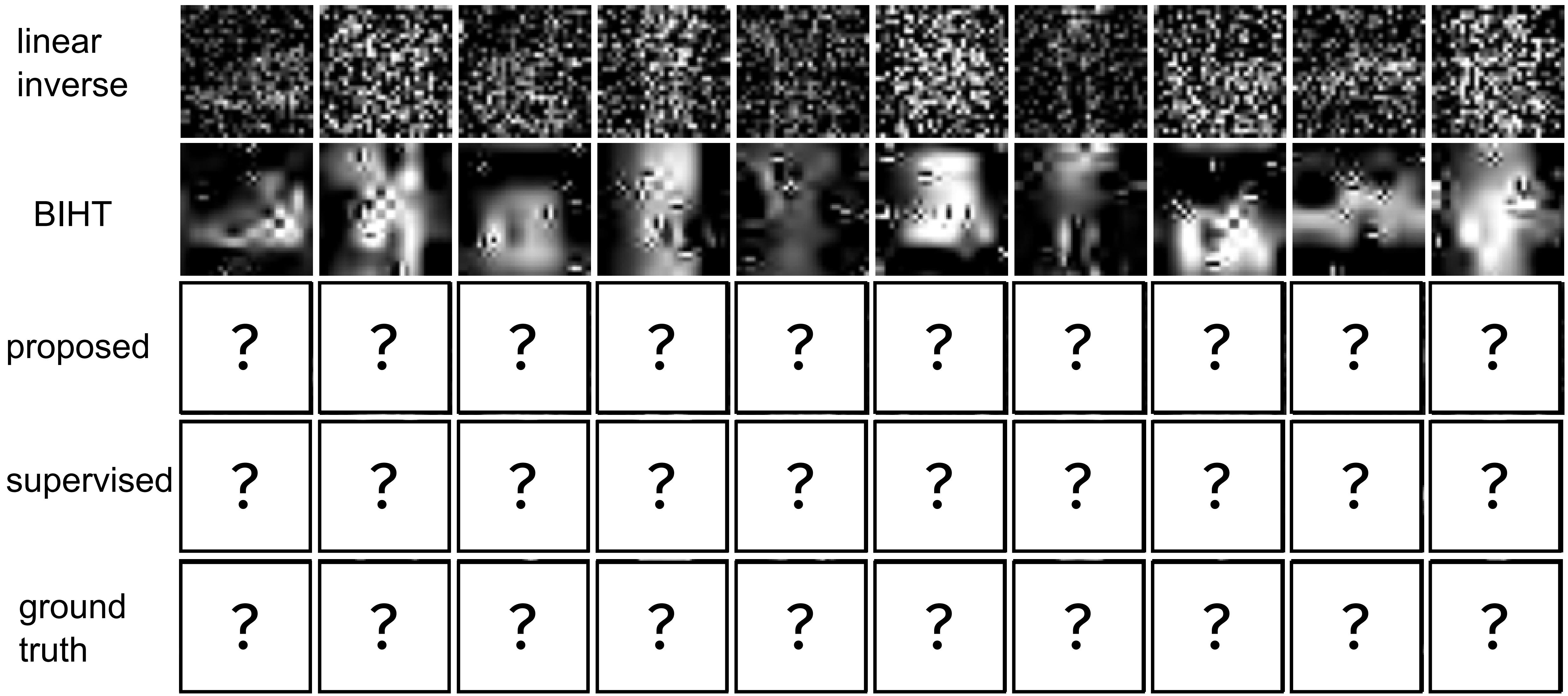




$$\text{Test PSNR} := \frac{1}{N'} \sum_{i=1}^{N'} \text{PSNR}\left(x_i', f_\theta\left(\text{sign}\left(A_{g_i} x_i'\right), A_{g_i}\right)\right),\ x_i' \in \text{“test set”}$$

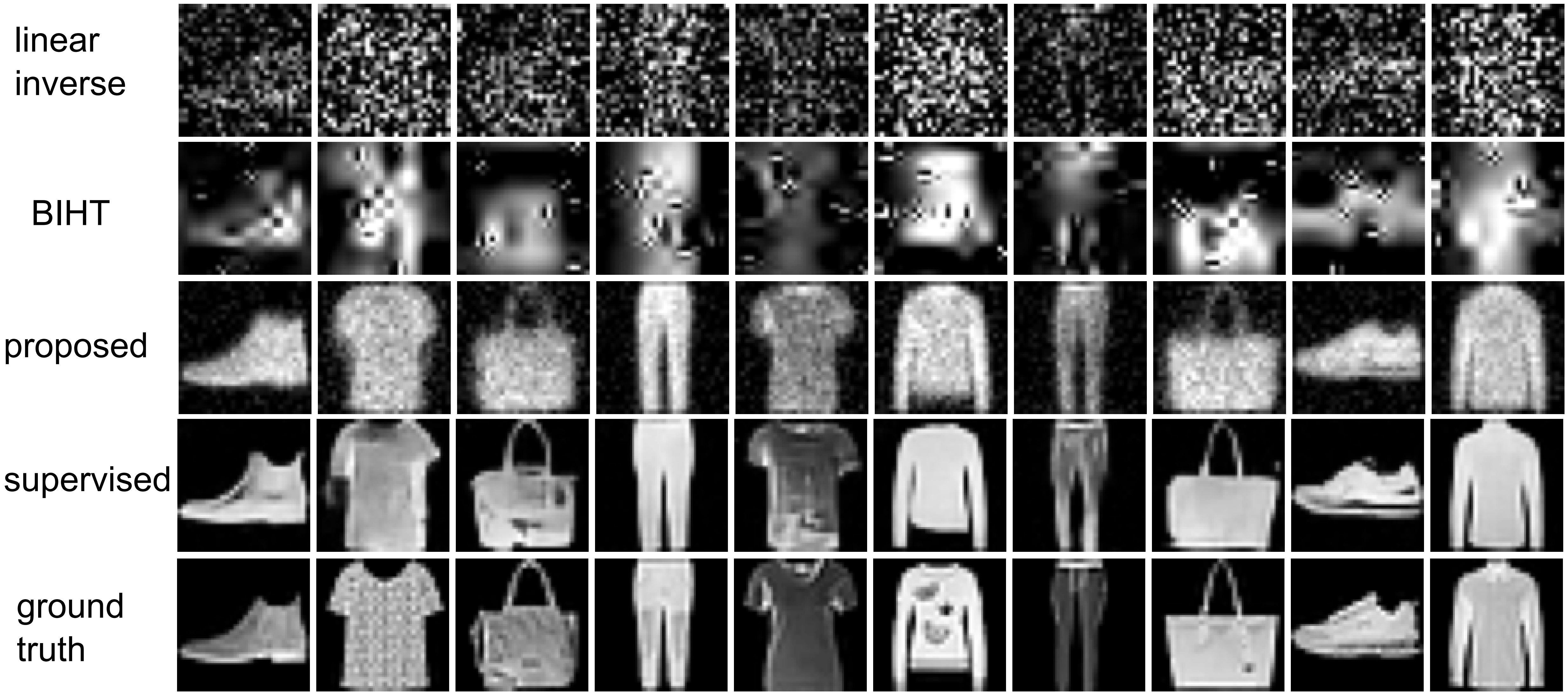
Fashion MNIST

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



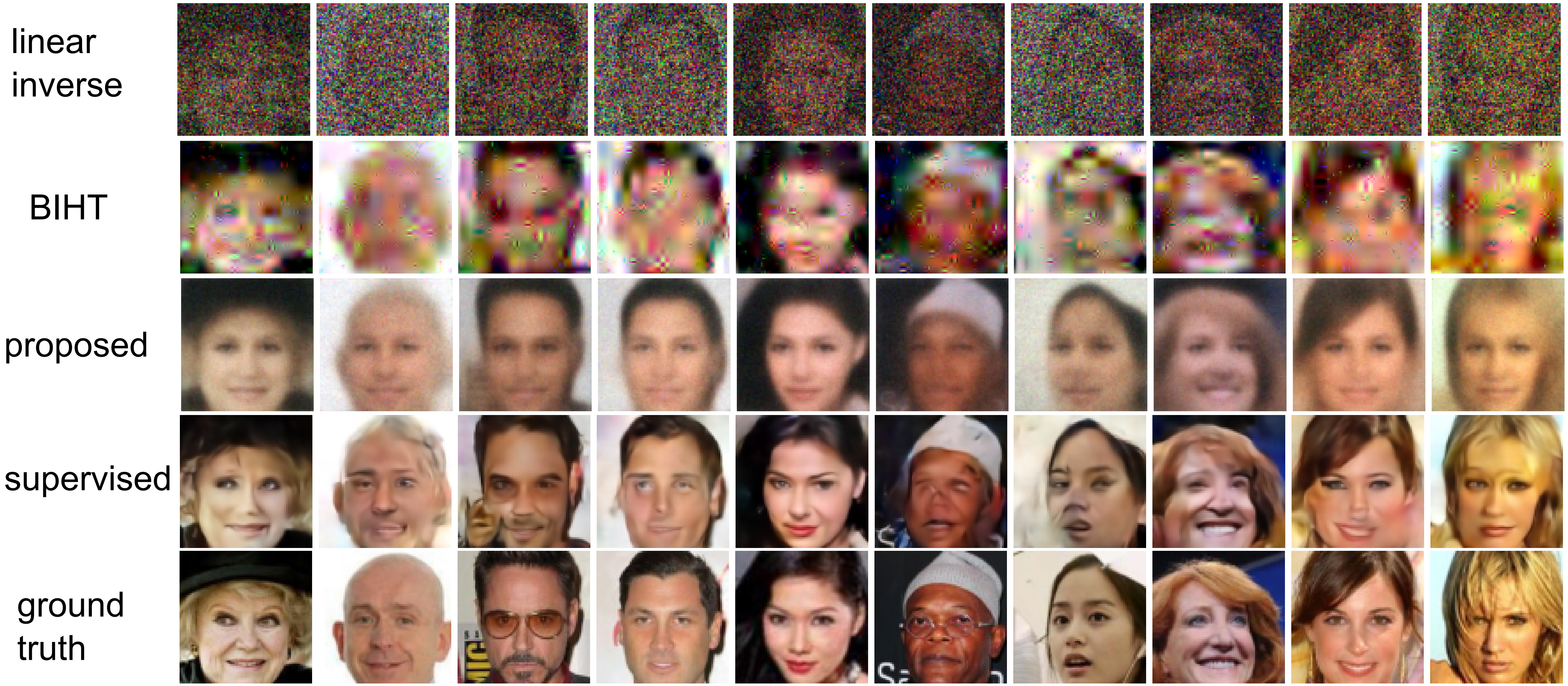
Fashion MNIST

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



CelebA

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: Conditions for learning

- Number of measurements
- L/U Bounds on global identification error

Practice: Deep learning approach

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

Ongoing/future work

- Other non-linear inverse problems (e.g., phase retrieval)
- Upper bounds for the invariant case
- Noise/dither $y = \text{sign}(Ax + \epsilon)$

Thank you!

References:

- 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", arXiv:2303.08691, 2023**