PHASE RETRIEVAL OF SPARSE SIGNALS USING OPTIMIZATION TRANSFER AND ADMM

Daniel S. Weller[*], *Ayelet Pnueli*[†], *Ori Radzyner*[†], *Gilad Divon*[†], *Yonina C. Eldar*[†], *Jeffrey A. Fessler*[*]

[*] University of Michigan, EECS Department, Ann Arbor, MI, 48109, USA
[†] Technion, Israel Institute of Technology, Department of EE, Haifa 32000, Israel

ABSTRACT

We propose a reconstruction method for the phase retrieval problem prevalent in optics, crystallography, and other imaging applications. Our approach uses signal sparsity to provide robust reconstruction, even in the presence of outliers. Our method is multi-layered, involving multiple random initial conditions, convex majorization, variable splitting, and alternating directions method of multipliers (ADMM)-based implementation. Monte Carlo simulations demonstrate that our algorithm can correctly and robustly detect sparse signals from full and undersampled sets of squared-magnitude-only measurements, corrupted by additive noise or outliers.

Index Terms— phase retrieval, sparse recovery, variable splitting, majorize-minimize, ADMM

1. INTRODUCTION

In "phase retrieval," [1, 2] we aim to reconstruct a length-N signal from $M \leq N$ magnitude-only samples of a linear transform of that signal. The problem of recovering a signal from the intensity of its Fourier spectrum is prevalent in crystallography [3–5], optics [6], and coherent diffraction imaging [7, 8]. Beyond lacking phase information, the measured magnitudes are often noisy in practice. However, noise and outliers pose substantial challenges for many existing methods.

Lacking phase information for the measurements, we require some assumptions, or prior information, to reconstruct the signal. A reasonable prior exploited broadly is signal sparsity: we assume that the signal can be represented by a small set of nonzero coefficients. While prominent earlier algorithms [9, 10] rely on oversampling, recent work use sparsity as prior information [11–14] or pose phase retrieval as a semidefinite programming problem [15–20]. Since phase retrieval is a nonconvex problem, global convergence usually is not guaranteed; therefore, in practice, multiple random initializations often are used to improve the solution.

Here, we investigate a phase retrieval approach based on optimization transfer [21–23]. Our algorithm uses a new convex majorizer for various nonconvex cost functions used in phase retrieval. Exploiting this flexibility, we choose a 1-norm penalty on the squared-magnitude measurements, providing robustness to outliers. We then develop an iterative algorithm for minimizing this majorizer using variable splitting and ADMM [24–27].

Section 2 describes the general problem. In Section 3, we derive a convex majorizer for the general problem and an ADMM algorithm for minimizing it in the sequel. Section 4 reports simulation results demonstrating the accuracy of our algorithm. We conclude with a discussion of these results and future work.

2. PROBLEM DESCRIPTION

Consider recovering a signal $\boldsymbol{x} \in \mathbb{C}^N$ from a vector of magnitude measurements $\boldsymbol{y} \in \mathbb{R}^M$, modeled as

$$y_i = |[\boldsymbol{A}\boldsymbol{x}]_i|^q + n_i, \tag{1}$$

where $\boldsymbol{A}$ denotes an arbitrary linear transform (e.g., the discrete Fourier transform (DFT)), $[\boldsymbol{A}\boldsymbol{x}]_i \triangleq \sum_{j=1}^{N} A_{ij}x_j$, q is a positive integer (usually 1 or 2), and n_i is additive noise.

To recover $\boldsymbol{x}$ from $\boldsymbol{y}$, we assume $\boldsymbol{x}$ is sparse: i.e., $\boldsymbol{x}$ has only $K \ll N$ nonzero values. Using the 1-norm $\|\cdot\|_1$ to promote sparsity, we estimate $\boldsymbol{x}$ by minimizing the regularized cost function

$$\Psi(\boldsymbol{x}) \triangleq \sum_{i=1}^{M} (h([\boldsymbol{A}\boldsymbol{x}]_i; y_i))^p + \beta\|\boldsymbol{x}\|_1, \tag{2}$$

where $h(t; y) \triangleq |y - |t|^q|$ is the absolute deviation for a single measurement y, the positive integer p controls the type of data discrepancy, and the regularization parameter $\beta > 0$ is selected to balance the error and sparsity terms. When $p = 2$, the data discrepancy term becomes a least squares penalty consistent with a Gaussian noise model; here, we focus on the choice $p = 1$ to improve robustness to outliers.

In either case, $\Psi(\boldsymbol{x})$ is nonconvex. Thus, to solve (2), we derive a convex majorizer $\Phi(\boldsymbol{x}; \boldsymbol{s})$ and use multiple random initial majorization points $\boldsymbol{s}$ to estimate $\boldsymbol{x}$.

3. OPTIMIZATION TRANSFER ALGORITHM

Optimization transfer [21–23] employs a surrogate function, or majorizer, in place of a more difficult objective function. The surrogate function $\Phi(\boldsymbol{x}; \boldsymbol{x}^{(n)})$ should satisfy $\Phi(\boldsymbol{x}; \boldsymbol{x}^{(n)}) \geq \Psi(\boldsymbol{x}), \forall \boldsymbol{x}$, with equality at $\boldsymbol{x} = \boldsymbol{x}^{(n)}$. Solving $\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \Phi(\boldsymbol{x}; \boldsymbol{x}^{(n)})$ for I_{ox} iterations yields a monotonically decreasing sequence $\{\Psi(\boldsymbol{x}^{(n)})\}$ for the original objective function.

3.1. Constructing the Majorizer

We begin constructing our majorizer by examining the single prediction error function $h(t; y)$, where $t \in \mathbb{C}$, and $y \in \mathbb{R}$. By definition,

$$h(t; y) = \max\{h_+(t; y) \triangleq |t|^q - y, \; h_-(t; y) \triangleq y - |t|^q\}. \tag{3}$$

Since $h_+(t; y)$ is already convex, but $h_-(t; y)$ is concave, we majorize $h_-(t; y)$ using a tangent plane $\phi_-(t; s, y)$ around $s \in \mathbb{C}$:

$$\phi_-(t; s, y) \triangleq y - |s|^q - q|s|^{q-1}\text{Re}\{e^{-i\angle s}(t - s)\} \qquad (4)$$

$$= y + (q-1)|s|^q - q|s|^{q-1}\text{Re}\{e^{-i\angle s}t\}. \qquad (5)$$

When $0 < y \leq |s|^q$, we get a tighter majorizer by replacing s in (5) with $y^{1/q}e^{i\angle s}$. The majorizer for $h(t; y)$ is

$$\phi(t; s, y) = \begin{cases} |t|^q - y, & y \leq 0, \\ \max\{|t|^q - y, y + (q-1)|s|^q - \\ \quad q|s|^{q-1}\text{Re}\{e^{-i\angle s}t\}\}, & 0 \leq |s|^q < y, \\ \max\{|t|^q - y, qy - \\ \quad qy^{(q-1)/q}\text{Re}\{e^{-i\angle s}t\}\}, & 0 < y \leq |s|^q. \end{cases} \qquad (6)$$

The function $h(t; y)$ plotted in Figure 1 is circularly symmetric and contains distinctly convex and nonconvex regions for $y > 0$. The majorizer also portrayed in this figure replaces the nonconvex region of $h(t; y)$ by the plane $\phi_-(t; s, y)$.
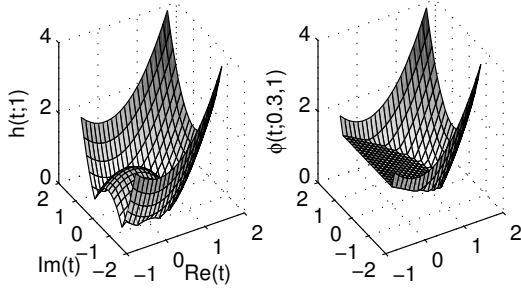

Fig. 1. The data discrepancy $h(t; y)$ (left) and majorizer $\phi(t; s, y)$ (right) are plotted for complex t, $y = 1$, and $s = 0.3$.

The complete majorizer for $\Psi(\boldsymbol{x})$ is then

$$\Phi(\boldsymbol{x}; \boldsymbol{s}) = \sum_{i=1}^M (\phi([\boldsymbol{A}\boldsymbol{x}]_i; s_i, y_i))^p + \beta\|\boldsymbol{x}\|_1, \qquad (7)$$

where $\boldsymbol{s} = \boldsymbol{A}\boldsymbol{x}^{(n)}$. Using $h_+(t; y) = -h_-(t; y)$, and $h_-(t; y) \leq \phi_-(t; s, y)$, one can easily see:

1. $h_+(t; y) \geq \phi_-(t; s, y)$ implies $h_+(t; y) \geq |\phi_-(t; s, y)|$.
2. $\phi_-(t; s, y) \geq h_+(t; y)$ implies $\phi_-(t; s, y) \geq |h_+(t; y)|$.
3. $(\max\{h_+(t; y), \phi_-(t; s, y)\})^p = \max\{(h_+(t; y))^p, (\phi_-(t; s, y))^p\}$, for any positive p.

Algorithm 1   Optimization transfer for minimizing (2)

Input: $\boldsymbol{s}_0 \in \mathbb{C}^M$, $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, $\boldsymbol{y} \in \mathbb{R}^M$, $\beta, \mu, \varepsilon_{\text{ox}}, \varepsilon_{\text{ADMM}} \in \mathbb{R}_+$, $I_{\text{ox}}, I_{\text{ADMM}} \in \mathbb{N}$
Output: $\hat{\boldsymbol{x}}$ - an approximate minimizer of (2)

Initialization: Set $\boldsymbol{x}^{(0)} = \boldsymbol{0}$, $\boldsymbol{s}^{(0)} = \boldsymbol{s}_0$, $f^{(0)} = \Psi(\boldsymbol{x}^{(0)})$
General Step: Given $\boldsymbol{s}^{(n)}$, do
1. $\boldsymbol{x}^{(n+1)} = \text{ADMM}(\boldsymbol{x}^{(n)}, \boldsymbol{s}^{(n)}, \boldsymbol{A}, \boldsymbol{y}, \beta, \mu, \varepsilon_{\text{ADMM}}, I_{\text{ADMM}})$, which is specified in Algorithm 2, to solve (7).
2. $\boldsymbol{s}^{(n+1)} = \boldsymbol{A}\boldsymbol{x}^{(n+1)}$
3. $f^{(n+1)} = \Psi(\boldsymbol{x}^{(n+1)})$
Stopping Rule: STOP if $f^{(n)} - f^{(n+1)} < \varepsilon_{\text{ox}}$, or $n + 1 \geq I_{\text{ox}}$

3.2. ADMM Algorithm

Minimizing (7) is challenging because each $\phi(\cdot; s_i, y_i)$ term generally depends on the entire vector $\boldsymbol{x}$. We introduce the auxiliary variable $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$ and write the augmented Lagrangian (AL)

$$\sum_{i=1}^M (\phi(u_i; s_i, y_i))^p + \beta\|\boldsymbol{x}\|_1 + \tfrac{\mu}{2}\|\boldsymbol{u} - \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2, \qquad (8)$$

where $\boldsymbol{b}$ and μ are the scaled dual vector and AL penalty parameter, respectively.

The alternating direction method of multipliers (ADMM) algorithm updates the primal variable $\boldsymbol{x}$, the auxiliary variable $\boldsymbol{u}$, and the scaled dual variable $\boldsymbol{b}$ separately and repeatedly, for I_{ADMM} iterations. The dual variable update $\boldsymbol{b}^{(n+1)} = \boldsymbol{b}^{(n)} + (\boldsymbol{A}\boldsymbol{x}^{(n+1)} - \boldsymbol{u}^{(n+1)})$ ensures we solve (7), with μ in (8) held fixed. The other steps are performed as follows.

3.2.1. Updating $\boldsymbol{x}$

Holding $\boldsymbol{u}$ and $\boldsymbol{b}$ fixed, the optimization problem for $\boldsymbol{x}$ becomes

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \beta\|\boldsymbol{x}\|_1 + \tfrac{\mu}{2}\|\boldsymbol{A}\boldsymbol{x} - (\boldsymbol{u}^{(n)} - \boldsymbol{b}^{(n)})\|_2^2. \qquad (9)$$

When $\boldsymbol{A}$ is left-unitary, such as a full DFT matrix, $\boldsymbol{x}^{(n+1)} = \text{S}_1(\boldsymbol{A}'(\boldsymbol{u}^{(n)} - \boldsymbol{b}^{(n)}); \tfrac{\beta}{\mu})$, where

$$\text{S}_1(x; \tau) = x(1 - \tfrac{\tau}{|x|}) \cdot \mathbf{1}_{\{|x| > \tau\}} \qquad (10)$$

is the element-wise soft-thresholding shrinkage operator. Otherwise, numerous algorithms exist for solving (9), including IST [28,29] and its momentum-accelerated variant FISTA [30]. To approximate the optimal $\boldsymbol{x}^{(n+1)}$, we run several iterations of FISTA:

$$\boldsymbol{x}^{(j+1)} = \text{S}_1(\boldsymbol{z}^{(j)} + \tfrac{\mu}{c}\boldsymbol{A}'(\boldsymbol{u} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{z}^{(j)}); \tfrac{\beta}{c}), \qquad (11)$$

$$t^{(j+1)} = \tfrac{1 + \sqrt{1 + 4t^{(j)2}}}{2}, \qquad (12)$$

$$\boldsymbol{z}^{(j+1)} = \boldsymbol{x}^{(j+1)} + \tfrac{t^{(j)} - 1}{t^{(j+1)}}(\boldsymbol{x}^{(j+1)} - \boldsymbol{x}^{(j)}), \qquad (13)$$

where $t^{(0)} = 1$, $\boldsymbol{z}^{(0)} = \boldsymbol{x}^{(0)}$, and the curvature c satisfies $c\boldsymbol{I} \succeq \mu\boldsymbol{A}'\boldsymbol{A}$ ($c = 1$ for right-unitary $\boldsymbol{A}$, such as the undersampled DFT).

3.2.2. Updating $\boldsymbol{u}$

Holding $\boldsymbol{x}$ and $\boldsymbol{b}$ fixed, the $\boldsymbol{u}$-update subproblem becomes additively separable, so each u_i can be updated independently:

$$u_i^{(n+1)} = \arg\min_u (\phi(u; s_i, y_i))^p + \tfrac{\mu}{2}|u - d_i|^2, \qquad (14)$$

where $d_i \triangleq [\boldsymbol{A}\boldsymbol{x}^{(n+1)} + \boldsymbol{b}^{(n)}]_i$. For convenience, we drop the subscripts in the following derivation.

When $y > 0$, assume $|s|^q \leq y$; otherwise, replace s with $y^{1/q}e^{i\angle s}$. Define $f_1(u) = (h_+(u; y))^p + \eta|u - d|^2$, and $f_2(u) = (\phi_-(u; s, y))^p + \eta|u - d|^2$, where $\eta \triangleq \tfrac{\mu}{2}$. Then, (14) is equivalent to the convex constrained problem

$$u^{(n+1)} = \arg\min_{u \in \mathbb{C}} \min_{T \in \mathbb{R}} T, \text{ s. t. } f_1(u) \leq T, f_2(u) \leq T. \qquad (15)$$

The Lagrangian of (15) is $T + \gamma_1(f_1(u) - T) + \gamma_2(f_2(u) - T)$, where $\gamma_1, \gamma_2 \geq 0$ are Lagrange multipliers. Differentiating with respect to T yields $\gamma_1 + \gamma_2 = 1$. Thus, we have three cases to consider: (1)

$\gamma_1 = 1, \gamma_2 = 0$; (2) $\gamma_1 = 0, \gamma_2 = 1$; or (3) $\gamma_1, \gamma_2 > 0$. In the first case, the optimal u minimizes $f_1(u)$; in the second, u minimizes $f_2(u)$. In the third, by complementary slackness, $f_1(u) = f_2(u) = T$, so we minimize $f_1(u)$ over the set of u where $f_1(u) = f_2(u)$. When, $y \leq 0$, the u-update simply minimizes $f_1(u)$. What follows are derivations for each case when $p = 1, q = 2$:

1. Minimization of $f_1(u)$. Completing the square, the objective function becomes $(1+\eta)|u - \frac{\eta}{1+\eta}d|^2 + (\frac{\eta}{1+\eta}|d|^2 - y)$. The minimizer is $u_{(1)} = \frac{\eta}{1+\eta}d$.

2. Minimization of $f_2(u)$. Again, completing the square,

$$\eta|u - e|^2 + (y + |s|^2 + \eta|d|^2 - \eta|e|^2),$$

where $e \triangleq \frac{1}{\eta}s + d$. The minimum is $u_{(2)} = e$.

3. Minimization of $f_1(u) = f_2(u)$. These two functions are equal on the circle $|u + s|^2 = 2(y + |s|^2)$. Defining $c_1 = \sqrt{2(y + |s|^2)}$, and parameterizing this circle using the angle θ, we minimize $f_1(u)$ along the curve $u(\theta) = c_1 e^{i\theta} - s$. Substituting and grouping the terms that do not contain θ,

$$f_1(u(\theta)) = c_2 - 2c_1 \operatorname{Re}\{((1+\eta)s + \eta d)^* e^{i\theta}\}, \quad (16)$$

where $c_2 \triangleq (1+\eta)c_1^2 + |s|^2 + \eta|s + d|^2 - y$. So, minimizing $f_1(u(\theta))$ is the same as maximizing $\cos(\theta - \angle((1+\eta)s + \eta d))$, which is accomplished at $\theta = \angle((1+\eta)s + \eta d)$, leaving $u_{(3)} = c_1 e^{i\angle((1+\eta)s+\eta d)} - s$.

If $f_1(u_{(1)}) \geq f_2(u_{(1)})$, then $u_{(1)}$ is optimal; conversely, if $f_2(u_{(2)}) \geq f_1(u_{(2)})$, then $u_{(2)}$ is optimal. If neither $u_{(1)}$ nor $u_{(2)}$ is optimal, then $u_{(3)}$ is the solution.

Algorithm 2 ADMM for minimizing AL in (7)

Input: $x_0 \in \mathbb{C}^N$, $s \in \mathbb{C}^M$, $A \in \mathbb{C}^{M \times N}$, $y \in \mathbb{R}^M$, $\beta, \mu, \varepsilon_{\text{ADMM}} \in \mathbb{R}_+$, $I_{\text{ADMM}} \in \mathbb{N}$
Output: $\hat{x}$ - a minimizer of (7)

Initialization: Set $x^{(0)} = x_0$, $u^{(0)} = Ax_0$, $b^{(0)} = 0$, $f^{(0)} = \Phi(x_0; s)$
General Step: Given $x^{(n)}, u^{(n)}, b^{(n)}$, calculate:
1. $x^{(n+1)} = \arg\min_x \frac{\mu}{2}\|Ax - (u^{(n)} - b^{(n)})\|_2^2 + \beta\|x\|_1$. The solution is given in Section 3.2.1.
2. FOR $i = 1, \ldots, M$,
  $u_i^{(n+1)} = \arg\min_u \phi(u; s_i, y_i) + \frac{\mu}{2}|u - d_i|^2$, where $d_i = [Ax^{(n+1)} + b^{(n)}]_i$. The solution is given in Section 3.2.2.
3. $b^{(n+1)} = b^{(n)} + (Ax^{(n+1)} - u^{(n+1)})$
4. $f^{(n+1)} = \Phi(x^{(n+1)}; s)$, $r = \|Ax^{(n+1)} - u^{(n+1)}\|_2$
Stopping Rule: STOP if either both $f^{(n)} - f^{(n+1)} < \varepsilon_{\text{ADMM}}$ and $r < \varepsilon_{\text{ADMM}}$, or $n + 1 \geq I_{\text{ADMM}}$

3.3. ADMM Parameter Selection

In theory, the AL penalty parameter μ affects the rate of convergence, but not the converged result. However, since we run ADMM for a finite number of iterations, we must choose μ appropriately to ensure both the majorizer objective $\Phi(x^{(n+1)}; s)$ and the primal residual $\|u^{(n+1)} - Ax^{(n+1)}\|$ are nearly minimized. We use an adaptive method, described in [27], for updating μ that balances the primal residual and the dual residual $\|\mu A'(u^{(n+1)} - u^{(n)})\|$. In our experiments, we start with $\mu = 1$, and double or halve μ when the primal or dual residual exceeds ten times the other. More problem-specific techniques may yield faster convergence and will be the subject of further research.

4. SIMULATION RESULTS

Throughout the simulations that follow, we generate K-sparse signals with randomly chosen support and uniformly random magnitudes and phases for those complex nonzeros. Noise-free samples of these signals correspond to squared-magnitude measurements of the full or randomly undersampled unitary DFT. In some simulations, we add *iid* Gaussian noise or introduce a fixed number of random outliers. These outliers are fixed to twice the maximum measurement value. We initialize our algorithm with multiple random s_0, where each s_i has magnitude $\max\{0, y_i\}^{1/q}$ and random phase uniformly distributed in $[0, 2\pi)$. We run our main algorithm for at most $I_{\text{ox}} = 10$ outer iterations and $I_{\text{ADMM}} = 50$ inner iterations, with stopping criteria $\varepsilon_{\text{ox}} = \varepsilon_{\text{ADMM}} = 10^{-10}$. When $M < N$, we compute $x^{(n+1)}$ using 5 iterations of FISTA [30].

To evaluate the recovered signal, the signal $\hat{x}$ corresponding to the lowest $\Psi(\hat{x})$ among outputs from multiple initializers (our best guess of the global minimizer) is compared to the true signal. After thresholding to identify the support, we find the best match accounting for circular shifts, reversal, and constant phase offsets. For this problem, errors in the actual nonzero values are less important than support recovery, as these errors can be corrected later, having estimated accurately the support.

4.1. Sensitivity to β

To recover a sparse signal, one must appropriately set the regularization parameter β; generally, $\hat{x}$ will have fewer nonzeros as β increases. We examine this relationship by simulating occurrences of correct support detection as a function of β for sparsity $K \in \{3, 5, 6, 8\}$. For 50 trials, we evaluate the outputs of the proposed method applied to noiseless (squared-magnitude) measurements of the full unitary DFT ($M = N = 128$) and the randomly undersampled unitary DFT ($M = 64, N = 128$) against the ground truth and report the percentage of correctly recovered supports. For each trial, we use 40 initializations and use $\hat{x}$ with the smallest $\Psi(\hat{x})$ as our reconstructed signal.
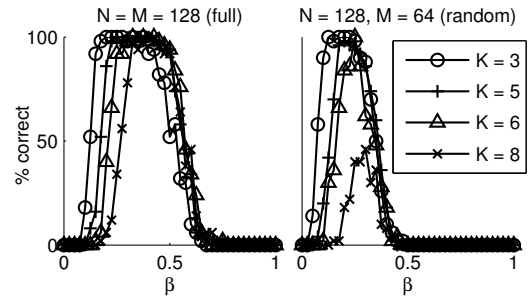

Fig. 2. The correct detection percentage varies with β; the proposed method is used to recover randomly generated signals ($N = 128$) with sparsities $K \in \{3, 5, 6, 8\}$ from $M = 128$ (left) and $M = 64$ randomly undersampled (right) noise-free DFT measurements.

Figure 2 shows that the correct detection probability depends strongly on β. In the fully sampled case, the algorithm tends to recover the support suboptimally outside the range of appropriate $\beta \in [0.3, 0.5]$. Thus, in the experiments that follow, we make sure to operate in the optimal β regime for each K used. In the $M = 64$ case, the optimal choices of β decrease significantly, and in the $K = 8$ case, even the best β yields $< 50\%$ correct detections.

4.2. Correctness and Robustness

Using the optimized choices of β determined previously for the noise-free case, we examine both the correctness and robustness of our proposed method. Consistent with previous experiments, our correctness measure corresponds to the probability of correct support recovery from the $\hat{x}$ with smallest $\Psi(\hat{x})$. We define robustness as the fraction of initializations that yield reconstructions with the correct support. The gap between correctness and robustness characterizes the effect of nonconvexity on the support recovery. When robustness is much less than the probability of correctness, more initializations are needed to achieve the best possible results. We run Monte Carlo simulations for 50 randomly generated signals with a given sparsity K, using M DFT measurements and run our algorithm with 40 random initial choices of s. We repeat these simulations for various K and M, for noise-free and noisy cases. Results for noise-free, white Gaussian noise (SNR = 40 dB), and two-outlier simulations are shown in Figures 3 and 4.
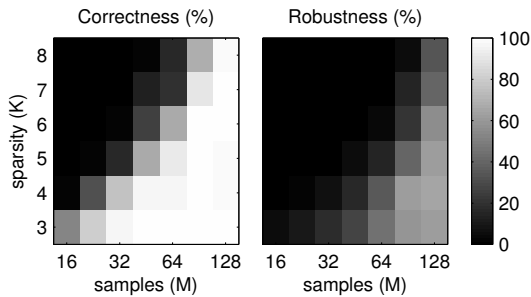

Fig. 3. The correctness (left) and robustness (right) are empirically determined using randomly generated signals ($N = 128$) with varying sparsities K from M noise-free DFT measurements (50 trials per (K, M) pair).

In the noiseless case, we observe a trend of increasing correctness and robustness as K decreases or M increases. The larger the gap between correctness and robustness, we hypothesize more initializations are necessary to minimize the chances of not finding a global minimizer. When too few initializations are used (not shown), we observe that correctness declines for large M and small K.
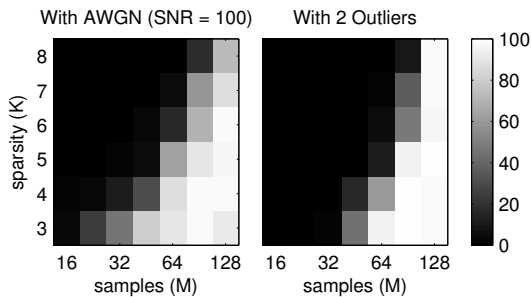

Fig. 4. The correctness is determined using random signals ($N = 128$) with varying sparsities K from M DFT measurements with either additive white Gaussian noise (SNR = 40 dB) (left) or two randomly distributed outliers (right) (50 trials per (K, M) pair).

We observe similar performance in both noisy cases, regardless of whether the noise is Gaussian or heavy-tailed in nature. Our proposed method attains similar levels of correctness with 40 dB Gaussian noise or with two outliers (set to twice the maximum measurement value), with the same β's as before in the fully sampled case

($M = N = 128$). However, more measurements are generally required for a given sparsity level to achieve a high detection fraction.

To demonstrate our algorithm's correctness in a more realistic environment, we simulated an image composed of circular discs arranged to mimic the Star of David, similar to the image used in [31]. The ground truth and reconstruction from $N = 512 \times 512$, $M = N/2$ randomly undersampled measurements corrupted by 10 outliers are shown in Figure 5. A dictionary of disc translates were employed in the reconstruction.
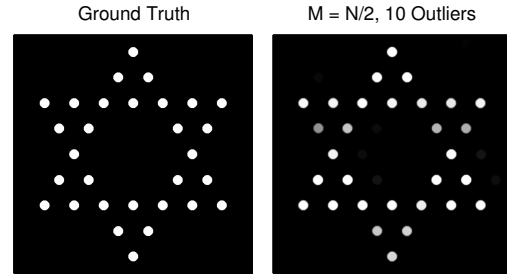

Fig. 5. The ground truth (left) and reconstruction (right) of the Star of David test image from undersampled ($M = N/2$) data corrupted by 10 outliers.

5. DISCUSSION AND CONCLUSION

The simulations demonstrate the effectiveness of the proposed method in terms of correct support recovery from magnitude-only measurements for a range of sparsities and undersampling factors, even in the presence of noise or outliers. However, the detection ability is highly dependent on appropriately choosing the regularization parameter β. Over- or under-regularization generally yields reconstructions without the desired sparse support. Automatic methods for choosing β for this problem, akin to Monte Carlo SURE [32] or nonlinear generalized cross-validation [33, 34], would simplify practical use. This challenge is similar to choosing the sparsity K for greedy sparse reconstruction methods like GESPAR [14]. Furthermore, the computational efficiency of this method is tied to the robustness measure of how likely a random initial condition is to yield the desired solution. While the per-iteration cost is similar to other iterative methods, a large number of iterations and initializations may be necessary for consistent support recovery. Our majorizer provides a convenient convex relaxation of the problem that may simplify the challenge of finding optimal initial conditions. This approach also generalizes to other separable convex regularizers beyond the 1-norm.

In future work, we plan to evaluate the proposed method against existing methods like GESPAR, extend the objective function (2) to analysis-form transform sparsity, compare different choices of p and q. and investigate the aforementioned problems of automatic parameter selection and initialization of the majorizer.

To conclude, we proposed a novel phase retrieval method based on majorization, variable splitting, and ADMM. Through Monte Carlo trials, we demonstrated our method yields correct and robust reconstructions in a wide range of cases, and we evaluated the sensitivity to the regularization parameter β.

6. REFERENCES

[1] J. R. Fienup, "Phase retrieval algorithms: a personal tour," *Appl. Optics*, vol. 52, no. 1, pp. 45–56, Jan. 2013, Invited

paper.

[2] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging," 2014, Submitted to IEEE Signal Processing Magazine. arXiv:1402.7350.

[3] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Am. A*, vol. 7, no. 3, pp. 394–411, Mar. 1990.

[4] H. A. Hauptman, "The phase problem of x-ray crystallography," *Rep. Prog. Phys.*, vol. 54, no. 11, pp. 1427–1454, Nov. 1991.

[5] R. W. Harrison, "Phase problem in crystallography," *JOSA A*, vol. 10, no. 5, pp. 1046–1055, May 1993.

[6] A. Walther, "The question of phase retrieval in optics," *Optica Acta*, vol. 10, no. 1, pp. 41–49, 1963.

[7] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, pp. 342–344, July 1999.

[8] T. Latychevskaia, J.-N. Longchamp, and H.-W. Fink, "Novel Fourier-domain constraint for fast phase retrieval in coherent diffraction imaging," *Optics Express*, vol. 19, no. 20, pp. 19330–19339, Sept. 2011.

[9] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–46, Apr. 1972.

[10] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Optics*, vol. 21, no. 15, pp. 2758–69, Aug. 1982.

[11] K. Jaganathan, S. Oymak, and B. Hassibi, "Recovery of sparse 1-d signals from the magnitudes of their Fourier transform," 2012, arXiv:1206.1405.

[12] S. Mukherjee and C. S. Seelamantula, "An iterative algorithm for phase retrieval with sparsity constraints: application to frequency domain optical coherence tomography," in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, 2012, pp. 553–6.

[13] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," *Applied and Computational Harmonic Analysis*, vol. 36, no. 3, pp. 473–494, May 2014.

[14] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: efficient phase retrieval of sparse signals," *IEEE Trans. Sig. Proc.*, vol. 62, no. 4, pp. 928–38, Feb. 2014.

[15] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "Compressive phase retrieval from squared output measurements via semidefinite programming," 2012, arxiv 1111.6323.

[16] Y. Shechtman, Y. C. Eldar, A. Szameit, and M. Segev, "Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing," *Optics Express*, vol. 19, no. 16, pp. 14807–22, Aug. 2011.

[17] I. Waldspurger, A. d'Aspremont, and Stéphane Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, 2014, To appear.

[18] X. Li and V. Voroninski, "Sparse signal recovery from quadratic measurements via convex programming," *SIAM J. Math. Anal.*, vol. 45, no. 5, pp. 3019–33, 2013.

[19] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming," *Comm. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–74, Aug. 2013.

[20] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 199–225, 2013.

[21] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *J. Computational and Graphical Stat.*, vol. 9, no. 1, pp. 1–20, Mar. 2000.

[22] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Trans. Im. Proc.*, vol. 16, no. 10, pp. 2411–22, Oct. 2007.

[23] E. Chouzenoux, J. C. Pesquet, and A. Repetti, "A block coordinate variable metric forward-backward algorithm," 2013, Available from Optimization Online.

[24] R. Glowinski and A. Marrocco, "Sur lapproximation par elements nis dordre un, et la resolution par penalisation-dualite dune classe de problemes de dirichlet nonlineaires, rev. francaise daut," *Inf. Rech. Oper.*, vol. R-2, pp. 41–76, 1975.

[25] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximations," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[26] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992.

[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[28] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Im. Proc.*, vol. 12, no. 8, pp. 906–16, Aug. 2003.

[29] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–57, Nov. 2004.

[30] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[31] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. C. Eldar, O. Cohen, and M. Segev, "Sparsity-based single-shot subwavelength coherent diffractive imaging," *Nature Materials*, vol. 11, pp. 455–9, Apr. 2012.

[32] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Im. Proc.*, vol. 17, no. 9, pp. 1540–54, Sept. 2008.

[33] D. A. Girard, "The fast Monte-Carlo cross-validation and CL procedures: Comments, new results and application to image recovery problems," *Comput. Statist.*, vol. 10, no. 3, pp. 205–31, 1995.

[34] S. Ramani, Z. Liu, J. Rosen, J-F. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Trans. Im. Proc.*, vol. 21, no. 8, pp. 3659–72, Aug. 2012.