Primal-Dual Optimization of Phase Retrieval Via Optimization Transfer

Daniel S. Weller

University of Virginia

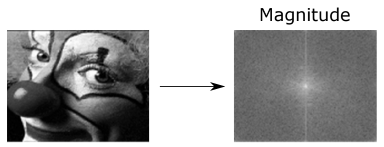
May 23, 2016

UNIVERSITY of VIRGINIA ENGINEERING

Virginia Imaging Theory and Algorithms Laboratory

Introduction to Phase Retrieval

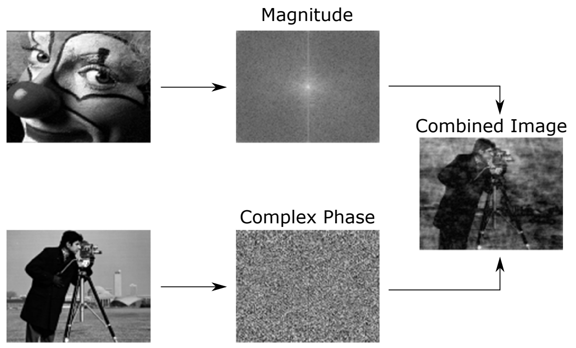
- In some signal/image acquisition applications, only the magnitudes of a complex-valued representation (e.g., a Fourier transform) of that image are available.



- Phase retrieval describes the problem of signal recovery, absent the complex phase information.

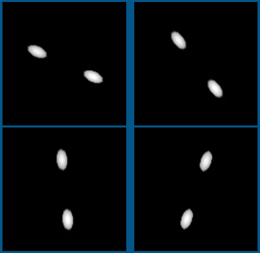
Introduction to Phase Retrieval

- Complex phase contains essential information about image features.
- To illustrate, replace the complex phase of the 2D DFT of the clown with that of the cameraman.



Introduction to Phase Retrieval

- Phase retrieval is ill-posed without additional information.
- Current methods use sparse or other regularizers.
- This presentation focuses on a primal-dual optimization method for sparse or compressive phase retrieval.



Point spread functions

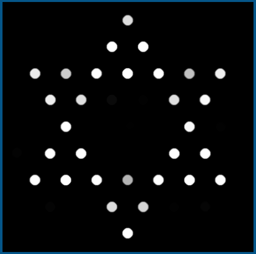


Image reconstruction

Problem Formulation

- A description of phase retrieval begins with the forward model, corresponding to the measurement process.
- The desired signal or columnized image $\boldsymbol{x}$ is related to squared-magnitude intensity measurements $\boldsymbol{y} = [y_1, \ldots, y_M]$ through a complex-valued transform $\boldsymbol{A}$:

$$y_m = |[\boldsymbol{A}\boldsymbol{x}]_m|^2, \quad m = 1, 2, \ldots, M.$$

- In the experiments that follow, $\boldsymbol{A}$ is primarily the discrete Fourier transform.
- In this work, the number of measurements M is less than or equal to the dimension of $\boldsymbol{x}$, denoted N.

Problem Formulation

- In real settings, these measurements are corrupted by noise.
- Depending on the type of imaging, noise may be appropriately added before (pre) or after (post) taking the magnitude in the forward model:

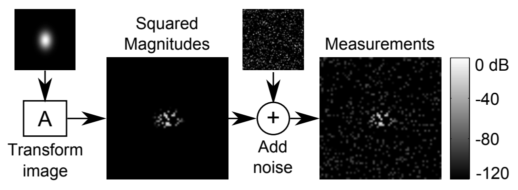
$$y_m = |[\boldsymbol{A}\boldsymbol{x}]_m + \text{pre}|^2 + \text{post}.$$

Problem Formulation

- In real settings, these measurements are corrupted by noise.
- Depending on the type of imaging, noise may be appropriately added before (pre) or after (post) taking the magnitude in the forward model:

$$y_m = |[\boldsymbol{A}\boldsymbol{x}]_m|^2 + \text{post}.$$

This work focuses on the post-magnitude noise model.



Problem Formulation

- Gaussian noise is frequently assumed, motivating the quadratic data fit term employed in many algorithms:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_{m=1}^{M} |y_m - |[\boldsymbol{A}\boldsymbol{x}]_m|^2|^2.$$

- Robust regression applies a 1-norm to the data fit term to avoid over-fitting to low-SNR data:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_{m=1}^{M} |y_m - |[\boldsymbol{A}\boldsymbol{x}]_m|^2|.$$

  This model appears in [DS Weller et al., IEEE ICIP, 2014] [P Hand, arXiv, 2015] [DS Weller et al., IEEE TCI, 2015].

- This work mainly considers the 1-norm data fit penalty.

Problem Formulation

- The data fit term corresponds to our forward model for the magnitude-only measurements.
- This problem remains ill-posed when M is not large enough, motivating regularization with function $R(\boldsymbol{x})$:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_{m=1}^{M} |y_m - |[\boldsymbol{A}\boldsymbol{x}]_m|^2|^p + \beta R(\boldsymbol{x}).$$

  ($p = 1$ or 2, for the 1-norm and quadratic data fit penalties, respectively)
- This work focuses on dictionary-based sparsity (synthesis frame is built into $\boldsymbol{A}$; $R(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$), as described in [DS Weller et al., IEEE TCI, 2015].

Relationship to Existing Work

A partial list of sparsity-regularized phase retrieval methods:

- Matrix-lifting and semidefinite relaxation methods (e.g., PhaseLift): [ML Moravec et al., SPIE Wavelets XII, 2007] [Y Shechtman et al., Opt Express, 2011] [H Ohlsson et al., arXiv, 2012] [EJ Candès et al., SIAM J Imag Sci, 2013] [X Li and V Voroninski, SIAM J Math Anal, 2013] [L Demanet and V Jugnon, arXiv, 2013] [I Waldspurger et al., Math Programm, 2015] [P Hand, arXiv, 2015]
- Alternating projections: [S Mukherjee and CS Seelamantula, IEEE ICASSP, 2012] [S Mukherjee and CS Seelamantula, IEEE TSP, 2014]

Relationship to Existing Work

More sparsity-regularized phase retrieval methods:

- Graphical model-based approximate message passing: [P Schniter and S Rangan, Allerton, 2012]
- Pursuit-type greedy methods: [Y Shechtman et al., IEEE TSP, 2014]
- Majorize-minimize methods: [DS Weller et al., IEEE ICIP, 2014] [DS Weller et al., IEEE TCI, 2015]

Relationship to Existing Work

More sparsity-regularized phase retrieval methods:

- Graphical model-based approximate message passing: [P Schniter and S Rangan, Allerton, 2012]
- Pursuit-type greedy methods: [Y Shechtman et al., IEEE TSP, 2014]
- Majorize-minimize methods: [DS Weller et al., IEEE ICIP, 2014] [DS Weller et al., IEEE TCI, 2015]

The last type forms the basis for this work.

Variable Splitting

- The coupling of x in each component of the data fit term (via A) complicates solving the inverse problem.
- Introducing an auxiliary variable $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$ decouples the components, making the data fit term separable:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{u}} \sum_{m=1}^{M} |y_m - |u_m|^2|^p + \beta \|\boldsymbol{x}\|_1, \quad \boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}.$$

- The Lagrange form of this problem is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\nu}) = \sum_{m=1}^{M} |y_m - |u_m|^2|^p + \beta \|\boldsymbol{x}\|_1 + \left[\boldsymbol{\nu}'(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{u})\right]_R.$$

The dual vector $\boldsymbol{\nu}$ is complex-valued.

Primal-Dual Optimization

- Solving the Lagrange dual problem directly,

$$\hat{\boldsymbol{\nu}} = \arg\max_{\boldsymbol{\nu}} \left[g(\boldsymbol{\nu}) = \inf_{\boldsymbol{x},\boldsymbol{u}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\nu}) \right]$$

- For a fixed $\boldsymbol{\nu}$, the Lagrangian function is additively separable. Minimizing each x_n individually,

$$\inf_{x_n} \beta|x_n| + \left[[\boldsymbol{A}'\boldsymbol{\nu}]_n^* x_n \right]_R = \begin{cases} 0, & |[\boldsymbol{A}'\boldsymbol{\nu}]_n| \leq \beta, \\ -\infty, & |[\boldsymbol{A}'\boldsymbol{\nu}]_n| > \beta. \end{cases}$$

- Therefore, $g(\boldsymbol{\nu})$ is finite only when $\|\boldsymbol{A}'\boldsymbol{\nu}\|_\infty \leq \beta$.

Primal-Dual Optimization

- For each u_m, the exact value is more complicated, but we can guarantee the infimum is always less than or equal to zero (at least when $y_m \geq 0$):

$$\inf_{u_m} |y_m - |u_m|^2|^p + \nu_m^* u_m = \inf_{|u_m|} |y_m - |u_m|^2|^p - |\nu_m||u_m| \\ \leq -|\nu_m|\sqrt{y_m}.$$

- Since $g(\boldsymbol{\nu}) \leq 0$, and $g(\mathbf{0}) = 0$, $\max_{\boldsymbol{\nu}} g(\boldsymbol{\nu}) = 0$.
- Returning to the primal problem:
  - The minimum feasible primal objective value is zero only when $\boldsymbol{x} = \mathbf{0}$, and $y_m = 0$ for all m.
  - In all other cases, the minimum is greater than zero, and we have a nonzero duality gap (weak duality).

Optimization Transfer

- Weak duality derives from the nonconvexity of the original problem. Instead, we use optimization transfer to produce a series of convex problems to solve.
- Optimization transfer refers to the iterative process of solving an optimization problem by solving a series of simpler problems built around a surrogate function.
- When minimizing an objective function, the function $f(\boldsymbol{x})$ is replaced by a surrogate called a majorizer $\phi(\boldsymbol{x}; \boldsymbol{x}^i)$, parameterized by the choice of majorization point $\boldsymbol{x}^i$, that satisfies two properties:
    1. The majorizer dominates the original objective: $\phi(\boldsymbol{x}; \boldsymbol{x}^i) \geq f(\boldsymbol{x})$, for all $\boldsymbol{x}$.
    2. The majorizer equals the objective at the majorization point: $\phi(\boldsymbol{x}^i; \boldsymbol{x}^i) = f(\boldsymbol{x}^i)$.

Optimization Transfer

- Given these properties, minimizing the majorizer is guaranteed not to increase the original objective function value relative to $\boldsymbol{x}^i$:

$$\boldsymbol{x}^{i+1} = \arg\min_{\boldsymbol{x}} \phi(\boldsymbol{x}; \boldsymbol{x}^i) \quad \Rightarrow$$
$$f(\boldsymbol{x}^{i+1}) \leq \phi(\boldsymbol{x}^{i+1}; \boldsymbol{x}^i) \leq \phi(\boldsymbol{x}^i; \boldsymbol{x}^i) = f(\boldsymbol{x}^i).$$

- Using the solution to each iteration as the majorization point for the next results in a monotonically nonincreasing sequence of objective function values $f(\boldsymbol{x}^1), f(\boldsymbol{x}^2), \ldots$.
- Furthermore, choosing a surrogate that is differentiable with respect to $\boldsymbol{x}$ around $\boldsymbol{x} = \boldsymbol{x}^i$ whenever $f(\boldsymbol{x})$ is differentiable at $\boldsymbol{x}^i$ ensures that the algorithm converges to a local extremum of the original function (if it is differentiable almost everywhere).

Optimization Transfer

Previous work [DS Weller et al., IEEE ICIP, 2014] [DS Weller et al., IEEE TCI, 2015] constructs a convex majorizer for the 1-norm and quadratic data fit terms:

- The data fit term can be rewritten as a pairwise maximum:

$$|y_m - |[\boldsymbol{Ax}]_m|^2|^p = (\max\{y_m - |[\boldsymbol{Ax}]_m|^2,\ |[\boldsymbol{Ax}]_m|^2 - y_m\})^p.$$

  The first term is concave; the second is convex.

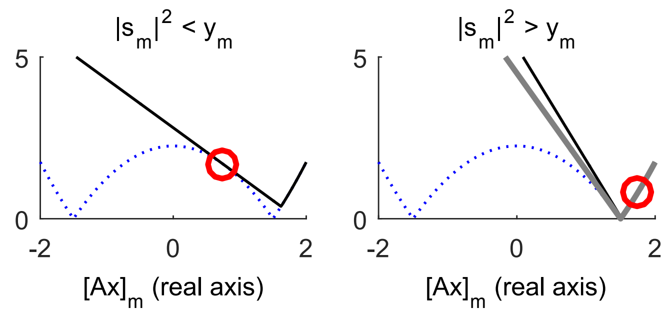
- Replace the first with its tangent plane around $s_m = [\boldsymbol{Ax}^i]_m$:

$$|y_m - |[\boldsymbol{Ax}]_m|^2|^p \leq (\max\{y_m + |s_m|^2 - 2[s_m^*[\boldsymbol{Ax}]_m]_R,\ |[\boldsymbol{Ax}]_m|^2 - y_m\})^p.$$

- The resulting function is a convex majorizer that tightly fits the original data fit term.

Optimization Transfer

When $|s_m|^2 < y_m$, the majorizer (black) is shown on the left. When $|s_m|^2$ is larger, the majorizer can be applied as if $|s_m|^2 = y_m$, yielding the gray curve on the right. This curve is a tighter fit than the function using the original s_m, in black.



Similar majorizers exist for the $p = 2$ case.

Minimizing the Majorizer

- Applying variable splitting to the majorizer yields a constrained optimization similar to that for the original problem. However, this problem is convex.
- Direct solution of the Lagrangian, or an efficient alternating minimization like ADMM can be applied here.
- The algorithm [DS Weller et al., IEEE TCI, 2015]:

$$\boldsymbol{x}^{(i+1)} = \arg\min_{\boldsymbol{x}} \beta\|\boldsymbol{x}\|_1 + \tfrac{\mu}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{u}^{(i)} + \boldsymbol{d}^{(i)}\|_2^2.$$

$$u_m^{(i+1)} = \arg\min_{u_m} \phi_m(u_m; s_m) + \tfrac{\mu}{2}|u_m - [\boldsymbol{A}\boldsymbol{x}^{(i+1)} + \boldsymbol{d}^{(i)}]_m|^2.$$

$$\boldsymbol{d}^{(i+1)} = \boldsymbol{d}^{(i)} + \boldsymbol{A}\boldsymbol{x}^{(i+1)} - \boldsymbol{u}^{(i+1)}.$$

The majorizer is
$\phi_m(u_m; s_m) = (\max\{y_m + |s_m|^2 - 2[s_m^* u_m]_R,\ |u_m|^2 - y_m\})$.

Incorporating Image Constraints

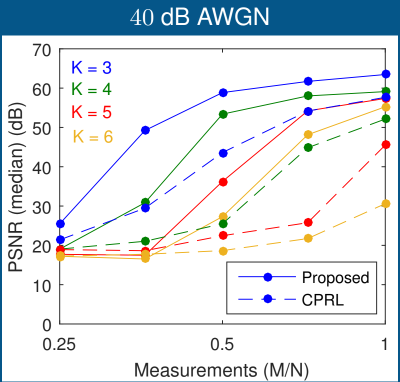
- Nonnegativity and support constraints can be enforced via projection between updating x and u.
- Alternatively, the x-update can be modified to explicitly include the constraints.
  - For a support mask $\mathcal{S}$, the matrix A can be modified to only include those columns in $\mathcal{S}$.
  - For nonnegativity, $\|x\|_1$ becomes $\mathbf{1}'x$, yielding a quadratic program.
- Incorporating other image domain constraints, like bounded magnitudes, is also possible.

Experimental Methods

- Both 1D Monte Carlo experiments on sparse random signals, and 2D image reconstructions for the Star of David phantom and a point spread function are conducted.
- Simulated 1D and 2D (image) data are generated with additive Gaussian noise and/or outliers.
- The reconstructed signals are analyzed quantitatively (for 1D Monte Carlo tests) and visually (for images). Note that errors are computed after accounting for global phase, spatial shifts, and image reversal (the objective function is insensitive to these operations).
- Some experiments are based on [DS Weller et al., IEEE TCI, 2015].

Monte Carlo Experiments (1D)

Compare matrix lifting-based compressive phase retrieval (CPRL) [H Ohlsson et al., arXiv, 2012] versus proposed optimization-transfer with 1-norm data fit term:



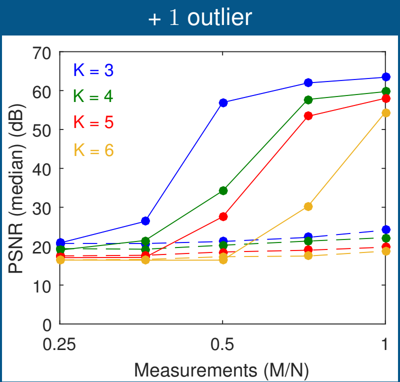
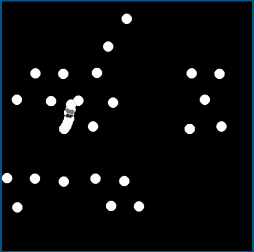


Image Reconstruction

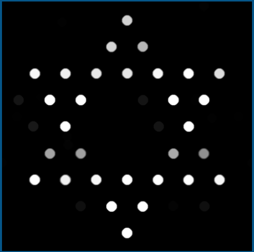
The Star-of-David phantom is reconstructed using both GESPAR and the proposed non-smooth-regularized phase retrieval method with sparsity only, and with an additional non-negativity constraint.

GESPAR



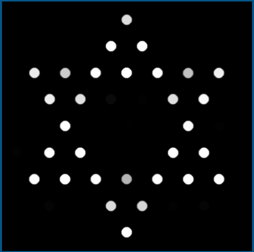
PSNR = 10.4 dB

Proposed
Sparsity Only



PSNR = 24.2 dB

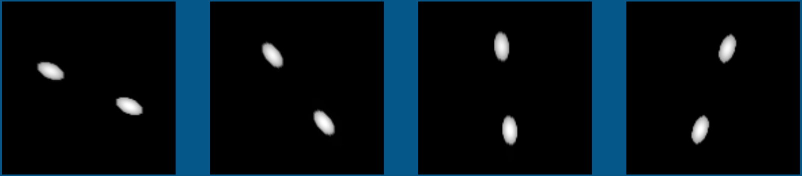
Proposed
and Non-Negativity



PSNR = 26.7 dB

Image Reconstruction

Elliptical cross sections of a simulated double-helix point spread functions (1% outliers) are reconstructed using the proposed phase retrieval method with sparsity only.



PSNR = 34.9 dB PSNR = 32.9 dB PSNR = 37.6 dB PSNR = 36.4 dB

Discussion

- The robust data fit term and non-smooth regularization greatly improves phase retrieval of both 1D signals and 2D images.
- The synthesis sparse prior extends to images of point spread functions. Other research concerning dictionary learning for phase retrieval like [AM Tillmann et al., arXiv, 2016] would be helpful for more arbitrary PSFs.
- The primal-dual ADMM-based solution to the majorizer in the optimization transfer algorithm is effective here without any manual tuning. Similarly, the regularization parameter was not manually adjusted for any of these experiments.

Conclusion

- A phase retrieval framework using variable splitting was introduced.
- A primal-dual ADMM-based optimization transfer algorithm was presented for robust phase retrieval.
- Both 1D and 2D image reconstruction results were presented and discussed.
- Ongoing research concerning synthesizing analysis transforms, image-domain constraints, more direct primal-dual methods, and theoretical analysis of such algorithms, continues.