Learning to reconstruct from clipped signals alone

EUSIPCO 2024
Lyon, France

Victor Sechaud[1], **Laurent Jacques**[2], **Patrice Abry**[1], **Julian Tachella**[1]

[1]CNRS, ENS de Lyon
[2]UCLouvain

August 27, 2024

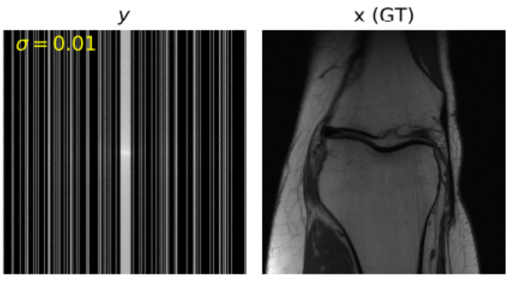
Introduction

What is an inverse problem?

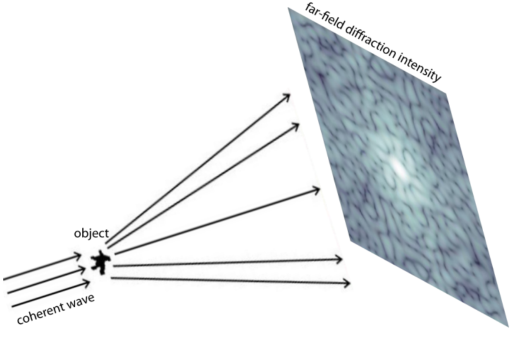
$$y = A(x) + \epsilon$$

- x original signal.
- y measurement.
- ϵ noise.
- A forward operator.

Magnetic resonance imaging (linear):



Coherent diffraction imaging (non-linear):



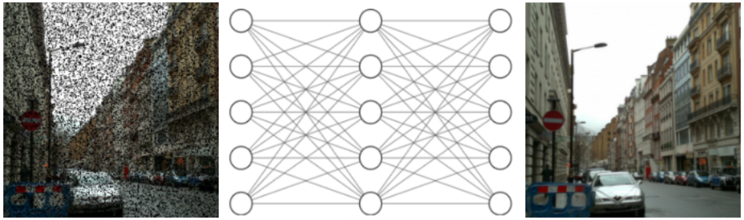
Deep learning approches

- Deep learning is now the state of the art.

Objective:

- **Learn** inverse forward operator of A, $f_\theta(y) \approx x$.
- **Using** a dataset $\{(x_i, y_i)\}_{i \in I}$ + a neural network.

f_θ



$$\arg\min_{\theta} \sum_{i \in I} \| f_\theta(y_i) - x_i \|^2$$

Limit of the method

- We need a large set of $(x_i)_{i \in I}$ (ground-truth).
- Training and testing datasets can be very different.

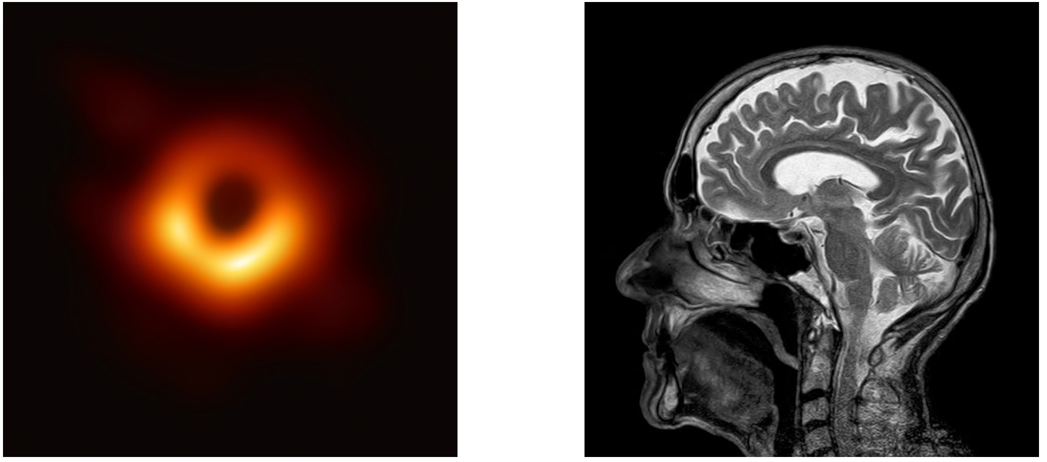
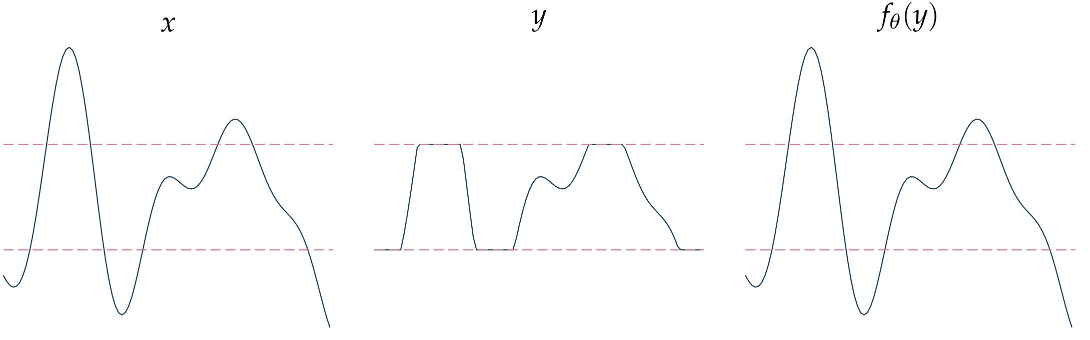


Figure 1. Astronomical and medical imaging.

Unsupervised learning for saturated signals

- **Non linear equation:** $y = \eta(x)$

$$y_i = \begin{cases} -\mu & \text{if} \quad x \leq -\mu \\ x & \text{if} \quad x \in [-\mu; \mu] \\ \mu & \text{if} \quad x \geq \mu \end{cases}$$



Unsupervised learning

- **Dataset:** $\cancel{(x_i, y_i)_{i \in I}}$ only $(y_i)_{i \in I}$.
- What can we do with only y_i?

$$y = \eta(x) \Rightarrow y \approx \eta(f_\theta(y))$$

$$\arg\min_\theta \sum_{i \in I} \|y_i - \eta(f_\theta(y_i))\|^2$$

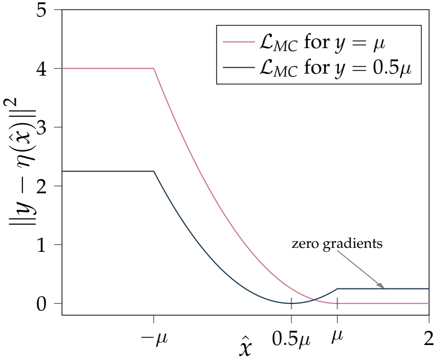
Unsupervised learning

- **Dataset:** $\cancel{(x_i, y_i)_{i\in I}}$ only $(y_i)_{i\in I}$.
- What can we do with only y_i?

$$y = \eta(x) \Rightarrow y \approx \eta(f_\theta(y))$$
$$\arg\min_\theta \sum_{i\in I} \|y_i - \eta(f_\theta(y_i))\|^2$$

Naive measurement consistency loss:

$$\mathcal{L}_{\text{NMC}}(\theta) = \sum_{i\in I} \|\eta(f_\theta(\boldsymbol{y}_i)) - \boldsymbol{y}_i\|^2 .$$



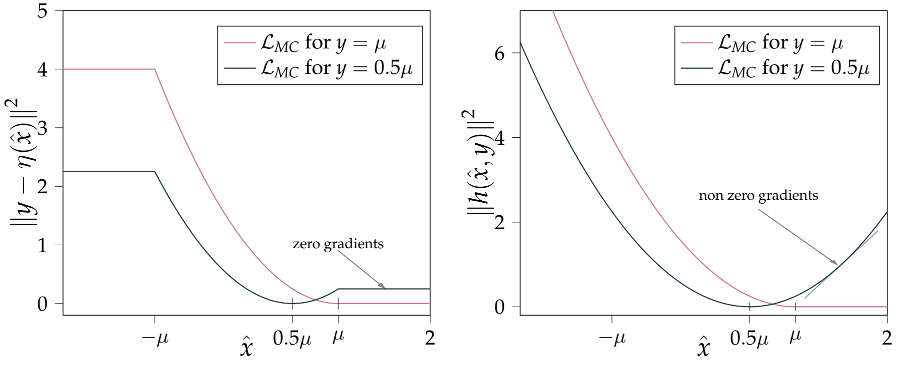
Unsupervised learning

Measurement consistency loss:

$$\mathcal{L}_{\mathrm{MC}}(\theta) = \sum_{i \in I} \|h(f_\theta(\boldsymbol{y}_i), \boldsymbol{y}_i)\|^2 .$$

with h defined as

$$h(\boldsymbol{u}, \boldsymbol{v})_j = \begin{cases} v_j - u_j & \text{if} \quad |v_j| < \mu \text{ or } \operatorname{sign}(v_j) u_j < \mu \\ 0 & \text{if} \quad |v_j| = \mu \text{ and } \operatorname{sign}(v_j) u_j \geq \mu. \end{cases}$$



Limit of measurement consistency

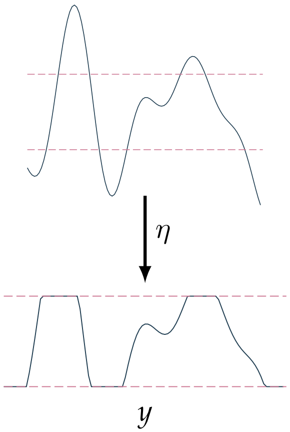
- **Dataset:** $\cancel{(x_i, y_i)_{i\in I}}$ only $(y_i)_{i\in I}$.
- What can we do with only y_i?

$$y = \eta(x) \Rightarrow y \approx \eta\left(f_\theta(y)\right)$$

$$\arg\min_\theta \sum_{i\in I} \left\| h\left(f_\theta(\boldsymbol{y}_i), \boldsymbol{y}_i\right)\right\|^2 .$$

Limit of measurement consistency

- **Dataset:** $\cancel{(x_i, y_i)_{i\in I}}$ only $(y_i)_{i\in I}$.
- What can we do with only y_i?



$$y = \eta(x) \Rightarrow y \approx \eta\left(f_\theta(y)\right)$$

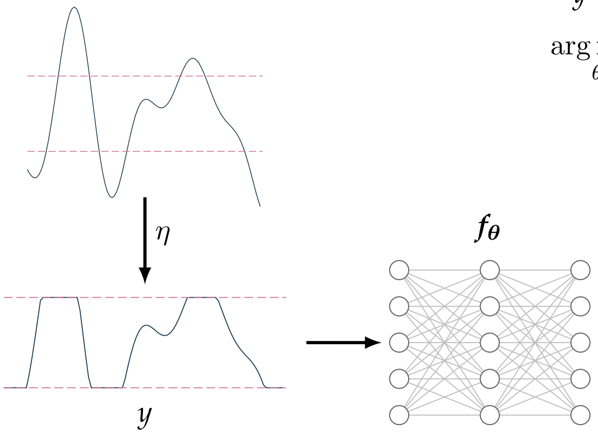
$$\arg\min_\theta \sum_{i\in I} \left\|h\left(f_\theta(\boldsymbol{y}_i), \boldsymbol{y}_i\right)\right\|^2 .$$

Limit of measurement consistency

- **Dataset:** $\cancel{(x_i, y_i)_{i \in I}}$ only $(y_i)_{i \in I}$.
- What can we do with only y_i?

$$y = \eta(x) \Rightarrow y \approx \eta\left(f_\theta(y)\right)$$

$$\arg\min_{\theta} \sum_{i \in I} \left\| h\left(f_\theta(\boldsymbol{y}_i)\, , \boldsymbol{y}_i\right) \right\|^2 .$$

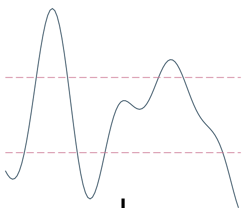


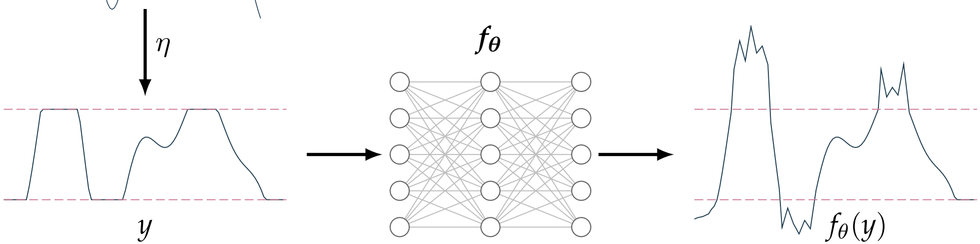
Limit of measurement consistency

- **Dataset:** $\cancel{(x_i, y_i)_{i\in I}}$ only $(y_i)_{i\in I}$.
- What can we do with only y_i?

$$y = \eta(x) \Rightarrow y \approx \eta\left(f_\theta(y)\right)$$

$$\arg\min_\theta \sum_{i\in I} \left\|h\left(f_\theta(\boldsymbol{y}_i), \boldsymbol{y}_i\right)\right\|^2.$$



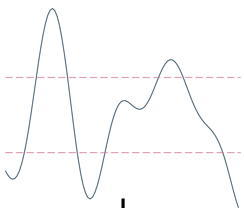


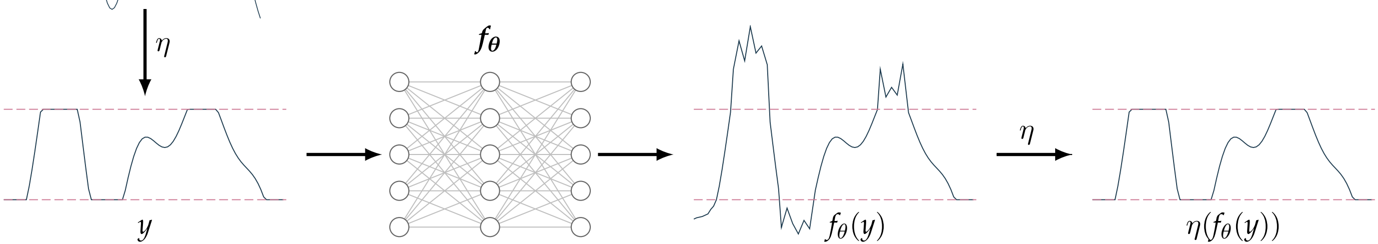
LIMIT OF MEASUREMENT CONSISTENCY

- **Dataset:** ~~$(x_i, y_i)_{i \in I}$~~ only $(y_i)_{i \in I}$.
- What can we do with only y_i?

$$y = \eta(x) \Rightarrow y \approx \eta\left(f_\theta(y)\right)$$

$$\arg\min_\theta \sum_{i \in I} \left\| h\left(f_\theta(\boldsymbol{y}_i), \boldsymbol{y}_i\right) \right\|^2 .$$



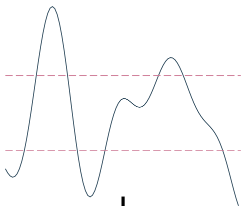


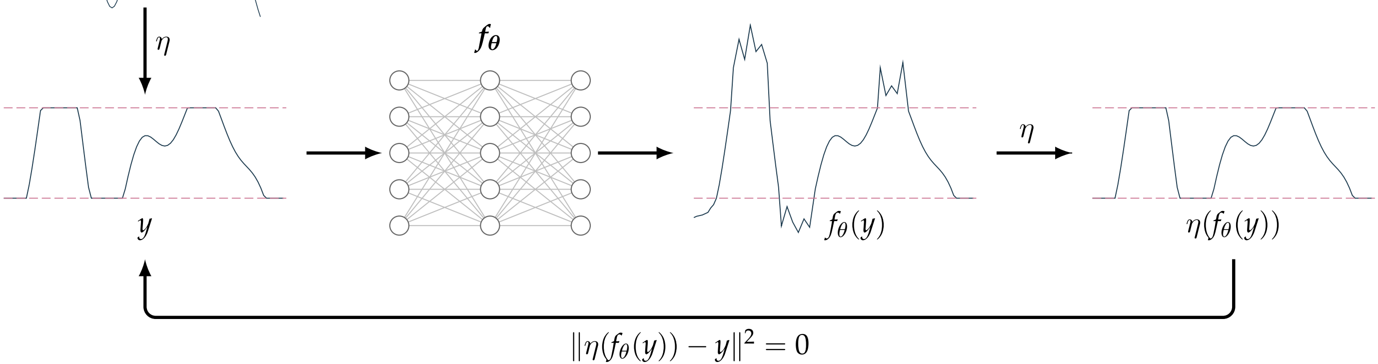
Limit of measurement consistency

- **Dataset:** $\cancel{(x_i, y_i)_{i\in I}}$ only $(y_i)_{i\in I}$.
- What can we do with only y_i?

$$y = \eta(x) \Rightarrow y \approx \eta\left(f_\theta(y)\right)$$

$$\arg\min_{\theta} \sum_{i\in I} \left\|h\left(f_\theta(\boldsymbol{y}_i)\,,\boldsymbol{y}_i\right)\right\|^2 .$$





Amplitude invariance

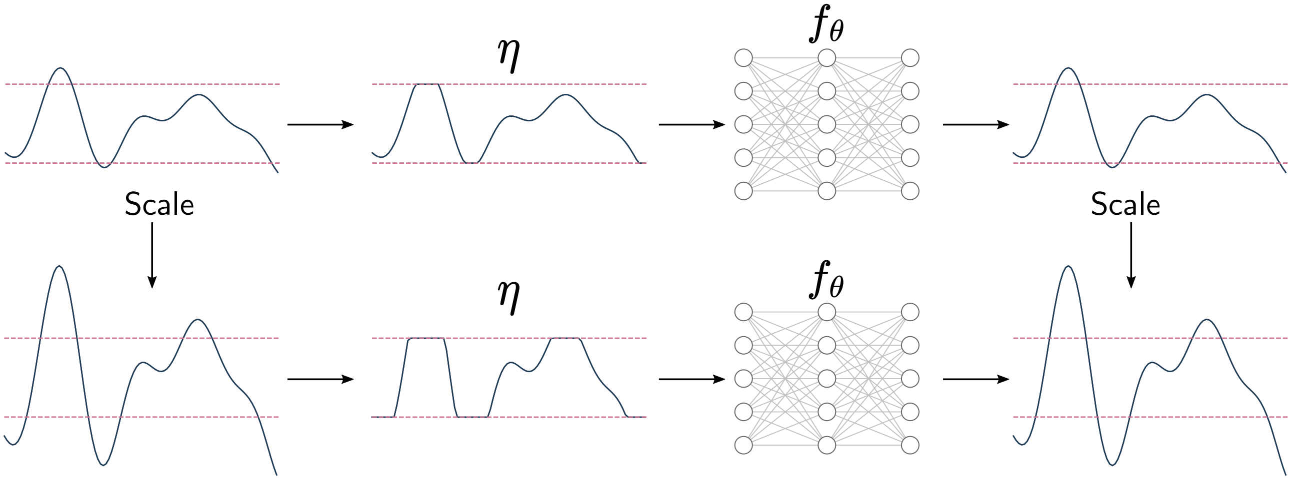
- **Amplitude invariance:** $\forall g \in \mathbb{R}^+, \forall x \in \mathcal{X},\ gx \in \mathcal{X}$
- **Equivariance property:**

$$\boxed{g f_\theta(\eta(x)) = f_\theta(\eta(gx))}$$

Amplitude invariance

- **Amplitude invariance:** $\forall g \in \mathbb{R}^+, \forall x \in \mathcal{X},\ gx \in \mathcal{X}$
- **Equivariance property:**

$$\boxed{g f_\theta(\eta(x)) = f_\theta(\eta(gx))}$$



Proposed method

- **Equivariance promoting loss:**

$$\mathcal{L}_{\mathrm{EI}}(\theta) = \sum_{i \in I} \mathbb{E}_{g \sim p_g} \left[\left\| g f_\theta \left(\boldsymbol{y}_i \right) - f_\theta \left(\eta \left(g f_\theta \left(\boldsymbol{y}_i \right) \right) \right) \right\|^2 \right]$$

Proposed method

- **Equivariance promoting loss:**

$$\mathcal{L}_{\mathrm{EI}}(\theta) = \sum_{i \in I} \mathbb{E}_{g \sim p_g} \left[\left\| g f_\theta\left(\boldsymbol{y}_i\right) - f_\theta\left(\eta\left(g f_\theta\left(\boldsymbol{y}_i\right)\right)\right) \right\|^2 \right]$$

- **Measurement consistency loss:**

$$\mathcal{L}_{\mathrm{MC}}(\theta) = \sum_{i \in I} \| h(f_\theta(\boldsymbol{y}_i), \boldsymbol{y}_i) \|^2 .$$

with h defined as

$$h(\boldsymbol{u}, \boldsymbol{v})_j = \begin{cases} v_j - u_j & \text{if} \quad |v_j| < \mu \text{ or } \operatorname{sign}(v_j) u_j < \mu \\ 0 & \text{if} \quad |v_j| = \mu \text{ and } \operatorname{sign}(v_j) u_j \geq \mu. \end{cases}$$

Proposed method

- **Equivariance promoting loss:**

$$\mathcal{L}_{\mathrm{EI}}(\theta) = \sum_{i \in I} \mathbb{E}_{g \sim p_g} \left[\left\| g f_\theta\left(\boldsymbol{y}_i\right) - f_\theta\left(\eta\left(g f_\theta\left(\boldsymbol{y}_i\right)\right)\right) \right\|^2 \right]$$

- **Measurement consistency loss:**

$$\mathcal{L}_{\mathrm{MC}}(\theta) = \sum_{i \in I} \left\| h(f_\theta(\boldsymbol{y}_i), \boldsymbol{y}_i) \right\|^2 .$$

with h defined as

$$h(\boldsymbol{u}, \boldsymbol{v})_j = \begin{cases} v_j - u_j & \text{if} \quad |v_j| < \mu \text{ or } \operatorname{sign}(v_j) u_j < \mu \\ 0 & \text{if} \quad |v_j| = \mu \text{ and } \operatorname{sign}(v_j) u_j \geq \mu. \end{cases}$$

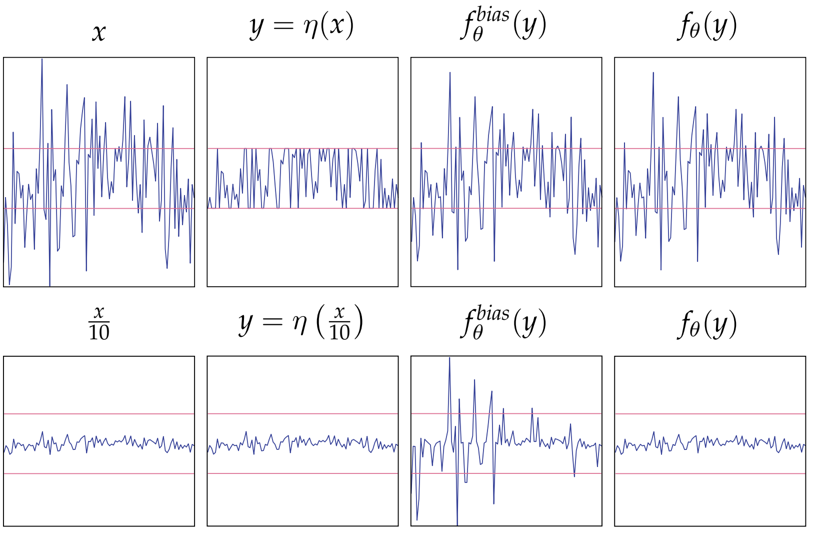
Training loss:

$$\mathcal{L}(\theta) = \mathcal{L}_{\mathrm{MC}}(\theta) + \mathcal{L}_{\mathrm{EI}}(\theta)$$

Implementation details

- **Bias-free** network:

$$f_\theta(\alpha y) = \alpha f_\theta(y) \quad \forall \alpha \in \mathbb{R}_+.$$



Implementation details

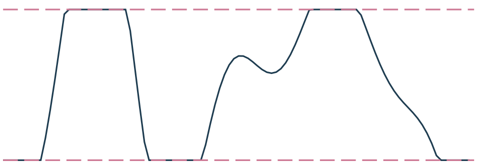
- **Bias-free** network:

$$f_\theta(\alpha y) = \alpha f_\theta(y) \quad \forall \alpha \in \mathbb{R}_+.$$

- **Masking:**

$$\hat{x}_j = (1 - b_j) y_j + b_j f_\theta(y)_j.$$

$$\text{with } b_j = \frac{\max(0, |y_j| - \tau\mu)}{(1-\tau)\mu}.$$



Implementation details

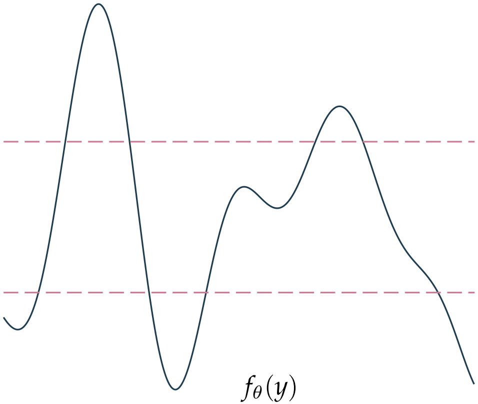
- **Bias-free** network:

$$f_\theta(\alpha y) = \alpha f_\theta(y) \quad \forall \alpha \in \mathbb{R}_+.$$

- **Masking:**

$$\hat{x}_j = (1 - b_j) y_j + b_j f_\theta(y)_j.$$

$$\text{with } b_j = \frac{\max(0, |y_j| - \tau\mu)}{(1-\tau)\mu}.$$



Implementation details

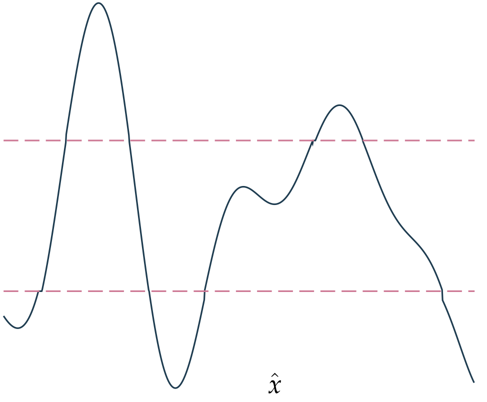
- **Bias-free** network:

$$f_\theta(\alpha y) = \alpha f_\theta(y) \quad \forall \alpha \in \mathbb{R}_+.$$

- **Masking:**

$$\hat{x}_j = (1 - b_j) y_j + b_j f_\theta(y)_j.$$

$$\text{with } b_j = \frac{\max(0, |y_j| - \tau\mu)}{(1-\tau)\mu}.$$



Experiments

Experiment: synthetic dataset

$$\mathrm{SDR}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 20 \log_{10} \left(\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2} \right) .$$

Experiment: synthetic dataset

$$\mathrm{SDR}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 20 \log_{10} \left(\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2} \right).$$

Synthetic dataset:

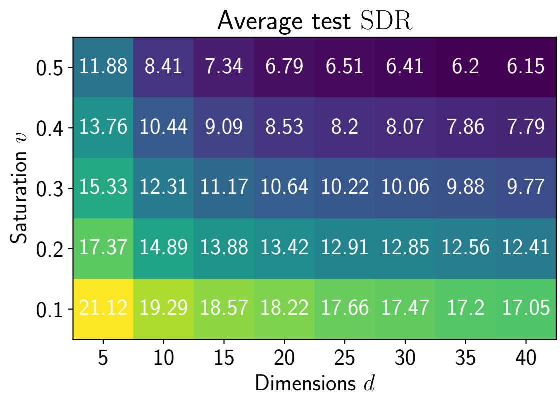
- Generate a random universal subspace of $\mathbb{R}^{100}$ with dimension d.
- Sample vectors from the subspace.
- Rescale vectors by a factor so that a proportion $v \in (0, 1)$ of measurements are clipped.

Experiment: synthetic dataset

$$\mathrm{SDR}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 20 \log_{10} \left(\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2} \right).$$

Synthetic dataset:

- Generate a random universal subspace of $\mathbb{R}^{100}$ with dimension d.
- Sample vectors from the subspace.
- Rescale vectors by a factor so that a proportion $v \in (0, 1)$ of measurements are clipped.



Experiments: real dataset

Real dataset:

Musics dataset: 21898 sounds of 1 sec each, 22050Hz.

Methods	SDR (dB)
Identity	5.28 ± 2.23
Social Sparsity	10.25 ± 4.64
Supervised	12.00 ± 2.31
Unsupervised (ours)	11.41 ± 3.02

Experiments: real dataset

Real dataset:

Musics dataset: 21898 sounds of 1 sec each, 22050Hz.

Methods	SDR (dB)
Identity	5.28 ± 2.23
Social Sparsity	10.25 ± 4.64
Supervised	12.00 ± 2.31
Unsupervised (ours)	11.41 ± 3.02

Robustness to distribution shift:

- **Supervised** train on music dataset (with ground-truth).
- **Unsupervised** train on music + voice dataset (without ground truth).

Methods	SDR (dB)
Identity	6.54 ± 2.34
Supervised	10.94 ± 2.00
Unsupervised (ours)	11.92 ± 2.46

Future work

- **Theoretical results:** signal recovery and model identification.
- **Extend to images:**