Self-supervised learning for phase retrieval problem

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

[1] *CNRS, ENS Lyon*
[2] *UCLouvain, Louvain-la-Neuve*

GRETSI'2025

Phase retrieval

What is phase retrieval?

$$\boldsymbol{y} = \left|\boldsymbol{A}\boldsymbol{x}\right|^2$$

Phase retrieval

What is phase retrieval?

$$\mathcal{Y} \subset \mathbb{R}^m \qquad \boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \qquad \mathcal{X} \subset \mathbb{C}^n$$

Phase retrieval

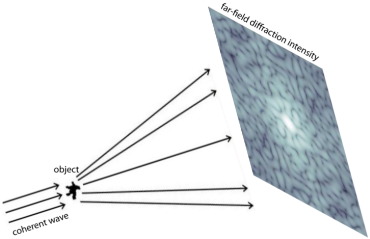
What is phase retrieval?

$$\mathcal{Y} \subset \mathbb{R}^m \qquad \boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \qquad \mathcal{X} \subset \mathbb{C}^n \qquad \mathbb{C}^{m \times n}$$

Phase retrieval

What is phase retrieval?

$$\mathcal{Y} \subset \mathbb{R}^m \qquad \boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \qquad \mathcal{X} \subset \mathbb{C}^n \qquad \mathbb{C}^{m \times n}$$

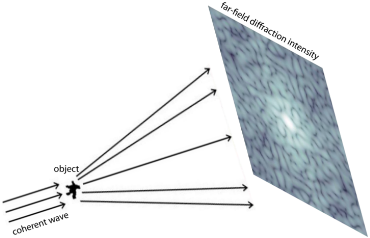


$\boldsymbol{A}$ the fourier transform.

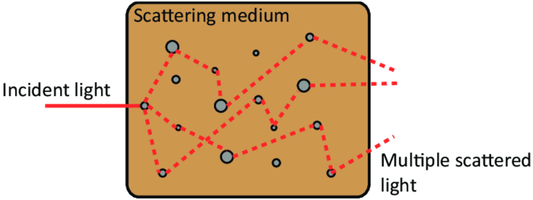
Phase retrieval

What is phase retrieval?

$$\mathcal{Y} \subset \mathbb{R}^m \quad \boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \quad \mathcal{X} \subset \mathbb{C}^n \quad \mathbb{C}^{m \times n}$$



$\boldsymbol{A}$ the fourier transform.

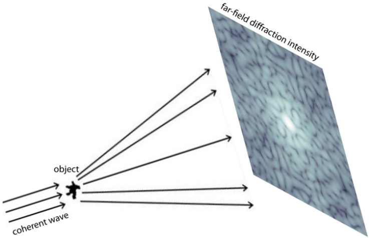


$\boldsymbol{A}$ random matrix.

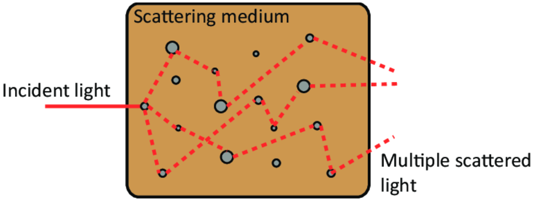
Phase retrieval

What is phase retrieval?

$$y = h(x)$$



A the fourier transform.



A random matrix.

Deep learning approches

- Deep learning is now the state of the art.

Deep learning approches

- Deep learning is now the state of the art.

Objective:

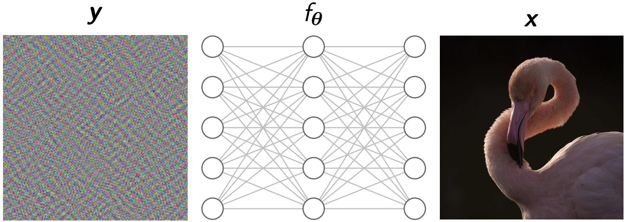
- **Learn** inverse forward operator of h, $f_{\boldsymbol{\theta}}(\boldsymbol{y}) \approx \boldsymbol{x}$.
- **Using** a dataset $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i \in I}$ + a neural network.

Deep learning approches

- Deep learning is now the state of the art.

Objective:

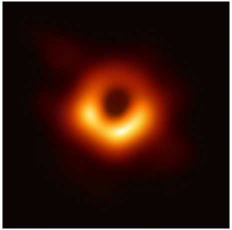
- **Learn** inverse forward operator of h, $f_{\boldsymbol{\theta}}(\boldsymbol{y}) \approx \boldsymbol{x}$.
- **Using** a dataset $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i \in I}$ + a neural network.



$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \| f_{\boldsymbol{\theta}}(\boldsymbol{y}_i) - \boldsymbol{x}_i \|^2$$

Limits of the method

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



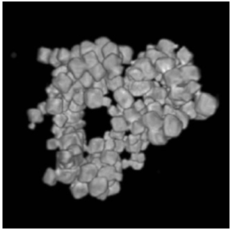


Figure: Astronomical and nanoscale imaging.

Unsupervised Learning, a challenge?

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

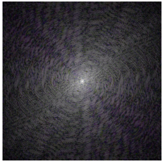
$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$
$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$

Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$
$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$

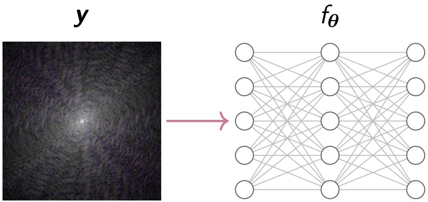
$\boldsymbol{y}$



Unsupervised Learning, a challenge?

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

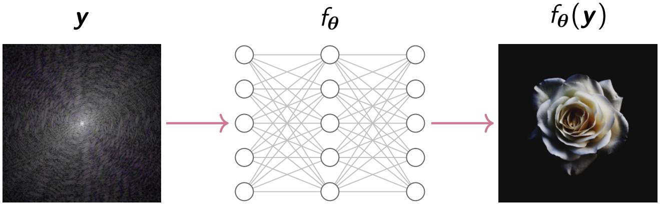
$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$
$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$



Unsupervised Learning, a challenge?

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

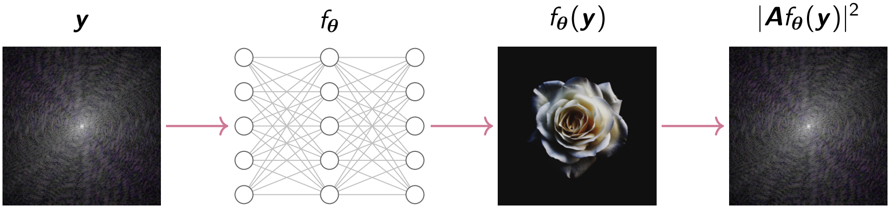
$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$
$$\arg\min_{\boldsymbol{\theta}} \sum_{i\in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$



Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$
$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \big\| \boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big) \big\|^2$$

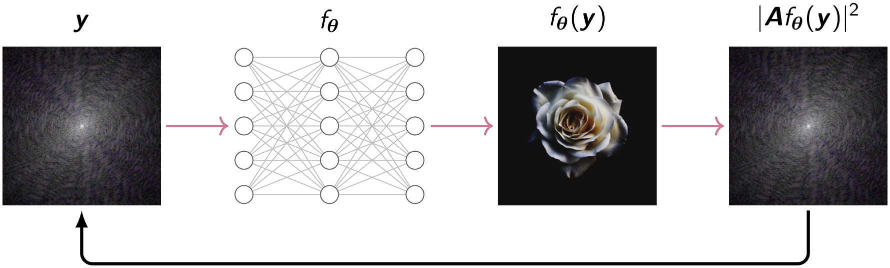


Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$

$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$

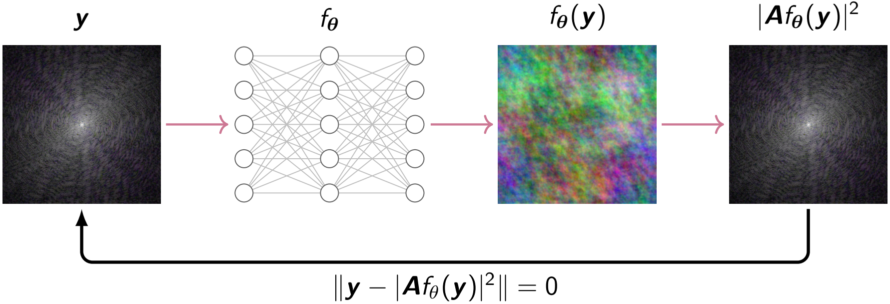


$$\|\boldsymbol{y} - |\boldsymbol{A} f_\theta(\boldsymbol{y})|^2\| = 0$$

Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$
$$\arg\min_{\boldsymbol{\theta}} \sum_{i\in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$

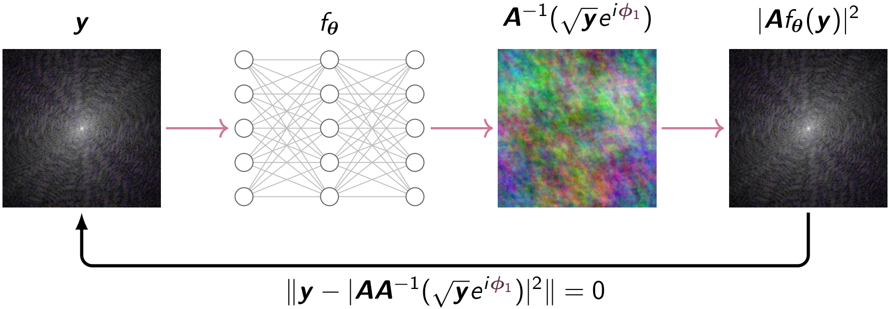


Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$

$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$

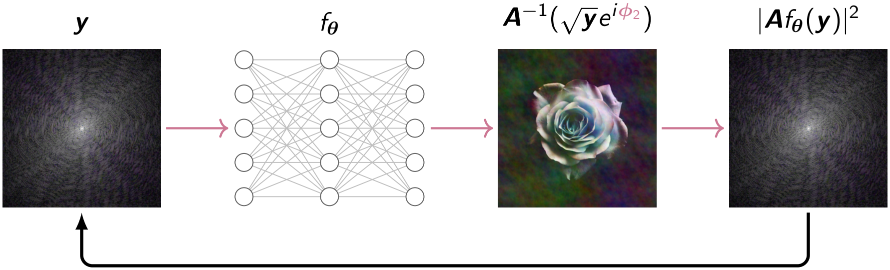


Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$

$$\arg\min_{\boldsymbol{\theta}} \sum_{i \in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$



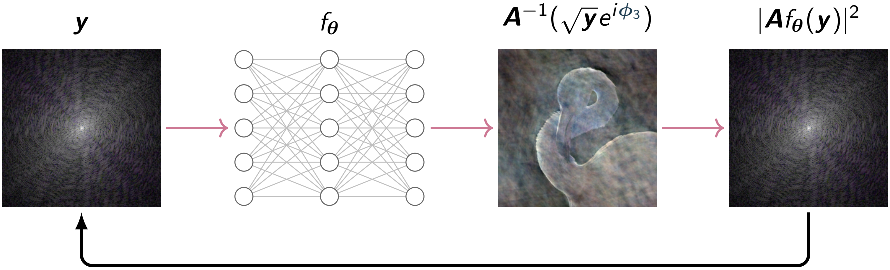
$$\|\boldsymbol{y} - |\boldsymbol{A}\boldsymbol{A}^{-1}(\sqrt{\boldsymbol{y}}e^{i\phi_2})|^2\| = 0$$

Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$

$$\arg\min_{\boldsymbol{\theta}} \sum_{i\in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$

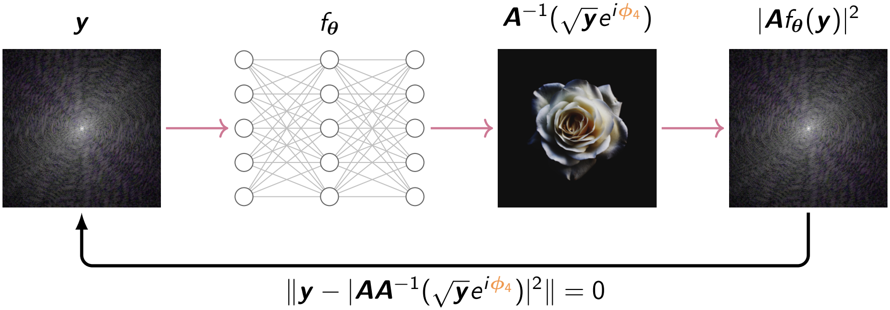


$$\|\boldsymbol{y} - |\boldsymbol{A}\boldsymbol{A}^{-1}(\sqrt{\boldsymbol{y}}e^{i\phi_3})|^2\| = 0$$

Unsupervised Learning, a challenge?

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

$$\boldsymbol{y} = h(\boldsymbol{x}) \Rightarrow \boldsymbol{y} \approx h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y})\big)$$
$$\arg\min_{\boldsymbol{\theta}} \sum_{i\in I} \|\boldsymbol{y}_i - h\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i)\big)\|^2$$



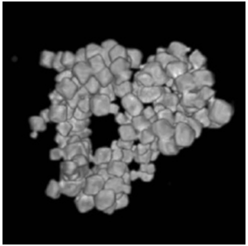
Invariance

- **A priori**: Set of signals invariant by transformation: $\forall g \in G, \forall \boldsymbol{x} \in \mathcal{X} : \boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$

Invariance

- **A priori**: Set of signals invariant by transformation: $\forall g \in G, \forall \boldsymbol{x} \in \mathcal{X} : \boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$

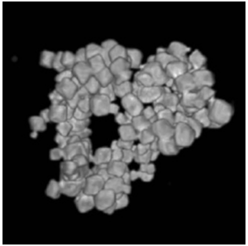
$\boldsymbol{x} \in \mathcal{X}$



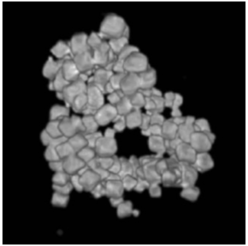
Invariance

- **A priori**: Set of signals invariant by transformation: $\forall g \in G, \forall \boldsymbol{x} \in \mathcal{X} : \boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$

$\boldsymbol{x} \in \mathcal{X}$



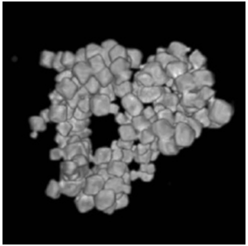
$\boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$



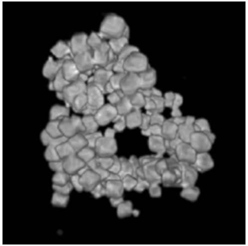
Invariance

- **A priori**: Set of signals invariant by transformation: $\forall g \in G, \forall \boldsymbol{x} \in \mathcal{X} : \boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$

$\boldsymbol{x} \in \mathcal{X}$



$\boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$



$$\boldsymbol{y} = h(\boldsymbol{x}) = h(\boldsymbol{T}_g \boldsymbol{T}_g^{-1} \boldsymbol{x}) = h_g(\boldsymbol{x}')$$

Equivariance

- **A priori**: Set of signals invariant by transformation.

Equivariance

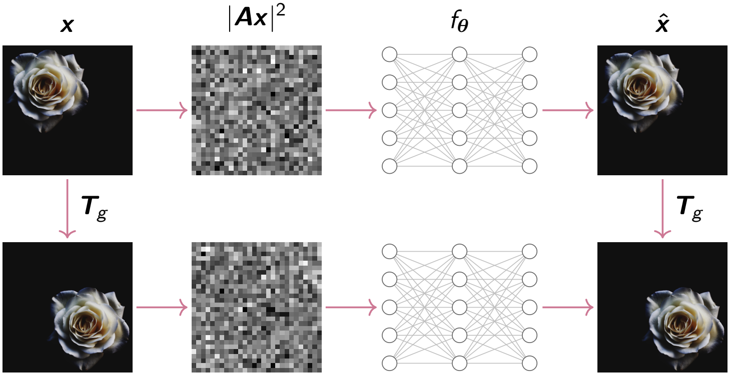
- **A priori**: Set of signals invariant by transformation.
- **Exploiting this information:**

$$\boxed{\boldsymbol{T}_g f_{\boldsymbol{\theta}}(h(\boldsymbol{x})) = f_{\theta}(h(\boldsymbol{T}_g \boldsymbol{x}))}$$

Equivariance

- **A priori**: Set of signals invariant by transformation.
- **Exploiting this information:**

$$\boxed{\boldsymbol{T}_g f_{\boldsymbol{\theta}}(h(\boldsymbol{x})) = f_{\boldsymbol{\theta}}(h(\boldsymbol{T}_g \boldsymbol{x}))}$$



Proposed method

Equivariance promoting loss:

$$\mathcal{L}_{\mathrm{EI}}(\boldsymbol{\theta}) = -\sum_{i \in I} \sum_{g \in G} \mathrm{CS}\Big(\boldsymbol{T}_g f_{\boldsymbol{\theta}}\left(\boldsymbol{y}_i\right), f_{\boldsymbol{\theta}}\left(h(\boldsymbol{T}_g f_{\boldsymbol{\theta}}\left(\boldsymbol{y}_i\right))\right)\Big)$$

$$\mathrm{CS}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{\left|\boldsymbol{x}^* \hat{\boldsymbol{x}}\right|}{\|\boldsymbol{x}\|_2 \|\hat{\boldsymbol{x}}\|_2}$$

Proposed method

Equivariance promoting loss:

$$\mathcal{L}_{\mathrm{EI}}(\boldsymbol{\theta}) = -\sum_{i \in I} \sum_{g \in G} \mathrm{CS}\Big(\boldsymbol{T}_g f_{\boldsymbol{\theta}}\left(\boldsymbol{y}_i\right), f_{\boldsymbol{\theta}}\left(h(\boldsymbol{T}_g f_{\boldsymbol{\theta}}\left(\boldsymbol{y}_i\right))\right)\Big)$$

$$\mathrm{CS}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{\left|\boldsymbol{x}^* \hat{\boldsymbol{x}}\right|}{\|\boldsymbol{x}\|_2 \|\hat{\boldsymbol{x}}\|_2}$$

Measurement consistency loss:

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

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

Proposed method

Equivariance promoting loss:

$$\mathcal{L}_{\mathrm{EI}}(\boldsymbol{\theta}) = -\sum_{i \in I} \sum_{g \in G} \mathrm{CS}\Big(\boldsymbol{T}_g f_{\boldsymbol{\theta}}\left(\boldsymbol{y}_i\right), f_{\boldsymbol{\theta}}\left(h(\boldsymbol{T}_g f_{\boldsymbol{\theta}}\left(\boldsymbol{y}_i\right))\right)\Big)$$

$$\mathrm{CS}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{|\boldsymbol{x}^* \hat{\boldsymbol{x}}|}{\|\boldsymbol{x}\|_2 \|\hat{\boldsymbol{x}}\|_2}$$

Measurement consistency loss:

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

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

Training loss: $\mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}_{\mathrm{A/I}}(\boldsymbol{\theta}) + \lambda \mathcal{L}_{\mathrm{EI}}(\boldsymbol{\theta})$.

Experiment

Experiments

- **"Phase MNIST"** dataset:
  $\boldsymbol{x} = e^{i\boldsymbol{x}_0}$ where $\boldsymbol{x}_0$ is an MNIST image.
- $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ **random** matrix with:

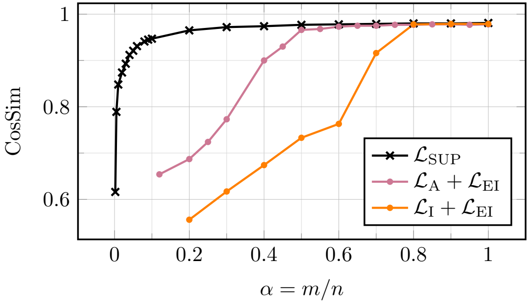
$$a_{k,l} \sim \mathcal{N}(0, 1/2m) + i\mathcal{N}(0, 1/2m).$$

- $\alpha = m/n$ **sampling rate**.
- Measurements: $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$.

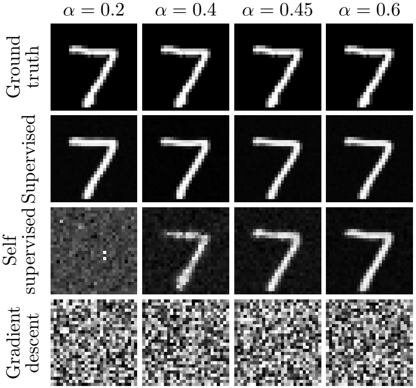
Implementation details

- **Loss parameter** $\lambda = 1$
- **image size** $n = 784$
- **Transformation** $\boldsymbol{T}_g$: shift
- Optimisation: **Learning rate**: $5e^{-5}$, **batch size**: 5.

Results



CosSim
1
0.8
0.6
0
0.2
0.4
0.6
0.8
1
$\alpha = m/n$
$\mathcal{L}_{\mathrm{SUP}}$
$\mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{EI}}$
$\mathcal{L}_{\mathrm{I}} + \mathcal{L}_{\mathrm{EI}}$



$\alpha = 0.2$
$\alpha = 0.4$
$\alpha = 0.45$
$\alpha = 0.6$
Ground truth
Supervised
Self supervised
Gradient descent

Conclusion

Future works

- **Improve** the performance for **low** sampling rate.
- **Adapt** the method to other linear forward operators (e.g. **Fourier** transform).
- **Explain** the difference between the **amplitude** $\mathcal{L}_{\mathrm{A}}$ and **intensity** $\mathcal{L}_{\mathrm{I}}$ loss.
- Provide **theoretical conditions** for learning depending on the **signal set** $\mathcal{X}$ and **sampling rate** α.