Phase Retrieval: Fast, Robust, and Data-Driven Algorithms for Computational Imaging

Chris Metzler

Manoj Sharma, Sudarshan Nagesh, Phil Schniter, Oliver Cossairt, Ashok Veeraraghavan, Richard Baraniuk
Computational Imaging

- Advances in computational imaging are driven by advances in
  - Computational algorithms
  - Optical system design
Computational Imaging

• Advances in **computational imaging** are driven by advances in
  – Computational algorithms
  – Optical system design

• **Case in point**
  Computational advance: **Compressive sensing**, an algorithmic breakthrough for solving certain ill-posed linear inverse problems (2004) [Donoho; Candes, Romberg, Tao, 2004]

• **Linear inverse problem**
  Given \( y = Ax + w \)
  solve for \( x \)

• When the problem is **ill-posed**, need to exploit **structure** in \( x \) (aka “prior”)

![Linear inverse problem diagram](image)
Single-Pixel Camera

Takhar et al., 2006
The Phase Retrieval Problem

- Nonlinear phase retrieval (PR) problem
  Given phaseless linear-measurements \( y = \|Ax + w\| \)
  solve for \( x \)

- Computational imaging applications of PR
  - X-ray imaging, Microscopy, Astronomy, Ptychography, Inverse Scattering, ...
Motivating Example: Imaging Through Scattering Media

• Learn the medium’s complex transmission matrix $A$
• Measure the magnitude squared of the output $y^2$ via a standard camera
• Solve PR inverse problem $y^2 = |Ax|^2$ to estimate input $x$

• **Ex: Double PR for imaging** [Drémeau et al. 2015, Rajaei et al. 2016]
  – Learning $A$ requires solving many thousands of PR problems
  – Computational complexity of PR limits the system resolution: $256^2 \times 64^2$ would require 100k CPU hours with today’s PR methods

Coherent Illumination

Amplitude-SLM Pattern

Scatterer

$z = Ax$

Camera

Measurements $y^2 = |Ax|^2$
Motivating Example: Long-Range High-Resolution Imaging

- Extend the **Synthetic Aperture Radar** (SAR) concept to light
  - Using a conventional small aperture camera, capture many coherently illuminated images
  - Stitch the images together to form a high-resolution image from a large synthetic aperture
- “Stitching” process employs PR
- Ex: **Long-range Fourier ptychography** [Holloway et al. 2017]
  - Computational complexity dominated by the large-scale PR problem
Motivating Example: Imaging Around Corners

- Illuminate object with randomized coherent illumination
- Measure intensities of far-field reflections to estimate object auto-correlation function
- Use PR to recover object from its auto-correlation (related through Fourier transform)
- Ex: **Around-the-corner Correlography** [Rangarajan et al. at SMU, in progress]
  - Requires a PR algorithm that is robust to significant amounts of sensor noise
Recall: Advances in computational imaging are driven by advances in:
- Computational algorithms
- Optical system design

In many applications involving phaseless/intensity measurements, the computational imaging system is only as good as its PR algorithm.

Many open challenges in PR algorithm design:
- Computational complexity
- Accuracy and robustness to noise
- Flexibility with respect to the distribution of the entries of $A$
PR Bottlenecks in Computational Imaging

• Recall: Advances in computational imaging are driven by advances in
  – Computational algorithms
  – Optical system design

• In many applications involving phaseless/intensity measurements, the computational imaging system is only as good as its PR algorithm

• Many open challenges in PR algorithm design
  – Computational complexity
  – Accuracy and robustness to noise
  – Flexibility with respect to the distribution of the entries of $A$
Flexibility wrt the Distribution of the Entries of $A$

- Given phaseless linear-measurements: $y = |Ax + w|$, solve for $x$
- Different applications feature different measurement matrices $A$
  - Ex:
    - Transmission matrices: $A$ contains binary 0/1 entries
    - Fourier Ptychography: $A$ applies many overlapping band-pass filters
    - Looking around the corner (and many more): $A$ is the Fourier transform

- Challenge: Most existing PR algorithms are limited to
  - Fourier measurements (Gerchberg-Saxton, Fienup, etc.) or
  - Random Gaussian measurements (PhaseX, etc.)

- Need new algorithms for emerging measurement distributions like binary 0/1
Gaussian Measurements $A$ are “Easy”
Binary 0/1 Measurements A are “Hard”
The Phase Retrieval Algorithm Zoo

- **Projection-based**
  - Gerchberg-Saxton (GS)
  - Hybrid Input Output (HIO)

- **Convex Relaxation**
  - PhaseLift
  - PhaseCut
  - PhaseMax

- **Stochastic Gradient Descent**
  - Wirtinger Flow (WF)
  - Reweighted Wirtinger Flow (RWF)
  - Truncated Wirtinger Flow (TWF)

- **AMP-based**
  - prGAMP
  - prVBEM
  - prSAMP
This Talk: Two New PR Algorithms

• Leverage recent algorithmic progress in *optimization* and *machine learning*

• **prVAMP** [M. et al., ICCP, 2017]
  – Extremely computationally efficient and parallelizable
  – Able to handle binary 0/1 and Gaussian measurements
  – Can improve accuracy and robustness further by imposing learned priors (not focus of this talk)

• **prDeep** [M. et al., SPIE, 2018]
  – Computationally efficient
  – Able to handle binary 0/1, Gaussian, and Fourier measurements (and many more)
  – Uses a deep neural network to exploits training data and boost performance
AMP = Approximate Message Passing algorithm for solving linear inverse problems

- Appropriated from statistical physics

- AMP *approximates* sum-product or max-sum belief propagation on the appropriate factor graph

- Accurately performs MMSE and MAP inference for linear inverse problems with (sub)Gaussian measurement matrices $A$

- **Extremely fast**: Solves LASSO problem in $\sim 30$ matrix-vector multiplies

**Linear inverse problem**

Given $y = Ax + w$

solve for $x$

given a model for $x$

* Access to phase *
Intuition Behind AMP

- Iterative thresholding algorithm
  - Model: $x$ has a **sparse representation** in some basis

Given an initial guess $x^0$

For $n = 0, 1, \ldots$ do

$$x^{n+1} = H_K \left[ x^n + A^T (y - Ax^n) \right]$$

1. Error in matching measurements
2. Improve signal estimate wrt matching measurements
3. Project estimated signal back onto (sparse) model by simple thresholding

- Linear inverse problem

Given $y = Ax + w$

solve for $x$
given a model for $x$
Intuition Behind AMP

• **Iterative thresholding algorithm**
  – Model: \( x \) has a **sparse representation** in some basis

Given an initial guess \( x^0 \)

For \( n = 0, 1, \ldots \) do

\[
x^{n+1} = H_K \left[ x^n + A^T (y - Ax^n) \right]
\]

1. Error in matching measurements
2. Improve signal estimate wrt matching measurements
3. Project estimated signal back onto (sparse) model by simple thresholding

**Problem:** Thresholding \( H_K \) is the optimal projection onto the sparse signal model only if the term \( A^T (y - Ax^n) \) can be modeled as WGN

\[
y = Ax + w
\]

solve for \( x \) given a model for \( x \)
AMP [Donoho et al. 2009]

• **AMP algorithm**
  – Model: \( x \) has a **sparse representation** in some basis

Given an initial guess \( x^0 \)

For \( n = 0, 1, \ldots \) do

\[
x^{n+1} = H_K \left[ x^n + A^T (y - Ax^n + O(x^{n-1})) \right]
\]

1. Error in matching measurements
2. Improve signal estimate wrt matching measurements
3. Project estimated signal back onto (sparse) model by simple thresholding

1A. **Onsager Correction Term** that makes the update term behave like white Gaussian noise

State-of-the-art performance for sparse recovery

- **Linear inverse problem**
  Given \( y = Ax + w \)
  solve for \( x \)
given a model for \( x \)
prGAMP: Accurate and Fast But Not Robust [Schniter and Rangan 2015]

• First extension of AMP to PR

• Approximates solution to

\[
\hat{x} = \mathbb{E}[x|y] \quad s.t. \quad y = |z + w| \\
\quad z = Ax \\
\quad w \sim CN(0, \sigma_w^2) \\
\quad x \sim px
\]

• Entries of A must be zero mean i.i.d. (sub)Gaussian
  – Cannot handle binary 0/1 nor Fourier measurements

• Complexity: \(O(N^2)\), where \(N\) is dimensionality of \(x\) (ex: number of pixels)
prSAMP: Accurate and Robust But Slow [Rajaei et al. 2016]

- Extends “Swept AMP” to PR [Manoel et al. 2015]
- Approximates solution to

\[
\hat{x} = \mathbb{E}[x|y] \quad s.t. \quad y = |z + w|
\]

\[
z = Ax
\]

\[
w \sim CN(0, \sigma_w^2)
\]

\[
x \sim p_x
\]

- Updates estimates elements of \(x\) sequentially

- Entries of \(A\) can be nonzero mean and even ill-conditioned
  - Works with binary 0/1 measurements
  - Exact requirements unknown [Manoel et al. 2015]

- Complexity: \(O(N^3)\)
  - Can be accelerated when \(A\) has a special “block” structure (but unrealistic for many applications)
Binary 0/1 Measurements A are “Hard”
prSAMP for Binary 0/1 Measurements [Rajaei et al. 2016]

... but unfortunately *computationally expensive*; a single $N=64^2$ PR problem takes 1 hour

Note: Results are with multiple restarts (non-convex problem)
New PR Algorithm: prVAMP

- Extends **VAMP-GLM** [Schniter et al. 2016] to PR

- Approximates solution to

\[
\hat{x} = \mathbb{E}[x|y] \quad s.t. \quad y = |z_2 + w|
\]

\[
z_1 = Ax_2
\]

\[
z_2 = z_1, \quad x_2 = x_1,
\]

\[
w \sim \mathcal{CN}(0, \sigma_w^2)
\]

\[
x_1 \sim p_x
\]

- Infuses **ADMM** concepts into AMP

- Updates the entire estimate of \(x\) **in parallel**

- Extends class of measurement matrices \(A\) to anything “right-rotationally invariant”
  - Works with binary 0/1 and Gaussian measurements
  - Does not work with Fourier measurements

- **Complexity:** \(O(N^2)\) + one-time SVD (and vector operations can be parallelized on GPUs)

- **Accurate, robust, and fast**
prVAMP Phase Transition (Binary 0/1 Measurements)

Note: Results are with multiple restarts (non-convex problem)
prVAMP vs. prSAMP

- Both can handle measurement matrices $A$ whose entries are not zero-mean Gaussian
- Both are robust to measurement noise

- **prSAMP** updates the elements of $x$ one at a time
  - Computationally demanding and not parallelizable
  - $O(N^3)$ complexity

- **prVAMP** updates two copies of the entire signal $x$, which are iteratively brought closer and closer together (ADMM concept)
  - Computationally efficient and parallelizable
  - $O(N^2)$ complexity
prVAMP vs. prSAMP: Error and Computation Time

- prVAMP is ~100x faster than prSAMP on the same CPU (0/1 measurements and low SNR)
- prVAMP easily maps onto a GPU for another 500x speed up
  - Unfair comparison below: prVAMP is running on one $1k GPU while prSAMP is running on $1M worth of CPUs
Double PR Imaging Experiments with prVAMP

256x256 measurements

Calibration requires solving 256² PR problems of size 64² (3 hours on one TitanX GPU)

Recovering 64x64 image on amplitude SLM (few seconds)
prVAMP: Pro and Con

**PRO**

- Robust
- Extremely computationally efficient and parallelizable
- Can handle both binary 0/1 and Gaussian measurements

**CON**

- Cannot handle Fourier measurements $A$
  - Say goodbye to X-ray diffraction, Correlography, Ptychography, ...

**Challenges:**

- How to deal with Fourier measurements?
- How to integrate signal structure into PR to make more robust?
prVAMP: Pro and Con

**PRO**
- Robust
- Extremely computationally efficient and parallelizable
- Can handle both binary 0/1 and Gaussian measurements

**CON**
- Cannot handle Fourier measurements $A$
  - Say goodbye to X-ray diffraction, Correlography, Ptychography, ...

**Challenges:**
- How to deal with Fourier measurements?
- How to integrate signal structure into PR to make more robust?

Deep Learning!
Standard Deep Learning Procedure

• Start with a particular inverse problem
• Propose an architecture for that problem
• Train the network to solve the inverse problem
• Apply to the specified inverse problem

Pros:
• Fast, one-step solution

Cons:
• Need to design and train for each inverse problem
• Black box

[Kappeler et al. 2016]
Proposed: Deep Learning as Part of an Algorithm

- Train a neural network to denoise noisy images
- Set up an algorithm that imposes priors with denoisers
- Test on arbitrary inverse problems

Pros:
- Encode and use knowledge of $A$
  - E.g., take advantage of FFT
- Train once, apply to many problems
- Optimization problem: Interpretable, (potentially) convergence guarantees, etc.

Cons:
- Need to iterate an algorithm

Example:

\[
\begin{align*}
x^{t+1} &= D \left( x^t + A^H z^t \right) \\
z^{t+1} &= y - A x^{t+1}
\end{align*}
\]
Step 1: Train a Neural Network to Denoise Noisy Images

- We use the DnCNN denoiser [Zhang et al. 2017]
- Train offline with 400 images divided into 300,000 50x50 image patches (3 hours)
Step 2: Set up an Alg that Imposes Priors with Denoisers

- L2 amplitude loss function
- RED: Regularization by Denoising [Romano et al. 2017]

\[ \| y - |Ax| \|^2 + \lambda x^t(x - D(x)) \]

Analytic subgradient

Is convex.\(^1\) Has proximal mapping.

- Need to initialize non-convex problem. We initialize with HIO algorithm [Fienup 82]
- Can solve with many different methods. We use FASTA [Goldstein et al. 2014]

---

1: Under some conditions on the denoiser
Simulation: Correlography (Fourier Measurements)

Hybrid Input/Output (HIO) (52 sec)

prDeep (52+27 sec)
Summary

- In many applications involving phaseless/intensity measurements, the computational imaging system is only as good as its PR algorithm.

- **prVAMP**
  - Robust to measurement noise
  - Very computationally efficient (1000s of times faster than current state of the art)
  - Able to handle 0/1 and Gaussian measurements
  - Easy to impose additional prior knowledge on signal structure
  - Code and dataset available: http://dsp.rice.edu/research/transmissionmatrices/

- **prDeep**
  - Robust to measurement noise
  - Moderately computationally efficient
  - Computationally demanding, one-time training
  - Able to handle Fourier measurements (and many more)
  - Easy to learn additional prior knowledge on signal structure
  - Code available soon
Acknowledgements

Team
- Manoj Sharma
- Sudarshan Nagesh
- Phil Schniter
- Oliver Cossairt
- Rich Baraniuk
- Ashok Veeraraghavan

Sponsors
- DARPA
- USN
- ONR
- NRC
- NSF
- Rice University
Summary

• In many applications involving phaseless/intensity measurements, the computational imaging system is only as good as its PR algorithm.

• prVAMP
  – Robust to measurement noise
  – Very computationally efficient (1000s of times faster than current state of the art)
  – Able to handle 0/1 and Gaussian measurements
  – Easy to impose additional prior knowledge on signal structure
  – Code and dataset available: http://dsp.rice.edu/research/transmissionmatrices/

• prDeep
  – Robust to measurement noise
  – Moderately computationally efficient
  – Computationally demanding, one-time training
  – Able to handle Fourier measurements (and many more)
  – Easy to learn additional prior knowledge on signal structure
  – Code available soon