Does deep learning solve the phase retrieval problem?

Ju Sun
Computer Science & Engineering
University of Minnesota, Twin Cities

July 17, 2020

Thanks to

Raunak Manekar

UMN

Kshitij Tayal

UMN

Chieh-Hsin Lai

UMN

Vipin Kumar

UMN

Stefano Marchesini

LBNL

More details in

- **Inverse Problems, Deep Learning, and Symmetry Breaking**
  Kshitij Tayal, Chieh-Hsin Lai, Raunak Manekar, Vipin Kumar, Ju Sun. ICML workshop on ML Interpretability for Scientific Discovery, 2020.
  https://sunju.org/pub/ICML20-WS-DL4INV.pdf
- **End-to-End Learning for Phase Retrieval**
  Raunak Manekar, Kshitij Tayal, Vipin Kumar, Ju Sun. ICML workshop on ML Interpretability for Scientific Discovery, 2020.
  https://sunju.org/pub/ICML20-WS-DL4FPR.pdf

More details in

- **Inverse Problems, Deep Learning, and Symmetry Breaking**
  Kshitij Tayal, Chieh-Hsin Lai, Raunak Manekar, Vipin Kumar, Ju Sun. ICML workshop on ML Interpretability for Scientific Discovery, 2020.
  https://sunju.org/pub/ICML20-WS-DL4INV.pdf
- **End-to-End Learning for Phase Retrieval**
  Raunak Manekar, Kshitij Tayal, Vipin Kumar, Ju Sun. ICML workshop on ML Interpretability for Scientific Discovery, 2020.
  https://sunju.org/pub/ICML20-WS-DL4FPR.pdf

see also

Phase Retrieval via Second-Order Nonsmooth Optimization
Zhong Zhuang, Gang Wang, Yash Travadi, Ju Sun. ICML workshop on Beyond First Order Methods in Machine Learning, 2020.
https://sunju.org/pub/ICML20-WS-ALM-FPR.pdf

DL for inverse problems

Given $\boldsymbol{y} = f(\boldsymbol{x})$, estimate $\boldsymbol{x}$ $\quad$ (f may be unknown)

– Traditional

$$\min_{\boldsymbol{x}} \ \ell(\boldsymbol{y}, f(\boldsymbol{x})) + \lambda \Omega(\boldsymbol{x})$$

DL for inverse problems

Given $\boldsymbol{y} = f(\boldsymbol{x})$, estimate $\boldsymbol{x}$ (f may be unknown)

- Traditional

$$\min_{\boldsymbol{x}} \; \ell(\boldsymbol{y}, f(\boldsymbol{x})) + \lambda \Omega(\boldsymbol{x})$$

- Modern
  * End-to-end: set up $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ to learn f^{-1} directly

DL for inverse problems

Given $\boldsymbol{y} = f(\boldsymbol{x})$, estimate $\boldsymbol{x}$ (f may be unknown)

- Traditional

$$\min_{\boldsymbol{x}} \ \ell(\boldsymbol{y}, f(\boldsymbol{x})) + \lambda \Omega(\boldsymbol{x})$$

- Modern
    * End-to-end: set up $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ to learn f^{-1} directly
    * Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA

DL for inverse problems

Given $\boldsymbol{y} = f(\boldsymbol{x})$, estimate $\boldsymbol{x}$ (f may be unknown)

- Traditional

$$\min_{\boldsymbol{x}} \ \ell(\boldsymbol{y}, f(\boldsymbol{x})) + \lambda \Omega(\boldsymbol{x})$$

- Modern
  * End-to-end: set up $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ to learn f^{-1} directly
  * Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA
- "Modern" works **better** when "traditional" **already works**

DL for inverse problems

Given $\boldsymbol{y} = f(\boldsymbol{x})$, estimate $\boldsymbol{x}$ (f may be unknown)

- Traditional

$$\min_{\boldsymbol{x}} \ \ell(\boldsymbol{y}, f(\boldsymbol{x})) + \lambda \Omega(\boldsymbol{x})$$

- Modern
  * End-to-end: set up $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ to learn f^{-1} directly
  * Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA
- "Modern" works **better** when "traditional" **already works**

Recent surveys: [McCann et al., 2017, Lucas et al., 2018, Arridge et al., 2019, Ongie et al., 2020]

Why DL for PR?

PR is difficult

Why DL for PR?

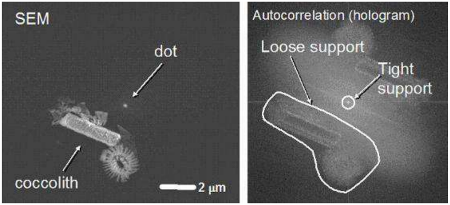
PR is difficult

- Most traditional methods fail
  * Effective methods: **proximal methods** [Luke et al., 2019]
  * Exceptions: saddle point optimization [Marchesini, 2007], **2nd order ALM** [Zhuang et al., 2020]

Why DL for PR?

PR is difficult

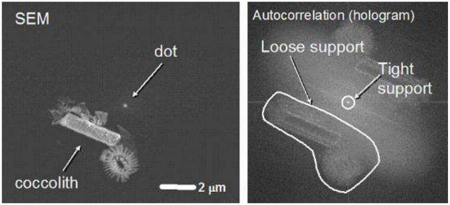
- Most traditional methods fail
  - * Effective methods: **proximal methods** [Luke et al., 2019]
  - * Exceptions: saddle point optimization [Marchesini, 2007], **2nd order ALM** [Zhuang et al., 2020]
- Complex-valued without accurate support info, e.g., [Marchesini et al., 2005]



Why DL for PR?

PR is difficult

- Most traditional methods fail
  * Effective methods: **proximal methods** [Luke et al., 2019]
  * Exceptions: saddle point optimization [Marchesini, 2007], **2nd order ALM** [Zhuang et al., 2020]
- Complex-valued without accurate support info, e.g., [Marchesini et al., 2005]



- Low-photon regime, beam stop, etc, e.g., [Chang et al., 2018]

How?

- Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA

  "modern" works **better** when traditionally **already works**

How?

– Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA

"modern" works **better** when traditionally **already works**

Attempts: [Metzler et al., 2018, Işıl et al., 2019], but HIO needed for initialization

How?

- Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA

  "modern" works **better** when traditionally **already works**

  Attempts: [Metzler et al., 2018, Işıl et al., 2019], but HIO needed for initialization
- End-to-end: set up $\{\boldsymbol{x}_i, |\mathcal{F}\boldsymbol{x}|^2\}$ to learn f^{-1} directly

How?

- Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA

  "modern" works **better** when traditionally **already works**

  Attempts: [Metzler et al., 2018, Işıl et al., 2019], but HIO needed for initialization

- End-to-end: set up $\{\boldsymbol{x}_i, |\mathcal{F}\boldsymbol{x}|^2\}$ to learn f^{-1} directly

  Attempts:
  [Goy et al., 2018, Uelwer et al., 2019, Metzler et al., 2020]
  with positive initial results

How?

- Hybrid: replace ℓ, Ω, or algorithmic components using **learned functions**, e.g., plug-and-play ADMM, unrolling ISTA

  "modern" works **better** when traditionally **already works**

  Attempts: [Metzler et al., 2018, Işıl et al., 2019], but HIO needed for initialization

- End-to-end: set up $\{\boldsymbol{x}_i, |\mathcal{F}\boldsymbol{x}|^2\}$ to learn f^{-1} directly

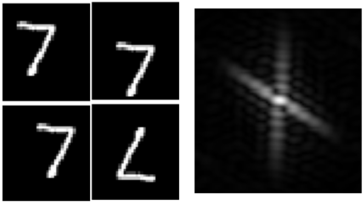
  Attempts:

  [Goy et al., 2018, Uelwer et al., 2019, Metzler et al., 2020]

  with positive initial results

Focus of this talk: **end-to-end approach**

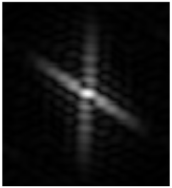
How good are they?



Symmetries in Fourier PR:

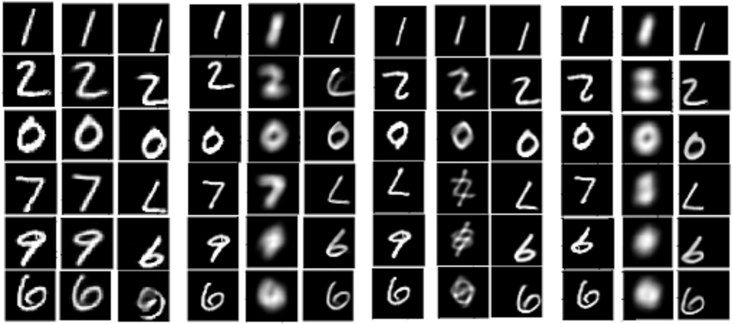
- shift
- 2D flipping
- global phase

How good are they?



Symmetries in Fourier PR:

- shift
- 2D flipping
- global phase



(a) No Symmetry

(b) Shift symmetry

(c) Flipping symmetry

(d) Shift and Flipping symmetries

Why (over)-optimistic results in practice?

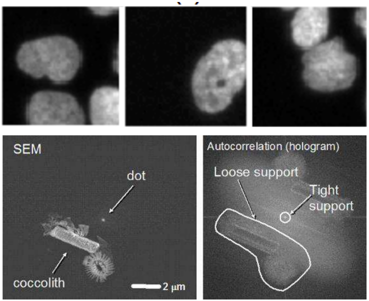
Data! Data! Data!

Why (over)-optimistic results in practice?

Data! Data! Data!

experimental data

naturally oriented and centered



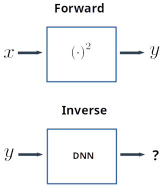
practical data

no natural orientation or centering

Dataset bias breaks problem symmetries

Why learning with symmetries is difficult?

Learning square roots!

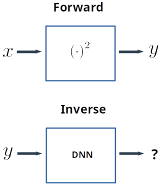


Original Dataset

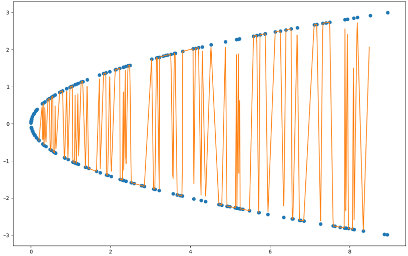
4	-2
9	3
16	-4
100	10
4	2
64	-8
36	6

Why learning with symmetries is difficult?

Learning square roots!

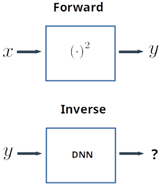


Original Dataset	
4	-2
9	3
16	-4
100	10
4	2
64	-8
36	6



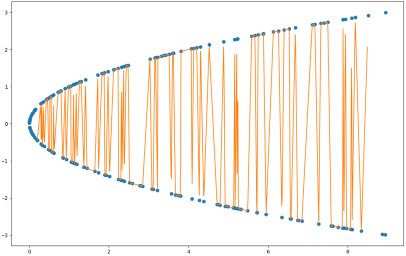
Why learning with symmetries is difficult?

Learning square roots!



Original Dataset

4	-2
9	3
16	-4
100	10
4	2
64	-8
36	6



nearby inputs mapped to remote outputs **due to symmetries**

Other examples

$\boldsymbol{y} = f(\boldsymbol{x})$ with f a many-to-one mapping

- symmetries in f

  - **Fourier phase retrieval** [BBE17] The forward model is $\boldsymbol{Y} = |\mathcal{F}(\boldsymbol{X})|^2$, where $\boldsymbol{X} \in \mathbb{C}^{n \times n}$ and $\boldsymbol{Y} \in \mathbb{R}^{m \times m}$ are matrices and $\mathcal{F}$ is a 2D oversampled Fourier transform. The operation $|\cdot|$ takes complex magnitudes of the entries elementwise. It is known that translations and conjugate flippings applied on $\boldsymbol{X}$, and also global phase transfer of the form $e^{i\theta}\boldsymbol{X}$ all lead to the same $\boldsymbol{Y}$.
  - **Blind deconvolution** [LG00, TB10] The forward model is $\boldsymbol{y} = \boldsymbol{a} \circledast \boldsymbol{x}$, where $\boldsymbol{a}$ is the convolution kernel, $\boldsymbol{x}$ is the signal (e.g., image) of interest, and $\circledast$ denotes the circular convolution. Both $\boldsymbol{a}$ and $\boldsymbol{x}$ are inputs. Here, $\boldsymbol{a} \circledast \boldsymbol{x} = (\lambda \boldsymbol{a}) \circledast (\boldsymbol{x}/\lambda)$ for any $\lambda \neq 0$, and circularly shifting $\boldsymbol{a}$ to the left and shifting $\boldsymbol{x}$ to the right by the same amount does not change $\boldsymbol{y}$.
  - **Synchronization over compact groups** [PWBM18] For $g_1, \dots, g_n$ over a compact group $\mathcal{G}$, the observation is a set of pairwise relative measurements $y_{ij} = g_i g_j^{-1}$ for all (i, j) in an index set $\mathcal{E} \subset \{1, \dots, n\} \times \{1, \dots, n\}$. Obviously, any global shift of the form $g_k \mapsto g_k g$ for all $k \in \{1, \dots, n\}$, for any $g \in \mathcal{G}$, leads to the same set of measurements.

Other examples

$\boldsymbol{y} = f(\boldsymbol{x})$ with f a many-to-one mapping

- symmetries in f

  - **Fourier phase retrieval** [BBE17] The forward model is $\boldsymbol{Y} = |\mathcal{F}(\boldsymbol{X})|^2$, where $\boldsymbol{X} \in \mathbb{C}^{n\times n}$ and $\boldsymbol{Y} \in \mathbb{R}^{m\times m}$ are matrices and $\mathcal{F}$ is a 2D oversampled Fourier transform. The operation $|\cdot|$ takes complex magnitudes of the entries elementwise. It is known that translations and conjugate flippings applied on $\boldsymbol{X}$, and also global phase transfer of the form $e^{i\theta}\boldsymbol{X}$ all lead to the same $\boldsymbol{Y}$.
  - **Blind deconvolution** [LG00, TB10] The forward model is $\boldsymbol{y} = \boldsymbol{a} \circledast \boldsymbol{x}$, where $\boldsymbol{a}$ is the convolution kernel, $\boldsymbol{x}$ is the signal (e.g., image) of interest, and $\circledast$ denotes the circular convolution. Both $\boldsymbol{a}$ and $\boldsymbol{x}$ are inputs. Here, $\boldsymbol{a} \circledast \boldsymbol{x} = (\lambda \boldsymbol{a}) \circledast (\boldsymbol{x}/\lambda)$ for any $\lambda \neq 0$, and circularly shifting $\boldsymbol{a}$ to the left and shifting $\boldsymbol{x}$ to the right by the same amount does not change $\boldsymbol{y}$.
  - **Synchronization over compact groups** [PWBM18] For $g_1, \ldots, g_n$ over a compact group $\mathcal{G}$, the observation is a set of pairwise relative measurements $y_{ij} = g_i g_j^{-1}$ for all (i,j) in an index set $\mathcal{E} \subset \{1,\ldots,n\} \times \{1,\ldots,n\}$. Obviously, any global shift of the form $g_k \mapsto g_k g$ for all $k \in \{1,\ldots,n\}$, for any $g \in \mathcal{G}$, leads to the same set of measurements.

- nontrivial kernel space, e.g. subsampled MRI imaging, e.g., [Gottschling et al., 2020]

Other examples

$\boldsymbol{y} = f(\boldsymbol{x})$ with f a many-to-one mapping

– symmetries in f

- **Fourier phase retrieval** [BBE17] The forward model is $\boldsymbol{Y} = |\mathcal{F}(\boldsymbol{X})|^2$, where $\boldsymbol{X} \in \mathbb{C}^{n \times n}$ and $\boldsymbol{Y} \in \mathbb{R}^{m \times m}$ are matrices and $\mathcal{F}$ is a 2D oversampled Fourier transform. The operation $|\cdot|$ takes complex magnitudes of the entries elementwise. It is known that translations and conjugate flippings applied on $\boldsymbol{X}$, and also global phase transfer of the form $e^{i\theta}\boldsymbol{X}$ all lead to the same $\boldsymbol{Y}$.
- **Blind deconvolution** [LG00, TB10] The forward model is $\boldsymbol{y} = \boldsymbol{a} \circledast \boldsymbol{x}$, where $\boldsymbol{a}$ is the convolution kernel, $\boldsymbol{x}$ is the signal (e.g., image) of interest, and $\circledast$ denotes the circular convolution. Both $\boldsymbol{a}$ and $\boldsymbol{x}$ are inputs. Here, $\boldsymbol{a} \circledast \boldsymbol{x} = (\lambda \boldsymbol{a}) \circledast (\boldsymbol{x}/\lambda)$ for any $\lambda \neq 0$, and circularly shifting $\boldsymbol{a}$ to the left and shifting $\boldsymbol{x}$ to the right by the same amount does not change $\boldsymbol{y}$.
- **Synchronization over compact groups** [PWBM18] For $g_1, \dots, g_n$ over a compact group $\mathcal{G}$, the observation is a set of pairwise relative measurements $y_{ij} = g_i g_j^{-1}$ for all (i, j) in an index set $\mathcal{E} \subset \{1, \dots, n\} \times \{1, \dots, n\}$. Obviously, any global shift of the form $g_k \mapsto g_k g$ for all $k \in \{1, \dots, n\}$, for any $g \in \mathcal{G}$, leads to the same set of measurements.

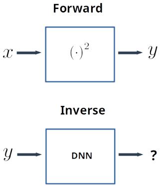
– nontrivial kernel space, e.g. subsampled MRI imaging, e.g., [Gottschling et al., 2020]

Inverse f^{-1} is one-to-many mapping

Get rid of the difficulty?

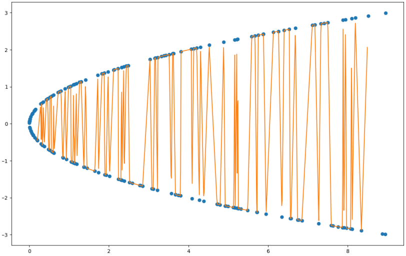
- active symmetry breaking
- passive symmetry breaking

An easy solution to the square root example

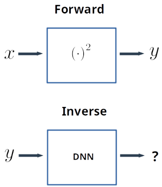


Original Dataset

4	-2
9	3
16	-4
100	10
4	2
64	-8
36	6

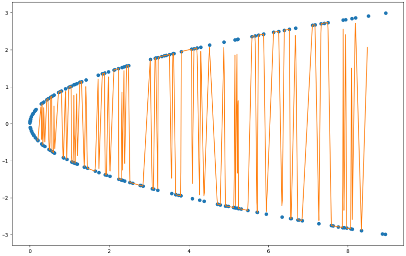


An easy solution to the square root example



Original Dataset

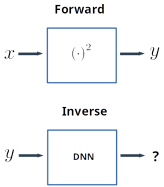
4	-2
9	3
16	-4
100	10
4	2
64	-8
36	6



Modified Dataset

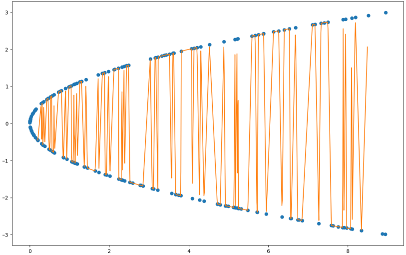
4	2
9	3
16	4
100	10
4	2
64	8
36	6

An easy solution to the square root example



Original Dataset

4	-2
9	3
16	-4
100	10
4	2
64	-8
36	6



Modified Dataset

4	2
9	3
16	4
100	10
4	2
64	8
36	6

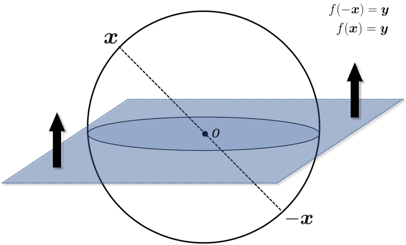
idea: fix the sign symmetry

Active symmetry breaking

Real Gaussian PR: $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$ for illustration

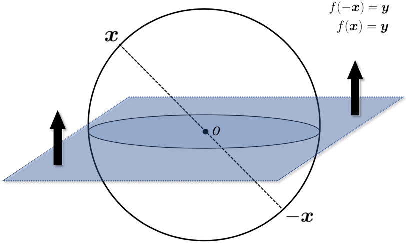
Active symmetry breaking

Real Gaussian PR: $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$ for illustration



Active symmetry breaking

Real Gaussian PR: $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$ for illustration



find a **smallest**, **representative**, and **connected** subset [Tayal et al., 2020]

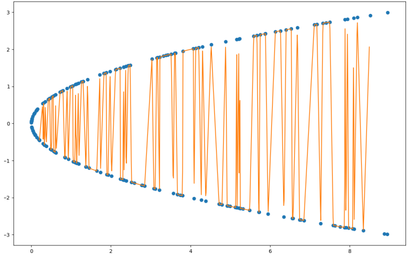
Does it work?

n	Sample	NN-A	K-NN	NN-B	WNN-A	K-NN	WNN-B	DNN-A	K-NN	DNN-B
5	2e4	10	17	283	**8**	18	283	10	19	284
	5e4	**6**	12	282	8	17	284	7	14	285
	1e5	**5**	10	284	**5**	12	283	13	18	284
	1e6	**4**	7	283	5	6	283	7	8	283
10	2e4	11	20	82	9	22	82	**8**	21	82
	5e4	9	16	82	**6**	18	82	9	20	82
	1e5	9	16	82	**6**	15	82	8	17	82
	1e6	7	13	82	**5**	10	82	9	11	82
15	2e4	12	17	38	**9**	16	38	**9**	16	38
	5e4	11	14	38	9	14	38	**8**	15	38
	1e5	10	13	38	8	13	38	**7**	13	38
	1e6	8	9	38	**7**	10	38	9	10	38

NN-A: **after** symmetry breaking — **more is worse**

NN-B: **before** symmetry breaking

K-NN: K-nearest neighbor baseline



More developments

– Complex Gaussian PR [Tayal et al., 2020]

More developments

- Complex Gaussian PR [Tayal et al., 2020]
- (Fourier) PR: forthcoming

More developments

- Complex Gaussian PR [Tayal et al., 2020]
- (Fourier) PR: forthcoming

Pros: 1) math. principled 2) only symmetry info needed even if f unknown [Krippendorf and Syvaeri, 2020]

Cons: math. involved

Math-free alternative?

passive symmetry breaking

- If $\mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i) \approx \boldsymbol{x}_i$, then $|\mathcal{F} \circ \mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i)| \approx |\mathcal{F}\boldsymbol{x}_i| = \boldsymbol{y}_i$

Math-free alternative?

passive symmetry breaking

- If $\mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i) \approx \boldsymbol{x}_i$, then $|\mathcal{F} \circ \mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i)| \approx |\mathcal{F}\boldsymbol{x}_i| = \boldsymbol{y}_i$
- Consider

$$\min_{\boldsymbol{W}} \sum_i \ell\left(\boldsymbol{y}_i, |\mathcal{F} \circ \mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i)|\right)$$

Math-free alternative?

passive symmetry breaking

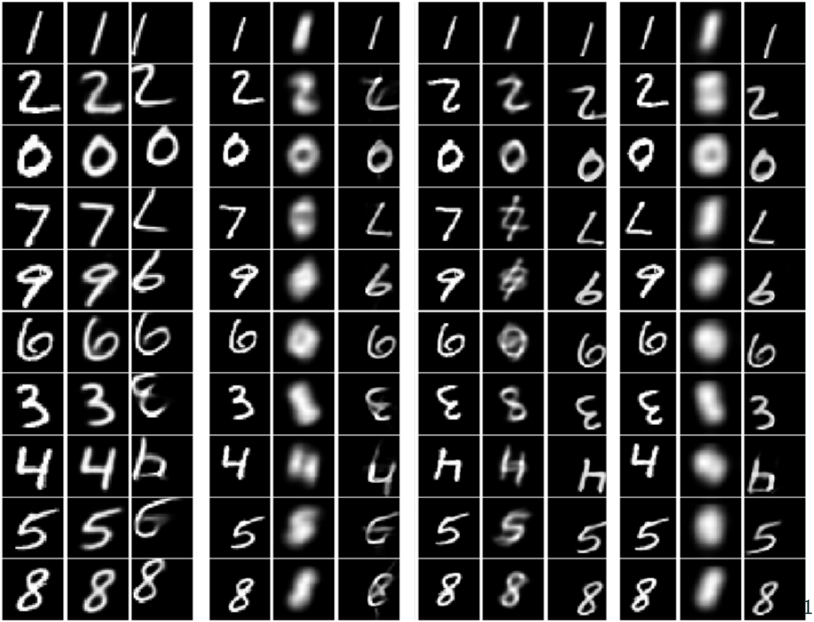
- If $\mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i) \approx \boldsymbol{x}_i$, then $|\mathcal{F} \circ \mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i)| \approx |\mathcal{F}\boldsymbol{x}_i| = \boldsymbol{y}_i$
- Consider

$$\min_{\boldsymbol{W}} \sum_i \ell\left(\boldsymbol{y}_i, |\mathcal{F} \circ \mathrm{DNN}_{\boldsymbol{W}}(\boldsymbol{y}_i)|\right)$$

- Why it might work?
  * $\mathrm{DNN}_{\boldsymbol{W}}$ is simple when symmetries are broken
  * **implicit regularization** means simple $\mathrm{DNN}_{\boldsymbol{W}}$ is preferred

similar idea appears in [Metzler et al., 2020]

Does it work?



More developments

– Complex-valued images

More developments

- Complex-valued images
- Other datasets

More developments

- Complex-valued images
- Other datasets

Pros: 1) lightweight 2) general

Cons: 1) f is needed 2) dense data needed (?)

Contribution

active and passive symmetry breaking for PR (and general inverse problems)

Contribution

active and passive symmetry breaking for PR (and general inverse problems)

- End-to-end learning offers new opportunities for solving difficult PR instances
- Current successes are contaminated by dataset biases
- Symmetry breaking seems to offer a way out

[Arridge et al., 2019] Arridge, S., Maass, P., Öktem, O., and Schönlieb, C.-B. (2019). **Solving inverse problems using data-driven models.** *Acta Numerica*, 28:1–174.

[Chang et al., 2018] Chang, H., Lou, Y., Duan, Y., and Marchesini, S. (2018). **Total variation–based phase retrieval for poisson noise removal.** *SIAM Journal on Imaging Sciences*, 11(1):24–55.

[Gottschling et al., 2020] Gottschling, N. M., Antun, V., Adcock, B., and Hansen, A. C. (2020). **The troublesome kernel: why deep learning for inverse problems is typically unstable.** *arXiv:2001.01258*.

[Goy et al., 2018] Goy, A., Arthur, K., Li, S., and Barbastathis, G. (2018). **Low photon count phase retrieval using deep learning.** *Physical Review Letters*, 121(24).

[Işıl et al., 2019] Işıl, Ç., Oktem, F. S., and Koç, A. (2019). **Deep iterative reconstruction for phase retrieval.** *Applied Optics*, 58(20):5422.

[Krippendorf and Syvaeri, 2020] Krippendorf, S. and Syvaeri, M. (2020). **Detecting symmetries with neural networks.** *arXiv preprint arXiv:2003.13679*.

[Lucas et al., 2018] Lucas, A., Iliadis, M., Molina, R., and Katsaggelos, A. K. (2018). **Using deep neural networks for inverse problems in imaging: Beyond analytical methods.** *IEEE Signal Processing Magazine*, 35(1):20–36.

[Luke et al., 2019] Luke, D. R., Sabach, S., and Teboulle, M. (2019). **Optimization on spheres: Models and proximal algorithms with computational performance comparisons.** *SIAM Journal on Mathematics of Data Science*, 1(3):408–445.

[Marchesini, 2007] Marchesini, S. (2007). **Phase retrieval and saddle-point optimization.** *Journal of the Optical Society of America A*, 24(10):3289.

[Marchesini et al., 2005] Marchesini, S., Chapman, H. N., Barty, A., Cui, C., Howells, M. R., Spence, J. C. H., Weierstall, U., and Minor, A. M. (2005). **Phase aberrations in diffraction microscopy.** *arXiv:physics/0510033*.

[McCann et al., 2017] McCann, M. T., Jin, K. H., and Unser, M. (2017). **Convolutional neural networks for inverse problems in imaging: A review.** *IEEE Signal Processing Magazine*, 34(6):85–95.

[Metzler et al., 2020] Metzler, C. A., Heide, F., Rangarajan, P., Balaji, M. M., Viswanath, A., Veeraraghavan, A., and Baraniuk, R. G. (2020). **Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging.** *Optica*, 7(1):63.

[Metzler et al., 2018] Metzler, C. A., Schniter, P., Veeraraghavan, A., and Baraniuk, R. G. (2018). **prdeep: Robust phase retrieval with a flexible deep network.** *arXiv preprint arXiv:1803.00212*.

[Ongie et al., 2020] Ongie, G., Jalal, A., Metzler, C. A., Baraniuk, R. G., Dimakis, A. G., and Willett, R. (2020). **Deep learning techniques for inverse problems in imaging.** *arXiv:2005.06001*.

[Tayal et al., 2020] Tayal, K., Lai, C.-H., Kumar, V., and Sun, J. (2020). **Inverse problems, deep learning, and symmetry breaking.** *arXiv:2003.09077*.

[Uelwer et al., 2019] Uelwer, T., Oberstraß, A., and Harmeling, S. (2019). **Phase retrieval using conditional generative adversarial networks.** *arXiv:1912.04981*.

[Zhuang et al., 2020] Zhuang, Z., Wang, G., Travadi, Y., and Sun, J. (2020). **Phase retrieval via second-order nonsmooth optimization.** In *ICML workshopon Beyond First Order Methods in Machine Learning*.