Toward practical phase retrieval: To learn or not, and how to learn?

Ju Sun

Computer Science & Engineering

University of Minnesota, Twin Cities

UNIVERSITY OF MINNESOTA
Driven to Discover®
Crookston Duluth Morris Rochester Twin Cities

ORS 中国运筹学会 Operations Research Society of China

September 27, 2020

Thanks to

Raunak Manekar

UMN, CS&E

Kshitij Tayal

UMN, CS&E

Zhong Zhuang

UMN, ECE

Chieh-Hsin Lai

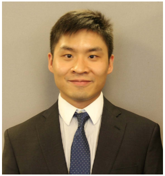
UMN, Math

Vipin Kumar

UMN, CS&E

Stefano Marchesini

LBNL



Gang Wang

UMN/BIT

Outline

Why phase retrieval?

How people solve PR?

Deep learning for PR?

Phase retrieval (PR): Given $|\mathcal{F}(\boldsymbol{x})|^2$, recover $\boldsymbol{x}$

Phase retrieval (PR): Given $|\mathcal{F}(\boldsymbol{x})|^2$, recover $\boldsymbol{x}$

– $\mathcal{F}$: (discrete) Fourier transform

Phase retrieval

Phase retrieval (PR): Given $|\mathcal{F}(\boldsymbol{x})|^2$, recover $\boldsymbol{x}$

- $\mathcal{F}$: (discrete) Fourier transform
- $\boldsymbol{x}$: 1D (vector), 2D (matrix), or 3D (tensor) signal

Phase retrieval

Phase retrieval (PR): Given $|\mathcal{F}(\boldsymbol{x})|^2$, recover $\boldsymbol{x}$

- $\mathcal{F}$: (discrete) Fourier transform
- $\boldsymbol{x}$: 1D (vector), 2D (matrix), or 3D (tensor) signal
- Without $|\cdot|^2$, a matter of $\mathcal{F}^{-1}$!

1D example: spectral factorization

In signal processing, control, and stochastic processes, etc: given an autocorrelation sequence $\boldsymbol{r} \in \mathbb{R}^{2n-1}$ and its Z transform $R(z)$

spectral factorization: find $X(z)$ so that $R(z) = \alpha X(z)\, X\left(z^{-1}\right)$ and $X(z)$ has all roots inside the unit circle.

1D example: spectral factorization

In signal processing, control, and stochastic processes, etc: given an autocorrelation sequence $\boldsymbol{r} \in \mathbb{R}^{2n-1}$ and its Z transform $R(z)$

spectral factorization: find $X(z)$ so that $R(z) = \alpha X(z) X\left(z^{-1}\right)$ and $X(z)$ has all roots inside the unit circle.

$$\Longleftrightarrow \text{ find } \boldsymbol{x} \in \mathbb{R}^n \text{ so that } \boldsymbol{r} = \boldsymbol{x} * \boldsymbol{x}$$

1D example: spectral factorization

In signal processing, control, and stochastic processes, etc: given an autocorrelation sequence $\boldsymbol{r} \in \mathbb{R}^{2n-1}$ and its Z transform $R(z)$

spectral factorization: find $X(z)$ so that $R(z) = \alpha X(z) X\left(z^{-1}\right)$ and $X(z)$ has all roots inside the unit circle.

$$\Longleftrightarrow \text{ find } \boldsymbol{x} \in \mathbb{R}^n \text{ so that } \boldsymbol{r} = \boldsymbol{x} * \boldsymbol{x}$$

$$\Longleftrightarrow \text{ find } \boldsymbol{x} \in \mathbb{R}^n \text{ so that } \mathcal{F}(\boldsymbol{r}) = \mathcal{F}(\boldsymbol{x} * \boldsymbol{x}) = |\mathcal{F}(\boldsymbol{x})|^2$$

1D example: spectral factorization

In signal processing, control, and stochastic processes, etc: given an autocorrelation sequence $\boldsymbol{r} \in \mathbb{R}^{2n-1}$ and its Z transform $R\left(z\right)$

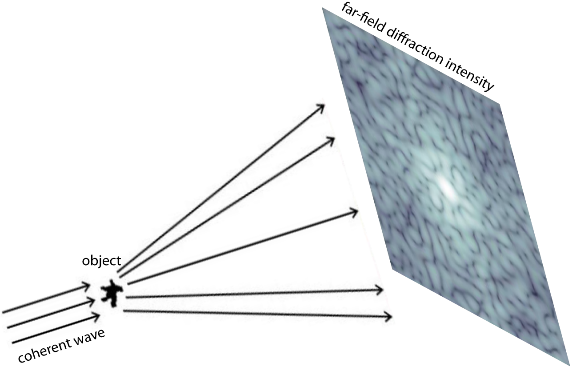
spectral factorization: find $X\left(z\right)$ so that $R\left(z\right) = \alpha X\left(z\right) X\left(z^{-1}\right)$ and $X(z)$ has all roots inside the unit circle.

$$\Longleftrightarrow \text{ find } \boldsymbol{x} \in \mathbb{R}^n \text{ so that } \boldsymbol{r} = \boldsymbol{x} * \boldsymbol{x}$$

$$\Longleftrightarrow \text{ find } \boldsymbol{x} \in \mathbb{R}^n \text{ so that } \mathcal{F}\left(\boldsymbol{r}\right) = \mathcal{F}\left(\boldsymbol{x} * \boldsymbol{x}\right) = \left|\mathcal{F}\left(\boldsymbol{x}\right)\right|^2$$

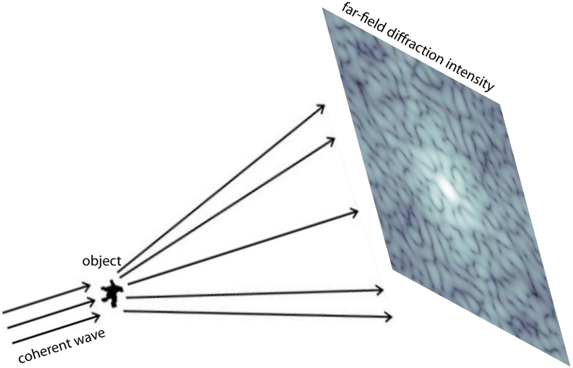
So: given $\left|\mathcal{F}\left(\boldsymbol{x}\right)\right|^2$, recover $\boldsymbol{x}$— 1D PR!

2D example: coherent diffraction imaging (CDI)



(Credit: [Shechtman et al., 2015])

2D example: coherent diffraction imaging (CDI)



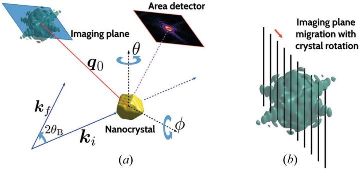
(Credit: [Shechtman et al., 2015])

Fraunhofer (far-field) approximation:

$$|f(x,y)|^2 \approx \frac{1}{\lambda^2 z^2} \left| \widehat{I}\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right) \right|^2,$$

where $I(x,y) = f(x,y,0)$ (complex-valued!).

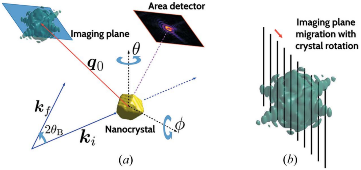
3D example: Bragg coherent diffraction imaging (BCDI)



single-reflection BCDI

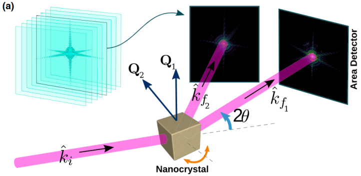
(Credit: [Maddali et al., 2020])

3D example: Bragg coherent diffraction imaging (BCDI)



single-reflection BCDI

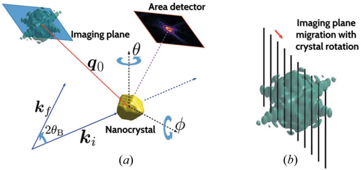
(Credit: [Maddali et al., 2020])



multi-reflection BCDI

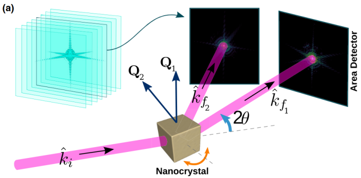
(Credit: [Newton, 2020])

3D example: Bragg coherent diffraction imaging (BCDI)



single-reflection BCDI

(Credit: [Maddali et al., 2020])



multi-reflection BCDI

(Credit: [Newton, 2020])

modern tools for x-ray crystallography, with application in chemistry, materials, medicine, etc

“Nobel-level problem”

Nobel Prizes involving X-ray crystallography [edit]

Year [hide]	Laureate	Prize	Rationale
1914	Max von Laue	Physics	"For his discovery of the diffraction of X-rays by crystals",[147] an important step in the development of X-ray spectroscopy.
1915	William Henry Bragg	Physics	"For their services in the analysis of crystal structure by means of X-rays"[148]
1915	William Lawrence Bragg	Physics	"For their services in the analysis of crystal structure by means of X-rays"[148]
1962	Max F. Perutz	Chemistry	"for their studies of the structures of globular proteins"[149]
1962	John C. Kendrew	Chemistry	"for their studies of the structures of globular proteins"[149]
1962	James Dewey Watson	Medicine	"For their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material"[150]
1962	Francis Harry Compton Crick	Medicine	"For their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material"[150]
1962	Maurice Hugh Frederick Wilkins	Medicine	"For their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material"[150]
1964	Dorothy Hodgkin	Chemistry	"For her determinations by X-ray techniques of the structures of important biochemical substances"[151]
1972	Stanford Moore	Chemistry	"For their contribution to the understanding of the connection between chemical structure and catalytic activity of the active centre of the ribonuclea
1972	William H. Stein	Chemistry	"For their contribution to the understanding of the connection between chemical structure and catalytic activity of the active centre of the ribonuclea
1976	William N. Lipscomb	Chemistry	"For his studies on the structure of boranes illuminating problems of chemical bonding"[153]
1985	Jerome Karle	Chemistry	"For their outstanding achievements in developing direct methods for the determination of crystal structures"[154]
1985	Herbert A. Hauptman	Chemistry	"For their outstanding achievements in developing direct methods for the determination of crystal structures"[154]
1988	Johann Deisenhofer	Chemistry	"For their determination of the three-dimensional structure of a photosynthetic reaction centre"[155]
1988	Hartmut Michel	Chemistry	"For their determination of the three-dimensional structure of a photosynthetic reaction centre"[155]
1988	Robert Huber	Chemistry	"For their determination of the three-dimensional structure of a photosynthetic reaction centre"[155]
1997	John E. Walker	Chemistry	"For their elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP)"[156]
2003	Roderick MacKinnon	Chemistry	"For discoveries concerning channels in cell membranes [...] for structural and mechanistic studies of ion channels"[157]
2003	Peter Agre	Chemistry	"For discoveries concerning channels in cell membranes [...] for the discovery of water channels"[157]
2006	Roger D. Kornberg	Chemistry	"For his studies of the molecular basis of eukaryotic transcription"[158]
2009	Ada E. Yonath	Chemistry	"For studies of the structure and function of the ribosome"[159]
2009	Thomas A. Steitz	Chemistry	"For studies of the structure and function of the ribosome"[159]
2009	Venkatraman Ramakrishnan	Chemistry	"For studies of the structure and function of the ribosome"[159]
2012	Brian Kobilka	Chemistry	"For studies of G-protein-coupled receptors"[160]

Outline

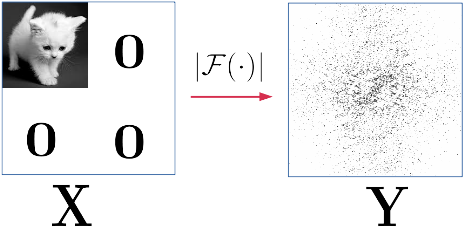
Why phase retrieval?

How people solve PR?

Deep learning for PR?

Possible at all?

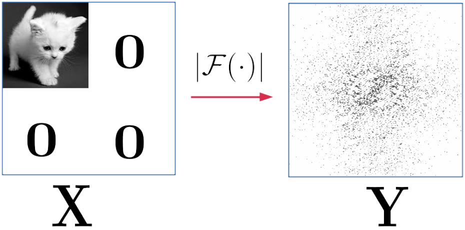
Phase retrieval: given $|\mathcal{F}(x)|$, recover x



$\mathcal{F}$ is oversampled Fourier transform: non-injective for 1D, but **generically** injective for 2D or higher [Hayes, 1982, Bendory et al., 2017]

Possible at all?

Phase retrieval: given $|\mathcal{F}(\boldsymbol{x})|$, recover $\boldsymbol{x}$

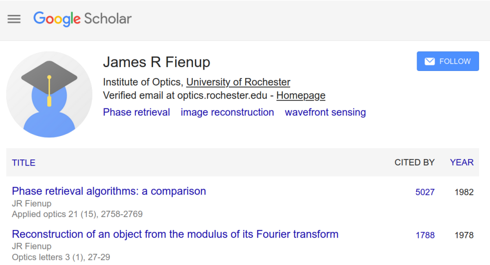


$\mathcal{F}$ is oversampled Fourier transform: non-injective for 1D, but **generically** injective for 2D or higher [Hayes, 1982, Bendory et al., 2017]

- $\mathcal{M}$ constraint: $|\mathcal{F}(\boldsymbol{X})| = \boldsymbol{Y}$
- $\mathcal{S}$ constraint: $\mathcal{A}(\boldsymbol{X}) = \mathbf{0}$

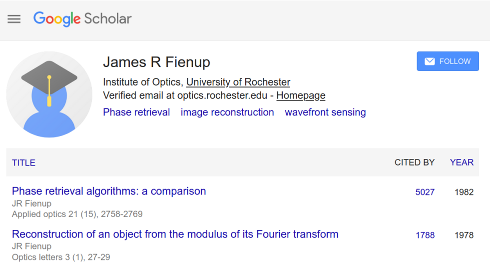
A brief history of algorithms

- Before 70's: error reduction method [Gerchberg and Saxton, 1972]
- Around 80's: hybrid input-output method [Fienup, 1982]



A brief history of algorithms

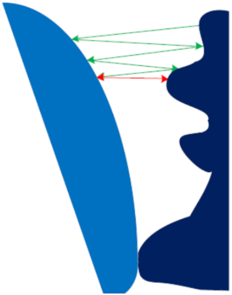
- Before 70's: error reduction method [Gerchberg and Saxton, 1972]
- Around 80's: hybrid input-output method [Fienup, 1982]



- Around 2000: connection to Douglas-Rachford method identified [Bauschke et al., 2002]
- Later variants: RAAR [Luke, 2004], difference map [Elser et al., 2007], see recent review [Luke et al., 2019]

PR algorithms

- Standard: alternating projection methods
- Popular: Fienup's hybrid input-output (HIO) and variants
- No guaranteed recovery (projection onto **nonconvex** sets)
- Often slow in practice, and sensitive to optimization parameters



Hybrid Input-Output (HIO) = Applying Douglas-Rachford splitting to $\delta_{\mathcal{M}} + \delta_{\mathcal{S}}$—ADMM! [Wen et al., 2012]

Insights from randomness?

(Fourier) phase retrieval:

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{F}\boldsymbol{x}|^2$, recover $\boldsymbol{x}$.

Insights from randomness?

(Fourier) phase retrieval:

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{F}\boldsymbol{x}|^2$, recover $\boldsymbol{x}$.

Generalized phase retrieval:

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{A}\boldsymbol{x}|^2$ where $\mathcal{A}$ contains randomness, recover $\boldsymbol{x}$.

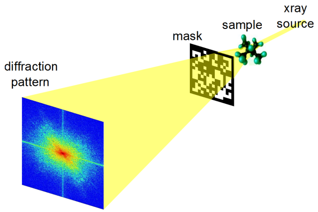
Insights from randomness?

(Fourier) phase retrieval:

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{F}\boldsymbol{x}|^2$, recover $\boldsymbol{x}$.

Generalized phase retrieval:

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{A}\boldsymbol{x}|^2$ where $\mathcal{A}$ contains randomness, recover $\boldsymbol{x}$.



coded-diffraction
CDI [Candès et al., 2015]

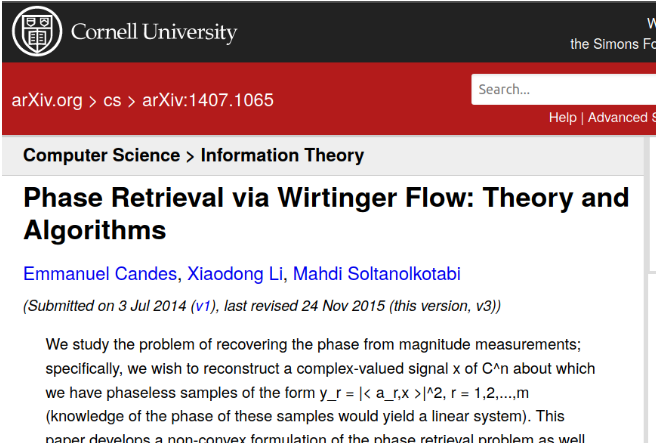
Insights from the Gaussian case?

$\boldsymbol{y} = |\boldsymbol{a}_i^* \boldsymbol{x}|$ for $i = 1, \ldots, m$ where $\boldsymbol{a}_i$'s complex Gaussian vectors

Insights from the Gaussian case?

$\boldsymbol{y} = |\boldsymbol{a}_i^* \boldsymbol{x}|$ for $i = 1, \ldots, m$ where $\boldsymbol{a}_i$'s complex Gaussian vectors

a beautiful **init + local descent** result



Cornell University

the Simons Fo

arXiv.org > cs > arXiv:1407.1065

Search...

Help | Advanced S

Computer Science > Information Theory

Phase Retrieval via Wirtinger Flow: Theory and Algorithms

Emmanuel Candes, Xiaodong Li, Mahdi Soltanolkotabi

(Submitted on 3 Jul 2014 (v1), last revised 24 Nov 2015 (this version, v3))

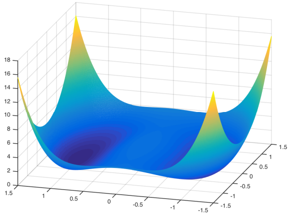
We study the problem of recovering the phase from magnitude measurements; specifically, we wish to reconstruct a complex-valued signal x of C^n about which we have phaseless samples of the form y_r = |< a_r,x >|^2, r = 1,2,...,m (knowledge of the phase of these samples would yield a linear system). This paper develops a non-convex formulation of the phase retrieval problem as well

My own results

Given $y_k = |\boldsymbol{a}_k^* \boldsymbol{x}|$ for $k = 1, \ldots, m$, $\boldsymbol{a}_k$'s iid complex Gaussians, recover $\boldsymbol{x}$ (**up to a global phase**).

My own results

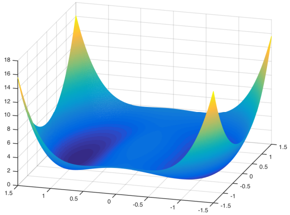
Given $y_k = |\boldsymbol{a}_k^* \boldsymbol{x}|$ for $k = 1, \dots, m$, $\boldsymbol{a}_k$'s iid complex Gaussians, recover $\boldsymbol{x}$ (**up to a global phase**).



$$\boxed{\min_{\boldsymbol{z} \in \mathbb{C}^n} f(\boldsymbol{z}) \doteq \frac{1}{2m} \sum_{k=1}^{m} (y_k^2 - |\boldsymbol{a}_k^* \boldsymbol{z}|^2)^2.}$$

My own results

Given $y_k = |\boldsymbol{a}_k^* \boldsymbol{x}|$ for $k = 1, \dots, m$, $\boldsymbol{a}_k$'s iid complex Gaussians, recover $\boldsymbol{x}$ (**up to a global phase**).



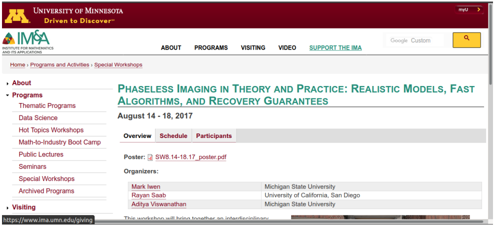
$$\boxed{\min_{\boldsymbol{z} \in \mathbb{C}^n} f(\boldsymbol{z}) \doteq \frac{1}{2m} \sum_{k=1}^{m} (y_k^2 - |\boldsymbol{a}_k^* \boldsymbol{z}|^2)^2.}$$

Theorem ([Sun et al., 2016])

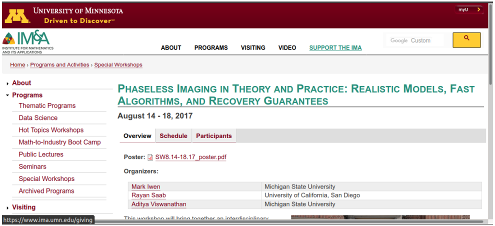
When $\boldsymbol{a}_k$'s generic and m large, with high probability

all local minimizers are global, all saddles are nice.

I was happy until ...



I was happy until ...



Take-home messages

James R Fienup

Fienup: I find it interesting people have tried to analyze Gaussian phase retrieval.

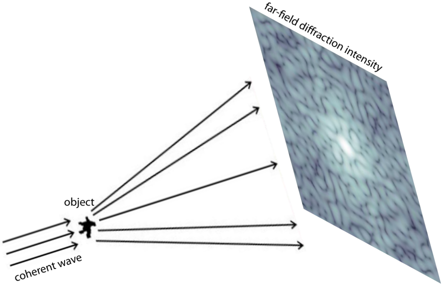
Take-home messages

James R Fienup

Fienup: I find it interesting people have tried to analyze Gaussian phase retrieval.

Beautiful mathematical results gathered so far [Chi et al., 2018, Fannjiang and Strohmer, 2020]

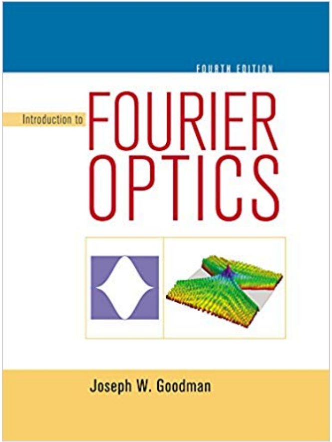
PR is about Fourier measurements



Fraunhofer (far-field) approximation:

$$\left|f\left(x,y\right)\right|^2 \approx \frac{1}{\lambda^2 z^2}\left|\widehat{I}\left(\frac{x}{\lambda z},\frac{y}{\lambda z}\right)\right|^2,$$

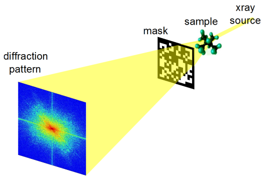
where $I\left(x,y\right) = f\left(x,y,0\right)$ (complex-valued!).



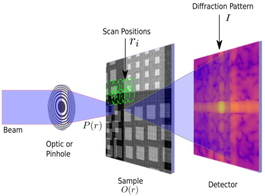
Variants

All variants are about Fourier measurements also

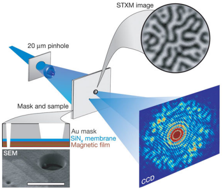
Coded diffraction



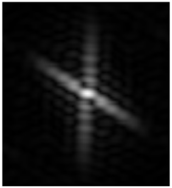
Ptychography



Fourier Holography



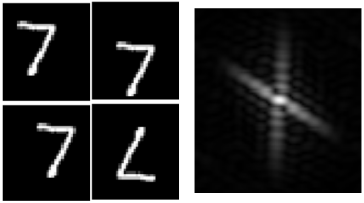
Where's the gap?



Symmetries in Fourier PR:

- translation
- 2D flipping
- global phase

Where's the gap?

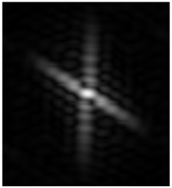


Symmetries in Fourier PR:

- translation
- 2D flipping
- global phase

GPR: For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{A}\boldsymbol{x}|^2$ where $\mathcal{A}$ contains randomness, recover $\boldsymbol{x}$.

Where's the gap?



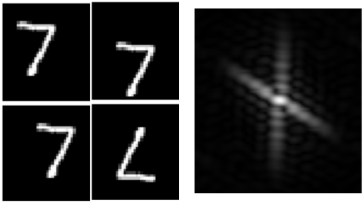
Symmetries in Fourier PR:

- translation
- 2D flipping
- global phase

GPR: For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{A}\boldsymbol{x}|^2$ where $\mathcal{A}$ contains randomness, recover $\boldsymbol{x}$.

GPR doesn't contain the translation and flipping symmetries!

Where's the gap?



Symmetries in Fourier PR:

- translation
- 2D flipping
- global phase

GPR: For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{A}\boldsymbol{x}|^2$ where $\mathcal{A}$ contains randomness, recover $\boldsymbol{x}$.

GPR doesn't contain the translation and flipping symmetries!
Albert Einstein: Everything should be made as simple as possible, but **no simpler**.

Outline

Why phase retrieval?

How people solve PR?

Deep learning for PR?

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

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

- 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?

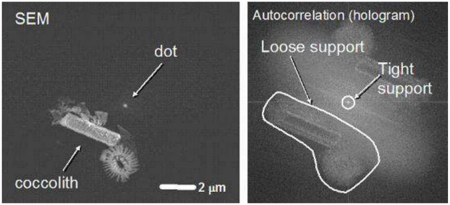
PR is difficult

- 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

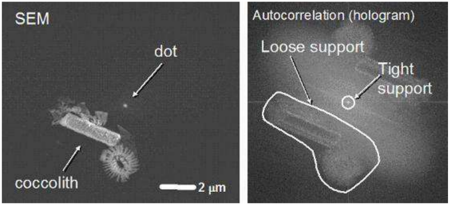
- Most traditional methods fail
  * Effective methods: **proximal methods** [Luke et al., 2019]
  * Exceptions: saddle point optimization [Marchesini, 2007], **2nd order ALM** [Zhuang et al., 2020]
- Complex-valued without accurate support info, e.g., [Marchesini et al., 2005]



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**

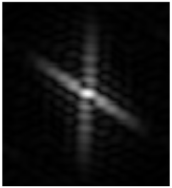
  Attempts: [Metzler et al., 2018, Işıl et al., 2019], but HIO needed for initialization

- 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**

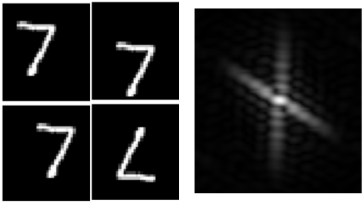
How good are they?



Symmetries in Fourier PR:

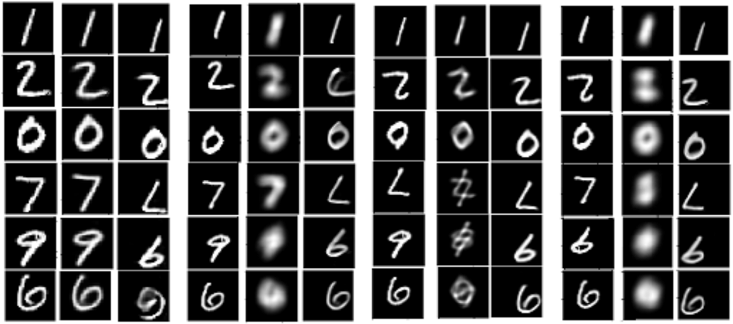
- shift
- 2D flipping
- global phase

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?

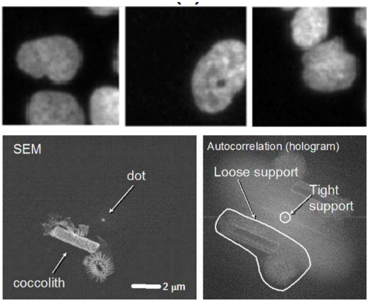
Data! Data! Data!

Why (over)-optimistic results in practice?

Data! Data! Data!

experimental data

naturally oriented and centered



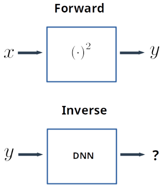
practical data

no natural orientation or centering

Dataset bias breaks problem symmetries

Why learning with symmetries is difficult?

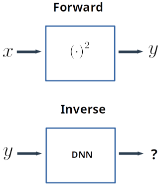
Learning square roots!



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

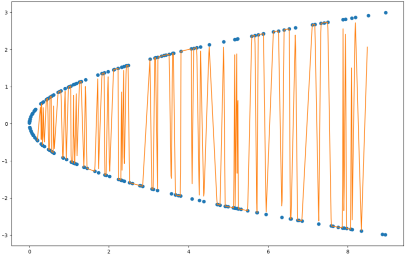
Why learning with symmetries is difficult?

Learning square roots!



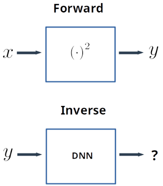
Original Dataset

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



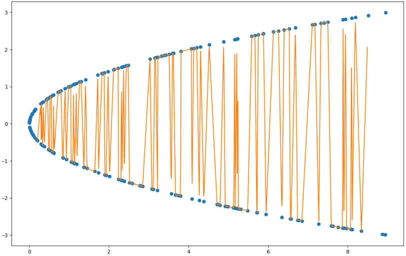
Why learning with symmetries is difficult?

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.

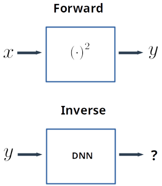
– nontrivial kernel space, e.g. subsampled MRI imaging, e.g., [Gottschling et al., 2020]

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

Get rid of the difficulty?

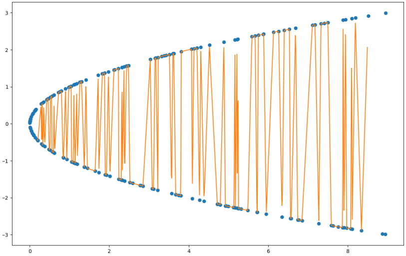
- active symmetry breaking
- passive symmetry breaking

An easy solution to the square root example

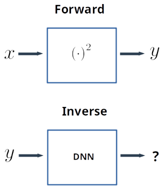


Original Dataset

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

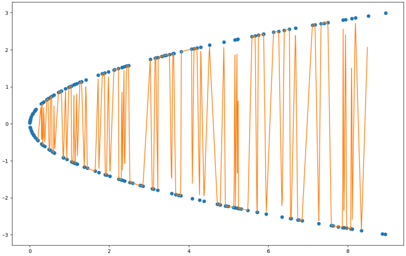


An easy solution to the square root example



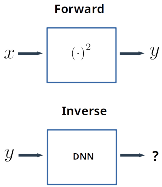
Original Dataset

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



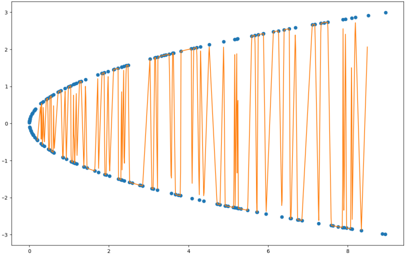
Modified Dataset

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



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

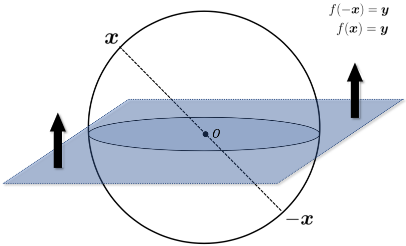
idea: fix the sign symmetry

Active symmetry breaking

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

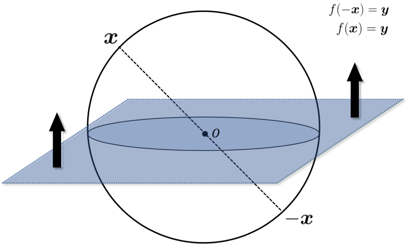
Active symmetry breaking

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]

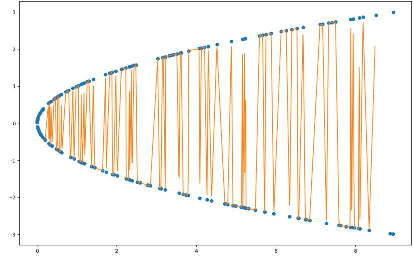
Does it work?

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

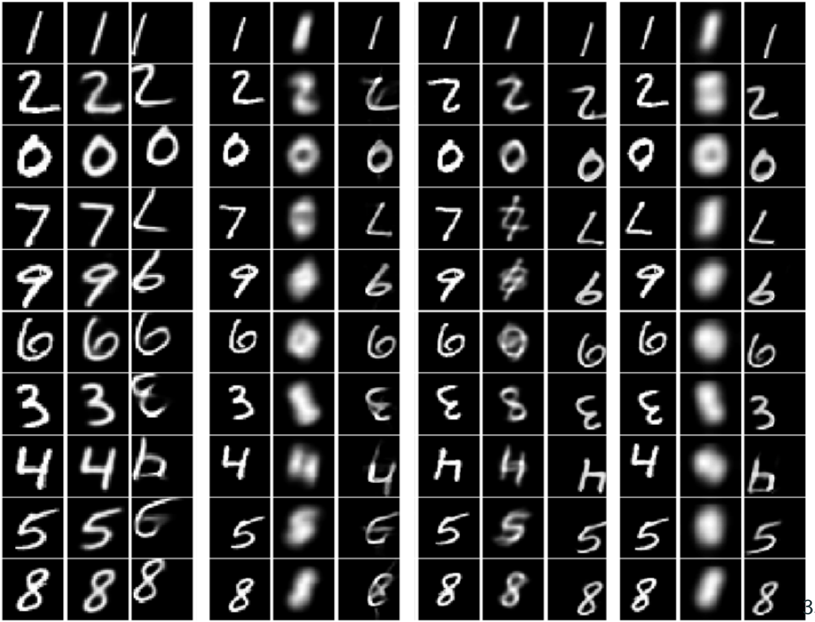
- If $\mathrm{DNN}_{\boldsymbol{W}}\left(\boldsymbol{y}_i\right) \approx \boldsymbol{x}_i$, then $\left|\mathcal{F} \circ \mathrm{DNN}_{\boldsymbol{W}}\left(\boldsymbol{y}_i\right)\right| \approx \left|\mathcal{F} \boldsymbol{x}_i\right| = \boldsymbol{y}_i$
- Consider

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

- 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
- For high accuracy, use the DL result to initialize a local method, e.g., 2nd order ALM [Zhuang et al., 2020]

More developments

- Complex-valued images
- Other datasets
- For high accuracy, use the DL result to initialize a local method, e.g., 2nd order ALM [Zhuang et al., 2020]

Pros: 1) lightweight 2) general
Cons: 1) f is needed 2) dense data needed—Jacobian regularization to the rescue

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

Thoughts

- Essential difficulty: use DL to approximate **one-to-many** mapping
  - When there is forward symmetry (this talk)
  - When the forward mapping under-determined (super-resolution, 3D structure from a single image)
  - or Both

Thoughts

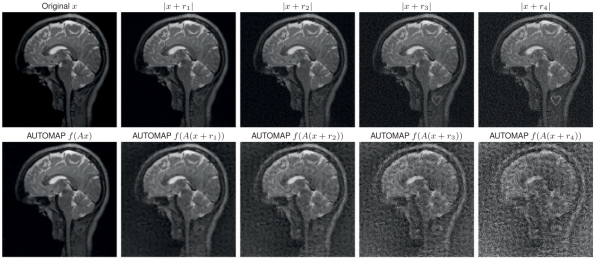
- Essential difficulty: use DL to approximate **one-to-many** mapping

  When there is forward symmetry (this talk)

  When the forward mapping under-determined (super-resolution, 3D structure from a single image)

  or Both

- Not only learning difficulty, but also **robustness** [Antun et al., 2020, Gottschling et al., 2020]



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
- **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

References i

[Antun et al., 2020] Antun, V., Renna, F., Poon, C., Adcock, B., and Hansen, A. C. (2020). **On instabilities of deep learning in image reconstruction and the potential costs of AI.** *Proceedings of the National Academy of Sciences*, page 201907377.

[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.

[Bauschke et al., 2002] Bauschke, H. H., Combettes, P. L., and Luke, D. R. (2002). **Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization.** *Journal of the Optical Society of America A*, 19(7):1334.

[Bendory et al., 2017] Bendory, T., Beinert, R., and Eldar, Y. C. (2017). **Fourier phase retrieval: Uniqueness and algorithms.** In *Compressed Sensing and its Applications*, pages 55–91. Springer International Publishing.

[Candès et al., 2015] Candès, E. J., Li, X., and Soltanolkotabi, M. (2015). **Phase retrieval from coded diffraction patterns.** *Applied and Computational Harmonic Analysis*, 39(2):277–299.

[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.

References ii

[Chi et al., 2018] Chi, Y., Lu, Y. M., and Chen, Y. (2018). **Nonconvex optimization meets low-rank matrix factorization: An overview.** *arXiv:1809.09573*.

[Elser et al., 2007] Elser, V., Rankenburg, I., and Thibault, P. (2007). **Searching with iterated maps.** *Proceedings of the National Academy of Sciences*, 104(2):418–423.

[Fannjiang and Strohmer, 2020] Fannjiang, A. and Strohmer, T. (2020). **The numerics of phase retrieval.** *arXiv:2004.05788*.

[Fienup, 1982] Fienup, J. R. (1982). **Phase retrieval algorithms: a comparison.** *Applied Optics*, 21(15):2758–2769.

[Gerchberg and Saxton, 1972] Gerchberg, R. W. and Saxton, W. (1972). **A practical algorithm for the determination of phase from image and diffraction plane pictures.** *Optik*.

[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).

References iii

[Hayes, 1982] Hayes, M. (1982). **The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform.** *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(2):140–154.

[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, 2004] Luke, D. R. (2004). **Relaxed averaged alternating reflections for diffraction imaging.** *Inverse Problems*, 21(1):37–50.

[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.

[Maddali et al., 2020] Maddali, S., Li, P., Pateras, A., Timbie, D., Delegan, N., Crook, A. L., Lee, H., Calvo-Almazan, I., Sheyfer, D., Cha, W., Heremans, F. J., Awschalom, D. D., Chamard, V., Allain, M., and Hruszkewycz, S. O. (2020). **General approaches for shear-correcting coordinate transformations in bragg coherent diffraction imaging. part i.** *Journal of Applied Crystallography*, 53(2):393–403.

[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*.

[Newton, 2020] Newton, M. C. (2020). **Concurrent phase retrieval for imaging strain in nanocrystals.** *Physical Review B*, 102(1).

[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*.

[Shechtman et al., 2015] Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J., and Segev, M. (2015). **Phase retrieval with application to optical imaging: A contemporary overview.** *Signal Processing Magazine, IEEE*, 32(3):87–109.

[Sun et al., 2016] Sun, J., Qu, Q., and Wright, J. (2016). **A geometric analysis of phase retreival.** *arXiv preprint arXiv:1602.06664*.

[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*.

[Wen et al., 2012] Wen, Z., Yang, C., Liu, X., and Marchesini, S. (2012). **Alternating direction methods for classical and ptychographic phase retrieval.** *Inverse Problems*, 28(11):115010.

[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*.