Compressive Phase Retrieval From Squared Output Measurements Via Semidefinite Programming

Henrik Ohlsson[†,*], Allen Y. Yang[*], Roy Dong[*], and
S. Shankar Sastry[*]

[†]Division of Automatic Control, Department of Electrical
Engineering, Linköping University, Sweden.
[*]Department of Electrical Engineering and Computer Sciences,
University of California at Berkeley, CA, USA,
{ohlsson,yang,roydong,sastry}@eecs.berkeley.edu.

Abstract

Given a linear system in a real or complex domain, linear regression aims to recover the model parameters from a set of observations. Recent studies in compressive sensing have successfully shown that under certain conditions, a linear program, namely, ℓ_1-minimization, guarantees recovery of sparse parameter signals even when the system is underdetermined. In this paper, we consider a more challenging problem: when the phase of the output measurements from a linear system is omitted. Using a lifting technique, we show that even though the phase information is missing, the sparse signal can be recovered exactly by solving a simple semidefinite program when the sampling rate is sufficiently high, albeit the exact solutions to both sparse signal recovery and phase retrieval are combinatorial. The results extend the type of applications that compressive sensing can be applied to those where only output magnitudes can be observed. We demonstrate the accuracy of the algorithms through theoretical analysis, extensive simulations and a practical experiment.

1   Introduction

Linear models, e.g. $\boldsymbol{y} = A\boldsymbol{x}$, are by far the most used and useful type of model. The main reasons for this are their simplicity of use and identification. For the identification, the least-squares (LS) estimate in a complex

1

domain is computed by[1]

$$\boldsymbol{x}_{\mathrm{ls}} = \operatorname*{argmin}_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 \in \mathbb{C}^n, \tag{1}$$

assuming the output $\boldsymbol{y} \in \mathbb{C}^N$ and $A \in \mathbb{C}^{N \times n}$ are given. Further, the LS problem has a unique solution if the system is full rank and not underdetermined, i.e. $N \geq n$.

Consider the alternative scenario when the system is underdetermined, i.e. $n > N$. The least squares solution is no longer unique in this case, and additional knowledge has to be used to determine a unique model parameter. Ridge regression or Tikhonov regression [Hoerl and Kennard, 1970] is one of the traditional methods to apply in this case, which takes the form

$$\boldsymbol{x}_{\mathrm{r}} = \operatorname*{argmin}_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_2^2, \tag{2}$$

where $\lambda > 0$ is a scalar parameter that decides the trade off between fit in the first term and the ℓ_2-norm of $\boldsymbol{x}$ in the second term.

Thanks to the ℓ_2-norm regularization, ridge regression is known to pick up solutions with small energy that satisfy the linear model. In a more recent approach stemming from the LASSO [Tibsharani, 1996] and compressive sensing (CS) [Candès et al., 2006, Donoho, 2006], another convex regularization criterion has been widely used to seek the *sparsest* parameter vector, which takes the form

$$\boldsymbol{x}_{\ell_1} = \operatorname*{argmin}_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1. \tag{3}$$

Depending on the choice of the weight parameter λ, the program (3) has been known as the LASSO by Tibsharani [1996], *basis pursuit denoising* (BPDN) by Chen et al. [1998], or ℓ_1-minimization (ℓ_1-min) by Candès et al. [2006]. In recent years, several pioneering works have contributed to efficiently solving sparsity minimization problems such as [Tropp, 2004, Beck and Teboulle, 2009, Bruckstein et al., 2009], especially when the system parameters and observations are in high-dimensional spaces.

In this paper, we consider a more challenging problem. In a linear model $\boldsymbol{y} = A\boldsymbol{x}$, rather than assuming that $\boldsymbol{y}$ is given, we will assume that only the squared magnitude of the output is observed:

$$b_i = |y_i|^2 = |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle|^2, \quad i = 1, \cdots, N, \tag{4}$$

---

[1]Our derivation in this paper is primarily focused on complex signals, but the results should be easily extended to real domain signals.

2

where $A^H = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_N] \in \mathbb{C}^{n \times N}$, $\boldsymbol{y}^T = [y_1, \cdots, y_N] \in \mathbb{C}^{1 \times N}$ and A^H denotes the Hermitian transpose of A. This is clearly a more challenging problem since the phase of $\boldsymbol{y}$ is lost when only its (squared) magnitude is available. A classical example is that $\boldsymbol{y}$ represents the Fourier transform of $\boldsymbol{x}$, and that only the Fourier transform modulus is observable. This scenario arises naturally in several practical applications such as optics [Walther, 1963, Millane, 1990], coherent diffraction imaging [Fienup, 1987], and astronomical imaging [Dainty and Fienup, 1987] and is known as the *phase retrieval* problem.

We note that in general phase cannot be uniquely recovered regardless whether the linear model is overdetermined or not. A simple example to see this, is if $\boldsymbol{x}_0 \in \mathbb{C}^n$ is a solution to $\boldsymbol{y} = A\boldsymbol{x}$, then for any scalar $c \in \mathbb{C}$ on the unit circle $c\boldsymbol{x}_0$ leads to the same squared output $\boldsymbol{b}$. As mentioned in [Candès et al., 2011a], when the dictionary A represents the unitary discrete Fourier transform (DFT), the ambiguities may represent time-reversed solutions or time-shifted solutions of the ground truth signal $\boldsymbol{x}_0$. These global ambiguities caused by losing the phase information are considered acceptable in phase retrieval applications. From now on, when we talk about the solution to the phase retrieval problem, it is the solution up to a global phase ambiguity. Accordingly, a unique solution is a solution unique up to a global phase.

Further note that since (4) is nonlinear in the unknown $\boldsymbol{x}$, $N \gg n$ measurements are in general needed for a unique solution. When the number of measurements N are fewer than necessary for a unique solution, additional assumptions are needed to select one of the solutions (just like in Tikhonov, Lasso and CS).

Finally, we note that the exact solution to either CS and phase retrieval is combinatorially expensive [Chen et al., 1998, Candès et al., 2011c]. Therefore, the goal of this work is to answer the following question: *Can we effectively recover a sparse parameter vector $\boldsymbol{x}$ of a linear system up to a global ambiguity using its squared magnitude output measurements via convex programming?* The problem is referred as *compressive phase retrieval* (CPR) [Moravec et al., 2007].

The main contribution of the paper is a *convex formulation* of the sparse phase retrieval problem. Using a lifting technique, the NP-hard problem is relaxed as a semidefinite program. We also derive bounds for guaranteed recovery of the true signal and compare the performance of our CPR algorithm with traditional CS and PhaseLift [Candès et al., 2011a] algorithms through extensive experiments. The results extend the type of applications that compressive sensing can be applied to; namely, applications where only

3

magnitudes can be observed.

1.1 Background

Our work is motivated by the ℓ_1-min problem in CS and a recent PhaseLift technique in phase retrieval by Candès et al. [2011c]. On one hand, the theory of CS and ℓ_1-min has been one of the most visible research topics in recent years. There are several comprehensive review papers that cover the literature of CS and related optimization techniques in linear programming. The reader is referred to the works of [Candès and Wakin, 2008, Bruckstein et al., 2009, Loris, 2009, Yang et al., 2010]. On the other hand, the fusion of phase retrieval and matrix completion is a novel topic that has recently been studied in a selected few papers, such as [Candès et al., 2011c,a]. The fusion of phase retrieval and CS was discussed in [Moravec et al., 2007]. In the rest of the section, we briefly review the phase retrieval literature and its recent connections with CS and matrix completion.

Phase retrieval has been a longstanding problem in optics and x-ray crystallography since the 1970s [Kohler and Mandel, 1973, Gonsalves, 1976]. Early methods to recover the phase signal using Fourier transform mostly relied on additional information about the signal, such as band limitation, nonzero support, real-valuedness, and nonnegativity. The Gerchberg-Saxton algorithm was one of the popular algorithms that alternates between the Fourier and inverse Fourier transforms to obtain the phase estimate iteratively [Gerchberg and Saxton, 1972, Fienup, 1982]. One can also utilize steepest-descent methods to minimize the squared estimation error in the Fourier domain [Fienup, 1982, Marchesini, 2007]. Common drawbacks of these iterative methods are that they may not converge to the global solution, and the rate of convergence is often slow. Alternatively, Balan et al. [2006] have studied a frame-theoretical approach to phase retrieval, which necessarily relied on some special types of measurements.

More recently, phase retrieval has been framed as a low-rank matrix completion problem in [Chai et al., 2010, Candès et al., 2011a,c]. Given a system, a lifting technique was used to approximate the linear model constraint as a semidefinite program (SDP), which is similar to the objective function of the proposed method only without the sparsity constraint. The authors also derived the upper-bound for the sampling rate that guarantees exact recovery in the noise-free case and stable recovery in the noisy case.

We are aware of the work by Moravec et al. [2007], which has considered compressive phase retrieval on a random Fourier transform model. Leveraging the sparsity constraint, the authors proved that an upper-bound of

4

$O(k^2 \log(4n/k^2))$ random Fourier modulus measurements to uniquely specify k-sparse signals. Moravec et al. [2007] also proposed a greedy compressive phase retrieval algorithm. Their solution largely follows the development of ℓ_1-min in CS, and it alternates between the domain of solutions that give rise to the same squared output and the domain of an ℓ_1-ball with a fixed ℓ_1-norm. However, the main limitation of the algorithm is that it tries to solve a nonconvex optimization problem and that it assumes the ℓ_1-norm of the true signal is known. No guarantees for when the algorithm recovers the true signal can therefore be given.

2 CPR via SDP

In the noise free case, the phase retrieval problem takes the form of the feasibility problem:

$$\text{find } \boldsymbol{x} \quad \text{subj. to} \quad \boldsymbol{b} = |A\boldsymbol{x}|^2 = \{\boldsymbol{a}_i^H \boldsymbol{x} \boldsymbol{x}^H \boldsymbol{a}_i\}_{1 \leq i \leq N}, \tag{5}$$

where $\boldsymbol{b}^T = [b_1, \cdots, b_N] \in \mathbb{R}^{1 \times N}$. This is a combinatorial problem to solve: Even in the real domain with the sign of the measurements $\{\alpha_i\}_{i=1}^N \subset \{-1, 1\}$, one would have to try out combinations of sign sequences until one that satisfies

$$\alpha_i \sqrt{b_i} = \boldsymbol{a}_i^T \boldsymbol{x}, \quad i = 1, \cdots, N, \tag{6}$$

for some $\boldsymbol{x} \in \mathbb{R}^n$ has been found. For any practical size of data sets, this combinatorial problem is intractable.

Since (5) is nonlinear in the unknown $\boldsymbol{x}$, $N \gg n$ measurements are in general needed for a unique solution. When the number of measurements N are fewer than necessary for a unique solution, additional assumptions are needed to select one of the solutions. Motivated by compressive sensing, we here choose to seek the sparsest solution of CPR satisfying (5) or, equivalent, the solution to

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0, \quad \text{subj. to} \quad \boldsymbol{b} = |A\boldsymbol{x}|^2 = \{\boldsymbol{a}_i^H \boldsymbol{x} \boldsymbol{x}^H \boldsymbol{a}_i\}_{1 \leq i \leq N}. \tag{7}$$

As the counting norm $\| \cdot \|_0$ is not a convex function, following the ℓ_1-norm relaxation in CS, (7) can be relaxed as

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1, \quad \text{subj. to} \quad \boldsymbol{b} = |A\boldsymbol{x}|^2 = \{\boldsymbol{a}_i^H \boldsymbol{x} \boldsymbol{x}^H \boldsymbol{a}_i\}_{1 \leq i \leq N}. \tag{8}$$

Note that (8) is still not a linear program, as its equality constraint is not a linear equation. In the literature, a lifting technique has been extensively

5

used to reframe problems such as (8) to a standard form in semidefinite programming, such as in Sparse PCA [d'Aspremont et al., 2007].

More specifically, given the ground truth signal $\boldsymbol{x}_0 \in \mathbb{C}^n$, let $X_0 \doteq \boldsymbol{x}_0\boldsymbol{x}_0^H \in \mathbb{C}^{n \times n}$ be an induced rank-1 semidefinite matrix. Then the compressive phase retrieval (CPR) problem can be cast as[2]

$$
\begin{aligned}
\min_X \quad & \|X\|_1 \\
\text{subj. to} \quad & b_i = \text{Tr}(\boldsymbol{a}_i^H X \boldsymbol{a}_i),\ i = 1, \cdots, N, \\
& \text{rank}(X) = 1, X \succeq 0.
\end{aligned} \tag{9}
$$

This is of course still a non-convex problem due to the rank constraint. The lifting approach addresses this issue by replacing $\text{rank}(X)$ with $\text{Tr}(X)$. For a semidefinite matrix, $\text{Tr}(X)$ is equal to the sum of the eigenvalues of X (or the ℓ_1-norm on a vector containing all eigenvalues of X). This leads to an SDP

$$
\begin{aligned}
\min_X \quad & \text{Tr}(X) + \lambda\|X\|_1 \\
\text{subj. to} \quad & b_i = \text{Tr}(\Phi_i X),\ i = 1, \cdots, N, \\
& X \succeq 0,
\end{aligned} \tag{10}
$$

where we further denote $\Phi_i \doteq \boldsymbol{a}_i\boldsymbol{a}_i^H \in \mathbb{C}^{n \times n}$ and where $\lambda \geq 0$ is a design parameter. Finally, the estimate of $\boldsymbol{x}$ can be found by computing the rank-1 decomposition of X via singular value decomposition. We will refer to the formulation (10) as compressive phase retrieval via lifting (CPRL).

We compare (10) to a recent solution of PhaseLift by Chai et al. [2010], Candès et al. [2011c]. In Chai et al. [2010], Candès et al. [2011c], a similar objective function was employed for phase retrieval:

$$
\begin{aligned}
\min_X \quad & \text{Tr}(X) \\
\text{subj. to} \quad & b_i = \text{Tr}(\Phi_i X),\ i = 1, \cdots, N, \\
& X \succeq 0,
\end{aligned} \tag{11}
$$

albeit the source signal was not assumed sparse. Using the lifting technique to construct the SDP relaxation of the NP-hard phase retrieval problem, with high probability, the program (11) recovers the exact solution (sparse or dense) if the number of measurements N is at least of the order of $O(n \log n)$. The region of success is visualized in Figure 1 as region I.

If $\boldsymbol{x}$ is sufficiently sparse and random Fourier dictionaries are used for sampling, Moravec et al. [2007] showed that in general the signal is uniquely

---

[2]In this paper, $\|X\|_1$ for a matrix X denotes the entry-wise ℓ_1-norm, and $\|X\|_2$ denotes the Frobenius norm.

6

defined if the number of squared magnitude output measurements $\boldsymbol{b}$ exceeds the order of $O(k^2 \log(4n/k^2))$. This lower bound for the region of success of CPR is illustrated by the dash line in Figure 1.

Finally, the motivation for introducing the ℓ_1-norm regularization in (10) is to be able to solve the sparse phase retrieval problem for N smaller than what PhaseLift requires. However, one will not be able to solve the compressive phase retrieval problem in region III below the dashed curve. Therefore, our target problems lie in region II.
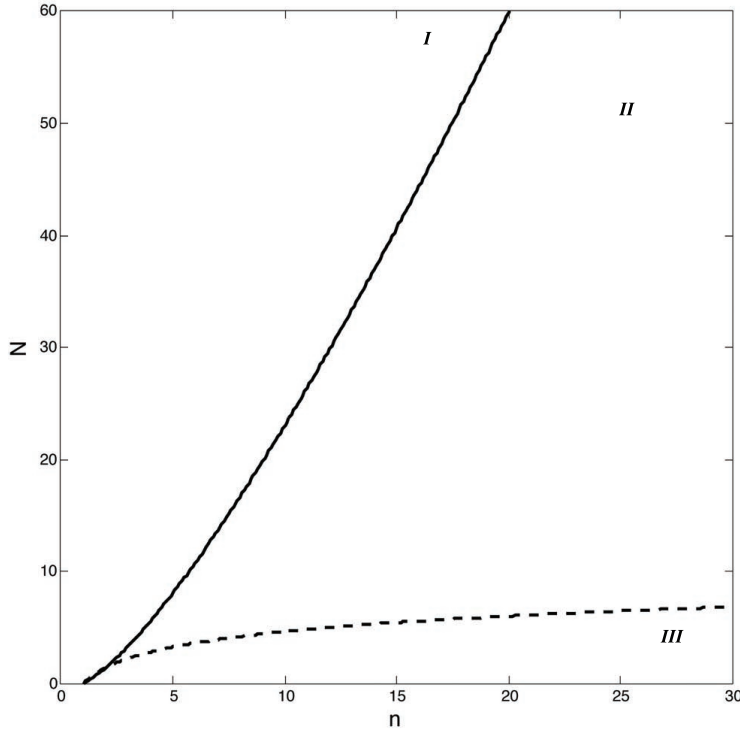
Figure 1: An illustration of the regions of importance in solving the phase retrieval problem. While PhaseLift primarily targets problems in region I, CPRL operates primarily in regions II and III.

Example 1 (Compressive Phase Retrieval). *In this example, we illustrate a simple CPR example, where a 2-sparse complex signal $\boldsymbol{x}_0 \in \mathbb{C}^{64}$ is first transformed by the Fourier transform $F \in \mathbb{C}^{64 \times 64}$ followed by random pro-*

7

jections $R \in \mathbb{C}^{32 \times 64}$ (generated by sampling a unit complex Gaussian):

$$\boldsymbol{b} = |RF\boldsymbol{x}_0|^2. \tag{12}$$

Given $\boldsymbol{b}$, F, and R, we first apply PhaseLift algorithm [Candès et al., 2011c] with $A = RF$ to the 32 squared observations $\boldsymbol{b}$. The recovered dense signal is shown in Figure 2. As seen in the figure, PhaseLift fails to identify the 2-sparse signal.

Next, we apply CPRL (15), and the recovered sparse signal is also shown in Figure 2. CPRL correctly identifies the two nonzero elements in $\boldsymbol{x}$.
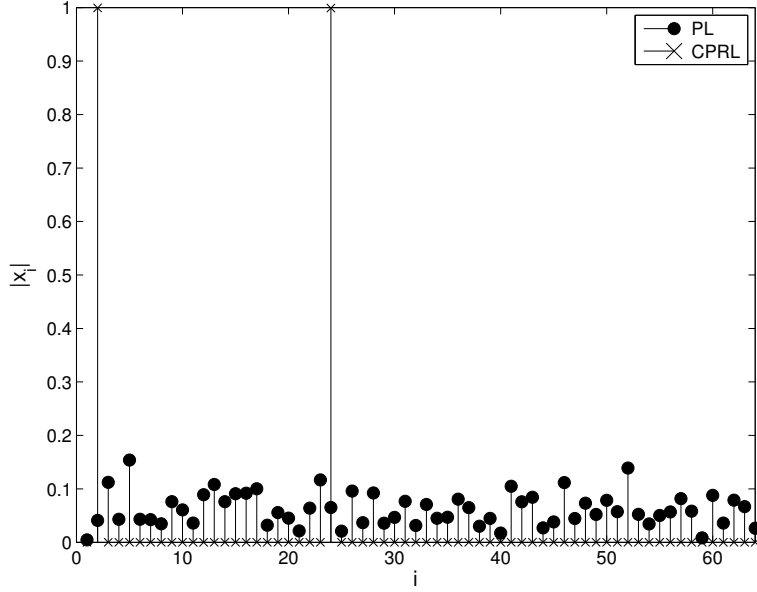
Figure 2: The magnitude of the estimated signal provided by CPRL and PhaseLift (PL). CPRL correctly identifies elements 2 and 24 to be nonzero while PhaseLift provides a dense estimate. It is also verified that the estimate from CPRL, after a global phase shift, is approximately equal the true $\boldsymbol{x}_0$.

3 Stable Numerical Solutions for Noisy Data

In this section, we consider the case that the measurements are contaminated by data noise. In a linear model, typically bounded random noise affects

8

the output of the system as $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{e} \in \mathbb{C}^N$ is a noise term with bounded ℓ_2-norm: $\|\boldsymbol{e}\|_2 \leq \epsilon$. However, in phase retrieval, we follow closely a more special noise model used in Candès et al. [2011c]:

$$b_i = |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle|^2 + e_i. \tag{13}$$

This nonstandard model avoids the need to calculate the squared magnitude output $|\boldsymbol{y}|^2$ with the added noise term. More importantly, in practical phase retrieval applications, measurement noise is introduced when the squared magnitudes or intensities of the linear system are measured, not on $\boldsymbol{y}$ itself [Candès et al., 2011c].

Accordingly, we denote a linear operator B of X as

$$B : X \in \mathbb{C}^{n \times n} \mapsto \{\mathrm{Tr}(\Phi_i X)\}_{1 \leq i \leq N} \in \mathbb{R}^N, \tag{14}$$

which measures the noise-free squared output. Then the approximate CPR problem with bounded ℓ_2 error model (13) can be solved by the following SDP program:

$$\begin{aligned} \min \quad & \mathrm{Tr}(X) + \lambda \|X\|_1 \\ \text{subj. to} \quad & \|B(X) - \boldsymbol{b}\|_2 \leq \epsilon, \\ & X \succeq 0. \end{aligned} \tag{15}$$

The estimate of $\boldsymbol{x}$, just as in noise free case, can finally be found by computing the rank-1 decomposition of X via singular value decomposition. We refer to the method as approximate CPRL. Due to the machine rounding error, in general a nonzero ϵ should be always assumed in the objective (15) and its termination condition during the optimization.

We should further discuss several numerical issues in the implementation of the SDP program. The constrained CPRL formulation (15) can be rewritten as an unconstrained objective function:

$$\min_{X \succeq 0} \mathrm{Tr}(X) + \lambda \|X\|_1 + \frac{\mu}{2} \|B(X) - \boldsymbol{b}\|_2^2, \tag{16}$$

where $\lambda > 0$ and $\mu > 0$ are two penalty parameters.

In (16), due to the lifting process, the rank-1 condition of X is approximated by its trace function $Tr(X)$. In Candès et al. [2011c], the authors considered phase retrieval of generic (dense) signal $\boldsymbol{x}$. They proved that if the number of measurements obeys $N \geq cn \log n$ for a sufficiently large constant c, with high probability, minimizing (16) without the sparsity constraint (i.e. $\lambda = 0$) recovers a unique rank-1 solution obeying $X^* = \boldsymbol{x}\boldsymbol{x}^H$. See also Recht et al. [2010].

9

In Section 7, we will show that using either random Fourier dictionaries or more general random projections, in practice, one needs much fewer measurements to exactly recover sparse signals if the measurements are noise free. Nevertheless, in the presence of noise, the recovered lifted matrix X may not be exactly rank-1. In this case, one can simply use its rank-1 approximation corresponding to the largest singular value of X.

We also note that in (16), there are two main parameters λ and μ that can be defined by the user. Typically μ is chosen depending on the level of noise that affects the measurements $\boldsymbol{b}$. For λ associated with the sparsity penalty $\|X\|_1$, one can adopt a *warm start* strategy to determine its value iteratively. The strategy has been widely used in other sparse optimization, such as in ℓ_1-min [Yang et al., 2010]. More specifically, the objective is solved iteratively with respect to a sequence of monotonically decreasing λ, and each iteration is initialized using the optimization results from the previous iteration. The procedure continues until a rank 1 solution has been found. When λ is large, the sparsity constraint outweighs the trace constraint and the estimation error constraint, and vice versa.

Example 2 (Noisy Compressive Phase Retrieval)**.** *Let us revisit Example 1 but now assume that the measurements are contaminated by noise. Using exactly the same data as in Example 1 but adding uniformly distributed measurement noise between -1 and 1, CPRL was able to recover the 2 nonzero elements. PhaseLift, just as in Example 1 gave a dense estimate of $\boldsymbol{x}$.*

4   Computational Aspects

In this section we discuss computational issus of the proposed SDP formulation, algorithms for solving the SDP and to efficient approximative solution algorithms.

4.1   A Greedy Algorithm

Since (10) is an SDP, it can be solved by standard software, such as CVX [Grant and Boyd, 2010]. However, it is well known that the standard toolboxes suffer when the dimension of X is large. We therefore propose a greedy approximate algorithm tailored to solve (10). If the number of nonzero elements in $\boldsymbol{x}$ is expected to be low, the following algorithm may be suitable and less computationally heavy compare to approaching the original SDP:

10

Algorithm 1: Greedy Compressive Phase Retrieval via Lifting (GCPRL)

Set $\mathcal{I} = \emptyset$ and let $\gamma > 0$, $\epsilon > 0$.
repeat
    for $k = 1, \cdots, N$, **do**
        Set $\mathcal{I}_k = \mathcal{I} \bigcup \{k\}$ and solve $X_k \doteq$
        $\arg\min_{X \succeq 0} \operatorname{Tr}(X^{(\mathcal{I}_k)}) + \gamma \sum_{i=1}^{N} (b_i - \operatorname{Tr}(\boldsymbol{a}_i^{(\mathcal{I}_k)} \boldsymbol{a}_i^{(\mathcal{I}_k)H} X^{(\mathcal{I}_k)}))^2$.
        Let W_k denote the corresponding objective value.
    Let p be such that $W_p \leq W_k, k = 1, \cdots, N$. Set $\mathcal{I} = \mathcal{I} \bigcup \{p\}$ and
    $X = X_p$.
until $W_p < \epsilon$;

Example 3 (GCPRL ability to solve the CPR problem). *To demonstrate the effectiveness of GCPRL let us consider a numerical example. Let the true $\boldsymbol{x}_0 \in \mathbb{C}^n$ be a k-sparse signal, let the nonzero elements be randomly chosen and their values randomly distributed on the complex unit circle. Let $A \in \mathbb{C}^{N \times n}$ be generated by sampling from a complex unit Gaussian distribution.*

If we fix $n/N = 2$, that is, twice as many unknowns as measurements, and apply GCPRL for different values of n, N and k we obtain the computational times visualized in the left plot of Figure 3. In all simulations $\gamma = 10$ and $\epsilon = 10^{-3}$ are used in GCPRL. The true sparsity pattern was always recovered. Since GCPRL can be executed in parallel, the simulation times can be divided by the number of cores used (the average run time in Figure 3 is computed on a standard laptop running Matlab, 2 cores, and using CVX to solve the low dimensional SDP of GCPRL). The algorithm is several magnitudes faster than the standard interior-point methods used in CVX.

5 The Dual

CPRL takes the form:

$$
\begin{aligned}
\min_X \quad & \operatorname{Tr}(X) + \lambda \|X\|_1 \\
\text{subj. to} \quad & b_i = \operatorname{Tr}(\Phi_i X); i = 1, \cdots, N, \\
& X \succeq 0.
\end{aligned}
\tag{17}
$$

11
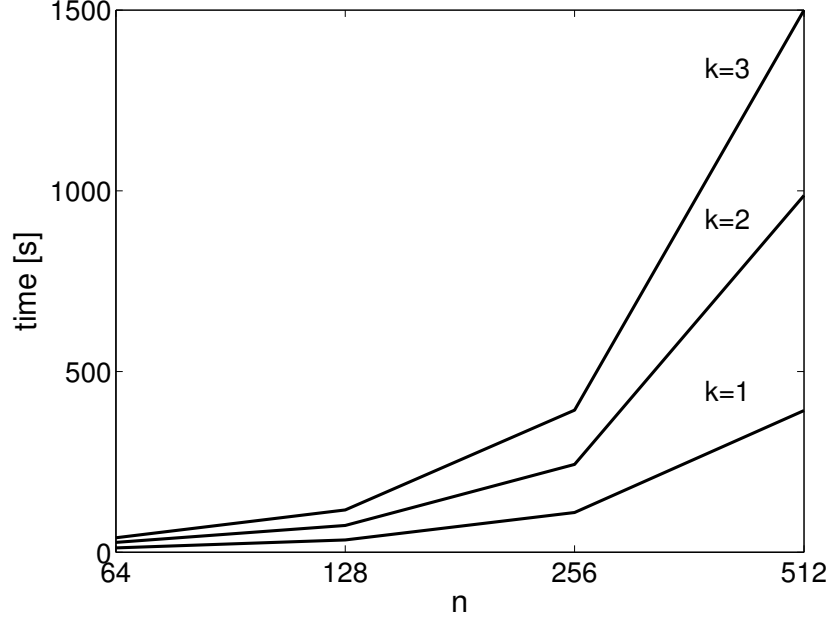
Figure 3: Average run time of GCPRL in Matlab CVX environment.

If we define $h(X)$ to be an N-dimensional vector such that our constraints are $h(X) = 0$, then we can equivalently write:

$$\min_X \max_{\mu,Y,Z=Z^H} \quad \text{Tr}(X) + \text{Tr}(ZX) + \mu^T h(X) - \text{Tr}(YX)$$
$$\text{subj. to} \quad Y \succeq 0, \tag{18}$$
$$\|Z\|_\infty \leq \lambda.$$

Then the dual becomes:

$$\max_{\mu,Z=Z^H} \quad \mu^T b$$
$$\text{subj. to} \quad \|Z\|_\infty \leq \lambda. \tag{19}$$
$$Y := I + Z - \sum_{i=1}^{N} \mu_i \Phi_i \succeq 0.$$

6   Analysis

This section contains various analysis results. The analysis follows that of CS and have been inspired by derivations given in [Candès et al., 2011c, 2006, Donoho, 2006, Candès, 2008, Berinde et al., 2008, Bruckstein et al., 2009]. The analysis is divided into three subsections. The first subsection

12

gives results based on RIP, the second based on RIP-1 and the third based on mutual coherence.

6.1 Analysis Using RIP

In order to state some theoretical properties, we need a generalization of the restricted isometry property (RIP).

Definition 4 (RIP). *We will say that a linear operator $B(\cdot)$ is (ϵ, k)-RIP if for all $X \neq 0$ s.t. $\|X\|_0 \leq k$ we have*

$$\left| \frac{\|B(X)\|_2^2}{\|X\|_2^2} - 1 \right| < \epsilon. \tag{20}$$

We can now state the following theorem:

Theorem 5 (Recoverability/Uniqueness). *Let $B(\cdot)$ be a $(\epsilon, 2\|X^*\|_0)$-RIP linear operator with $\epsilon < 1$ and let $\bar{\boldsymbol{x}}$ be the sparsest solution to (4). If X^* satisfies*

$$\begin{aligned} \mathbf{b} &= B(X^*), \\ X^* &\succeq 0, \\ \operatorname{rank}\{X^*\} &= 1, \end{aligned} \tag{21}$$

then X^* is unique and $X^* = \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$.

Proof of Theorem 5. Assume the contrary i.e., $X^* \neq \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$. It is clear that $\|\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\|_0 \leq \|X^*\|_0$ and hence $\|\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H - X^*\|_0 \leq 2\|X^*\|_0$. If we now apply the RIP inequality (20) on $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H - X^*$ and use that $B(\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H - X^*) = 0$ we are led to the contradiction $1 < \epsilon$. We therefore conclude that X^* is unique and $X^* = \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$. $\qquad\square$

We can also give a bound on the sparsity of $\bar{\boldsymbol{x}}$:

Theorem 6 (Bound on $\|\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\|_0$ from above). *Let $\bar{\boldsymbol{x}}$ be the sparsest solution to (4) and let $\tilde{X}$ be the solution of CPRL (10). If $\tilde{X}$ has rank 1 then $\|\tilde{X}\|_0 \geq \|\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\|_0$.*

Proof of Theorem 6. Let $\tilde{X}$ be a rank-1 solution of CPRL (10). By contradiction, assume $\|\tilde{X}\|_0 < \|\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\|_0$. Since $\tilde{X}$ satisfies the constraints of (4), it must give a lower objective value than $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$ in (4) . This is a contradiction since $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$ was assumed to be the solution of (4). Hence we must have that $\|\tilde{X}\|_0 \geq \|\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\|_0$. $\qquad\square$

13

Corollary 7 (Guaranteed Recovery Using RIP). *Let $\bar{\boldsymbol{x}}$ be the sparsest solution to (4). The solution of CPRL (10), $\tilde{X}$, is equal to $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$ if it has rank 1 and B is $(\epsilon, 2\|\tilde{X}\|_0)$-RIP with $\epsilon < 1$.*

Proof of Corollary 7. This follows trivially from Theorem 5 by realizing that $\tilde{X}$ satisfy all properties of X^*.

$\square$

If $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H = \tilde{X}$ can not be guaranteed, the following bound could come useful:

Theorem 8 (Bound on $\|X^* - \tilde{X}\|_2$). *Let $\epsilon < \frac{1}{1+\sqrt{2}}$ and assume $B(\cdot)$ to be a $(\epsilon, 2k)$-RIP linear operator. Let X^* be any matrix (sparse or dense) satisfying*

$$
\begin{aligned}
\mathbf{b} &= B(X^*), \\
X^* &\succeq 0, \\
\mathrm{rank}\{X^*\} &= 1,
\end{aligned}
\tag{22}
$$

let $\tilde{X}$ be the CPRL solution, (10), and form X_s from X^* by setting all but the k largest elements to zero, i.e.,

$$
X_s = \operatorname*{argmin}_{X : \|X\|_0 \le k} \|X^* - X\|_1.
\tag{23}
$$

Then,

$$
\begin{aligned}
\|\tilde{X} - X^*\|_2 \le &\frac{2}{(1-\rho)\sqrt{k}}\|X^* - X_s\|_1 \\
&+ (2(1-\rho)^{-1} + k^{-1/2})\frac{1}{\lambda}(\mathrm{Tr}\,X^* - \mathrm{Tr}\,\tilde{X})
\end{aligned}
\tag{24}
$$

with $\rho = \sqrt{2}\epsilon/(1-\epsilon)$.

Proof of Theorem 8. The proof is inspired by the work on compressed sensing presented in Candès [2008].

First, we introduce $\Delta = \tilde{X} - X^*$. For a matrix X and an index set T, we use the notation X_T to mean the matrix with all zeros except those indexed by T, which are set to the corresponding values of X. Then let T_0 be the index set of the k largest elements of X^* in absolute value, and $T_0^c = \{(1,1), (1,2), \ldots, (n,n)\} \setminus T_0$ be its complement. Let T_1 be the index set associated with the k largest elements in absolute value of $\Delta_{T_0^c}$ and $T_{0,1} \doteq T_0 \cup T_1$ be the union. Let T_2 be the index set associated with the k largest elements in absolute value of $\Delta_{T_{0,1}^c}$, and so on.

14

Notice that

$$\|\Delta\|_2 = \|\Delta_{T_{0,1}} + \Delta_{T_{0,1}^c}\|_2 \leq \|\Delta_{T_{0,1}}\|_2 + \|\Delta_{T_{0,1}^c}\|_2. \qquad (25)$$

We will now study each of the two terms on the right hand side separately.

We first consider $\|\Delta_{T_{0,1}^c}\|_2$. For $j > 1$ we have that for each $i \in T_j$ and $i' \in T_{j-1}$ that $|\Delta[i]| \leq |\Delta[i']|$. Hence $\|\Delta_{T_j}\|_\infty \leq \|\Delta_{T_{j-1}}\|_1/k$. Therefore,

$$\|\Delta_{T_j}\|_2 \leq k^{1/2}\|\Delta_{T_j}\|_\infty \leq k^{-1/2}\|\Delta_{T_{j-1}}\|_1 \qquad (26)$$

and

$$\|\Delta_{T_{0,1}^c}\|_2 \leq \sum_{j\geq 2}\|\Delta_{T_j}\|_2 \leq k^{-1/2}\|\Delta_{T_0^c}\|_1. \qquad (27)$$

Now, since $\tilde{X}$ minimizes $\operatorname{Tr} X + \lambda\|X\|_1$, we have

$$
\begin{aligned}
\operatorname{Tr} X^* + \lambda\|X^*\|_1 \;\geq\;& \operatorname{Tr}\tilde{X} + \lambda\|\tilde{X}\|_1 \\
\geq\;& \operatorname{Tr}\tilde{X} + \lambda(\|X_{T_0}^*\|_1 - \|\Delta_{T_0}\|_1 \\
+\;& \|\Delta_{T_0^c}\|_1 - \|X_{T_0^c}^*\|_1).
\end{aligned} \qquad (28)
$$

Hence,

$$\|\Delta_{T_0^c}\|_1 \leq \frac{-1}{\lambda}\operatorname{Tr}\Delta - \|X_{T_0}^*\|_1 + \|\Delta_{T_0}\|_1 + \|X_{T_0^c}^*\|_1 + \|X^*\|_1. \qquad (29)$$

Using the fact $\|X_{T_0^c}^*\|_1 = \|X^* - X_s\|_1 = \|X^*\|_1 - \|X_{T_0}^*\|_1$, we get a bound for $\|\Delta_{T_0^c}\|_1$:

$$\|\Delta_{T_0^c}\|_1 \leq \frac{-1}{\lambda}\operatorname{Tr}\Delta + \|\Delta_{T_0}\|_1 + 2\|X_{T_0^c}^*\|_1. \qquad (30)$$

Subsequently, the bound for $\|\Delta_{T_{0,1}^c}\|_2$ is given by

$$\|\Delta_{T_{0,1}^c}\|_2 \leq k^{-1/2}\left(\frac{-1}{\lambda}\operatorname{Tr}\Delta + \|\Delta_{T_0}\|_1 + 2\|X_{T_0^c}^*\|_1\right) \qquad (31)$$

$$\leq k^{-1/2}\left(\frac{-1}{\lambda}\operatorname{Tr}\Delta + 2\|X_{T_0^c}^*\|_1\right) + \|\Delta_{T_0}\|_2. \qquad (32)$$

Next, we consider $\|\Delta_{T_{0,1}}\|_2$. It can be shown by a similar derivation as in Candès [2008] that

$$\|\Delta_{T_{0,1}}\|_2 \leq \frac{\rho}{1-\rho}k^{-1/2}\|X^* - X_s\|_1 - \frac{1}{\lambda}\frac{1}{1-\rho}\operatorname{Tr}\Delta. \qquad (33)$$

15

Lastly, combine the bounds for $\|\Delta_{T_{0,1}^c}\|_2$ and $\|\Delta_{T_{0,1}}\|_2$, and we get the final result:

$$\|\Delta\|_2 \leq \|\Delta_{T_{0,1}}\|_2 + \|\Delta_{T_{0,1}^c}\|_2 \tag{34}$$

$$\leq -k^{-1/2}\frac{1}{\lambda}\operatorname{Tr}\Delta + 2k^{-1/2}\|X_{T_0^c}^*\|_1 + 2\|\Delta_{T_{0,1}}\|_2 \tag{35}$$

$$\leq -\left(2\left(1-\rho\right)^{-1} + k^{-1/2}\right)\frac{1}{\lambda}\operatorname{Tr}\Delta \tag{36}$$

$$+ 2(1-\rho)^{-1}k^{-1/2}\|X^* - X_s\|_1. \tag{37}$$

$$\square$$

The bound given in Theorem 8 is rather impractical since it contains both $\|\tilde{X} - X^*\|_2$ and $\operatorname{Tr}(\tilde{X} - X^*)$. The weaker bound given in the following corollary does not have this problem:

Corollary 9 (A Practical Bound on $\|\tilde{X} - X^*\|_2$). *The bound on $\|\tilde{X} - X^*\|_2$ in Theorem 8 can be relaxed to a weaker bound:*

$$\left(1 - \left(\frac{2k^{1/2}}{1-\rho} + 1\right)\frac{1}{\lambda}\right)\|\tilde{X} - X^*\|_2 \tag{38}$$

$$\leq \frac{2}{(1-\rho)\sqrt{k}}\|X^* - X_s\|_1. \tag{39}$$

If X^* is k-sparse, $\epsilon < \frac{1}{1+\sqrt{2}}$, and $B(\cdot)$ is an $(\epsilon, 2k)$-RIP linear operator, then we can guarantee that $\tilde{X} = X^*$ if

$$\lambda > \frac{2k^{1/2}}{1-\rho} + 1 \tag{40}$$

and $\tilde{X}$ has rank 1.

Proof of Corollary 9. It follows from the assumptions of Theorem 8 that

$$1 - \rho = 1 - \frac{\sqrt{2}\epsilon}{1-\epsilon} \geq 1 - \frac{\sqrt{2}\frac{1}{1+\sqrt{2}}}{1 - \frac{1}{1+\sqrt{2}}} = 0. \tag{41}$$

Hence,

$$\left(2\left(1-\rho\right)^{-1} + k^{-1/2}\right)\frac{1}{\lambda} \geq 0. \tag{42}$$

16

Therefore, we have

$$\|\tilde{X} - X^*\|_2 \leq \frac{2}{(1-\rho)\sqrt{k}}\|X^* - X_s\|_1 \tag{43}$$

$$+ \left(2\,(1-\rho)^{-1} + k^{-1/2}\right)\frac{1}{\lambda}(\operatorname{Tr} X^* - \operatorname{Tr} \tilde{X}) \tag{44}$$

$$\leq \frac{2}{(1-\rho)\sqrt{k}}\|X^* - X_s\|_1 \tag{45}$$

$$+ \left(2\,(1-\rho)^{-1} + k^{-1/2}\right)\frac{1}{\lambda}\|X^* - \tilde{X}\|_1 \tag{46}$$

$$\leq \frac{2}{(1-\rho)\sqrt{k}}\|X^* - X_s\|_1 \tag{47}$$

$$+ \left(2\,(1-\rho)^{-1} + k^{-1/2}\right)\frac{k^{1/2}}{\lambda}\|X^* - \tilde{X}\|_2 \tag{48}$$

which is equal to the proposed condition after a rearrangement of the terms. $\qquad\square$

Given the above analysis, however, it may be the case that the linear operator $B(\cdot)$ does not satisfy the RIP property defined in Definition 4 for a sufficiently small ϵ, as pointed out in Candès et al. [2011c]. Therefore, next we turn our attention to RIP-1 linear operators.

6.2   Analysis Using RIP-1

We define RIP-1 as follows:

Definition 10 (RIP-1)**.** *A linear operator $B(\cdot)$ is (ϵ, k)-RIP-1 if for all matrices $X \neq 0$ subject to $\|X\|_0 \leq k$, we have*

$$\left|\frac{\|B(X)\|_1}{\|X\|_1} - 1\right| < \epsilon. \tag{49}$$

Theorems 5–6 and Corollary 7 all hold with RIP replaced by RIP-1. The proofs follow those of the previous section with minor modifications (basically replace the 2-norm with the ℓ_1-norm). The RIP-1 counterparts of Theorems 5–6 and Corollary 7 are not restated in details here. Instead we summarize the most important property in the following theorem:

Theorem 11 (Upper Bound & Recoverability Through ℓ_1)**.** *Let $\bar{x}$ be the sparsest solution to (4). The solution of CPRL (10), $\tilde{X}$, is equal to $\bar{x}\bar{x}^H$ if it has rank 1 and $B(\cdot)$ is $(\epsilon, 2\|\tilde{X}\|_0)$-RIP-1 with $\epsilon < 1$.*

17

Proof of Theorem 11. The proof follows trivially for the proof Theorem 7 (basically replace the 2-norm with the ℓ_1-norm).

<div style="text-align: right">□</div>

6.3 Analysis Using Mutual Coherence

The RIP type of argument may be difficult to check for a given matrix and are more useful for claiming results for classes of matrices/linear operators. For instance, it has been shown that random Gaussian matrices satisfy the RIP with high probability. However, given one sample of a random Gaussian matrix, it is hard to check if it actually satisfies the RIP or not.

Two alternative arguments are spark Chen et al. [1998] and mutual coherence Donoho and Elad [2003], Candès et al. [2011b]. The spark condition usually gives tighter bounds but is known to be difficult to compute. On the other hand, mutual coherence may give less tight bounds, but is more tractable. We will focus on mutual coherence here.

Mutual coherence is defined as:

Definition 12 (Mutual Coherence). *For a matrix A, define the mutual coherence as*

$$\mu(A) = \max_{1 \leq k,j \leq n, k \neq j} \frac{|\boldsymbol{a}_k^H \boldsymbol{a}_j|}{\|\boldsymbol{a}_k\|_2 \|\boldsymbol{a}_j\|_2}. \tag{50}$$

By an abuse of notation, let B be the matrix satisfying $\boldsymbol{b} = BX^s$ with X^s being the vectorized version of X. We are now ready to state the following theorem:

Theorem 13 (Recovery Using Mutual Coherence). *Let $\bar{\boldsymbol{x}}$ be the sparsest solution to (4). The solution of CPRL (10), $\tilde{X}$, is equal to $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$ if it has rank 1 and $\|\tilde{X}\|_0 < 0.5(1 + 1/\mu(B))$.*

Proof of Theorem 13. Since $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$ satisfies $\boldsymbol{b} = B(\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H)$, it follows from [Donoho and Elad, 2003, Thm. 1] that

$$\|\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(B)}\right) \tag{51}$$

is a sufficient condition for $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$ to be a unique solution. It further follows that if $\tilde{X}$ also satisfies (51) then we must have that $\tilde{X} = \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H$ since $\tilde{X}$ also satisfies $\boldsymbol{b} = B(\tilde{X})$.

<div style="text-align: right">□</div>

18

7   Experiment

This section gives a number of comparisons with the other state-of-the-art methods in compressive phase retrieval. Code for the numerical illustrations will be made available.

7.1   Simulation

First, we repeat the simulation given in Example 1 for $k = 1, \ldots, 5$. For each k, $n = 64$ is fixed, and we increase the measurement dimension N until CPRL recovered the true sparse support in at least 95 out of 100 trials, i.e., 95% success rate. New data ($\boldsymbol{x}$, $\boldsymbol{b}$, and R) are generated in each trial. The curve of 95% success rate is shown in Figure 4.
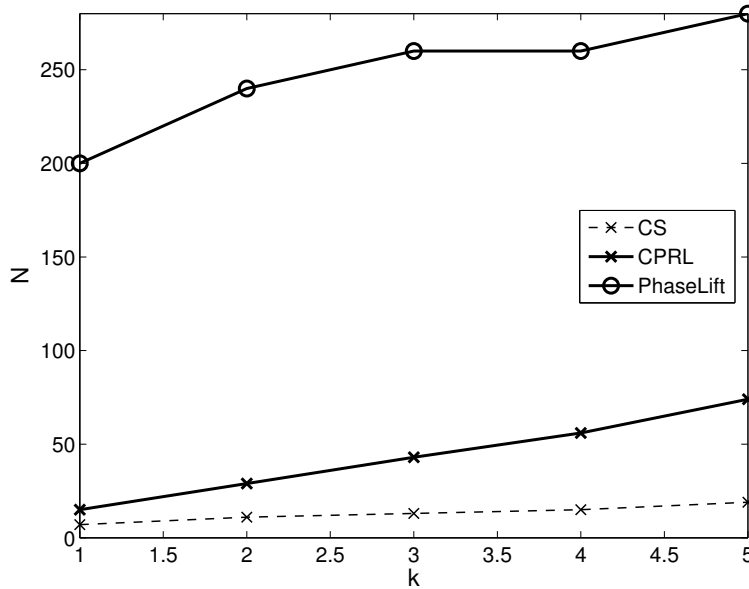
Figure 4: The curves of 95% success rates for CPRL, PhaseLift, and CS. Note that in the CS scenario, the simulation is given the complete output $\boldsymbol{y}$ instead of its squared magnitudes.

With the same simulation setup, we compare the accuracy of CPRL with the PhaseLift approach and the CS approach in Figure 4. First, note that CS is not applicable to phase retrieval problems in practice, since it assumes the phase of the observation is also given. Nevertheless, the simulation shows

19

CPRL via the SDP solution only requires a slightly higher sampling rate to achieve the same success rate as CS, even when the phase of the output is missing. Second, similar to the discussion in Example 1, without enforcing the sparsity constraint in (11), PhaseLift would fail to recover correct sparse signals in the low sampling rate regime.

It is also interesting to see the performance as n and N vary and k held fixed. We therefore use the same setup as in Figure 4 but now fixed $k = 2$ and for $n = 10, \ldots, 60$, gradually increased N until CPRL recovered the true sparsity pattern with 95% success rate. The same procedure is repeated to evaluate PhaseLift and CS. The results are shown in Figure 5.
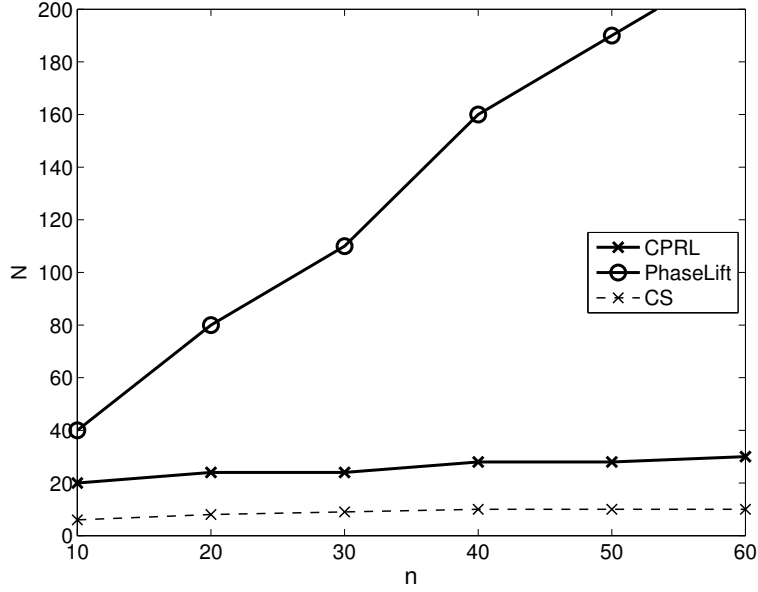
Figure 5: The curves of 95% success rate for CPRL, PhaseLift, and CS. Note that the CS simulation is given the complete output $\boldsymbol{y}$ instead of its squared magnitudes.

Compared to Figure 4, we can see that the degradation from CS to CPRL when the phase information is omitted is largely affected by the sparsity of the signal. More specifically, when the sparsity k is fixed, even when the dimension n of the signal increases dramatically, the number of squared observations to achieve accurate recovery does not increase significantly for both CS and CPRL.

20

Next, we calculate the quantity $\frac{1}{2}\left(1 + \frac{1}{\mu(B)}\right)$, as Theorem 13 shows that when

$$\|\tilde{X}\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(B)}\right) \tag{52}$$

and $\tilde{X}$ has rank 1, then $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H = \tilde{X}$. The quantity is plotted for a number of different N and n's in Figure 6. From the plot it can be concluded that if the solution $\tilde{X}$ has rank 1 and only a single nonzero component for a choice of $125 \geq n$, $N \geq 5$, Theorem 13 can guarantee that $\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H = \tilde{X}$. We also observer that Theorem 13 is pretty conservative, since from Figure 5 we have that with high probability we needed $N > 25$ to guarantee that a two sparse vector is recovered correctly for $n = 20$.
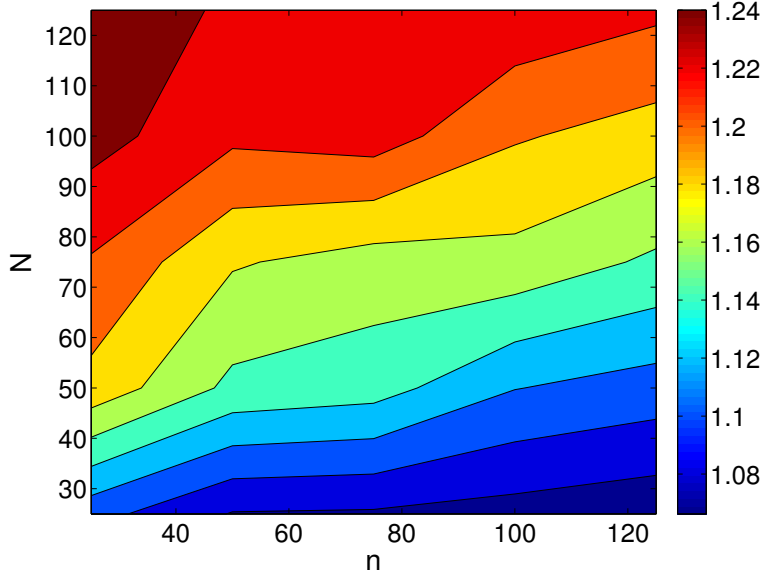
Figure 6: A contour plot of the quantity $\frac{1}{2}\left(1 + \frac{1}{\mu(B)}\right)$. μ is taken as the average over 10 realizations of B.

7.2 Audio Signals

In this section, we further demonstrate the performance of CPRL using signals from a real-world audio recording. The timbre of a particular note on an instrument is determined by the fundamental frequency, and several overtones. In a Fourier basis, such a signal is sparse, being the summation of

21

a few sine waves. Using the recording of a single note on an instrument will give us a naturally sparse signal, as opposed to synthesized sparse signals in the previous sections. Also, this experiment will let us analyze how robust our algorithm is in practical situations, where effects like room ambience might color our otherwise exactly sparse signal with noise.

Our recording $\boldsymbol{z} \in \mathbb{R}^s$ is a real signal, which is assumed to be sparse in a Fourier basis. That is, for some sparse $\boldsymbol{x} \in \mathbb{C}^n$, we have $\boldsymbol{z} = F_{inv}\boldsymbol{x}$, where $F_{inv} \in \mathbb{C}^{s \times n}$ is a matrix representing a transform from Fourier coefficients into the time domain. Then, we have a randomly generated mixing matrix with normalized rows, $R \in \mathbb{R}^{N \times s}$, with which our measurements are sampled in the time domain:

$$\boldsymbol{y} = R\boldsymbol{z} = RF_{inv}\boldsymbol{x}. \tag{53}$$

Finally, we are only given the magnitudes of our measurements, such that $\boldsymbol{b} = |\boldsymbol{y}|^2 = |R\boldsymbol{z}|^2$.

For our experiment, we choose a signal with $s = 32$ samples, $N = 30$ measurements, and it is represented with $n = 2s$ (overcomplete) Fourier coefficients. Also, to generate F_{inv}, the $\mathbb{C}^{n \times n}$ matrix representing the Fourier transform is generated, and s rows from this matrix are randomly chosen.

The experiment uses part of an audio file recording the sound of a tenor saxophone. The signal is cropped so that the signal only consists of a single sustained note, without silence. Using CPRL to recover the original audio signal given $\boldsymbol{b}$, R, and F_{inv}, the algorithm gives us a sparse estimate $\boldsymbol{x}$, which allows us to calculate $\boldsymbol{z}_{est} = F_{inv}\boldsymbol{x}$. We observe that all the elements of $\boldsymbol{z}_{est}$ have phases that are π apart, allowing for one global rotation to make $\boldsymbol{z}_{est}$ purely real. This matches our previous statements that CPRL will allow us to retrieve the signal up to a global phase.

We also find that the algorithm is able to achieve results that capture the trend of the signal using less than s measurements. In order to fully exploit the benefits of CPRL that allow us to achieve more precise estimates with smaller errors using fewer measurements relative to s, the problem should be formulated in a much higher ambient dimension. However, using the CVX Matlab toolbox by Grant and Boyd [2010], we already ran into computational and memory limitations with the current implementation of the CPRL algorithm. These results highlight the need for a more efficient numerical implementation of CPRL as an SDP problem.

22
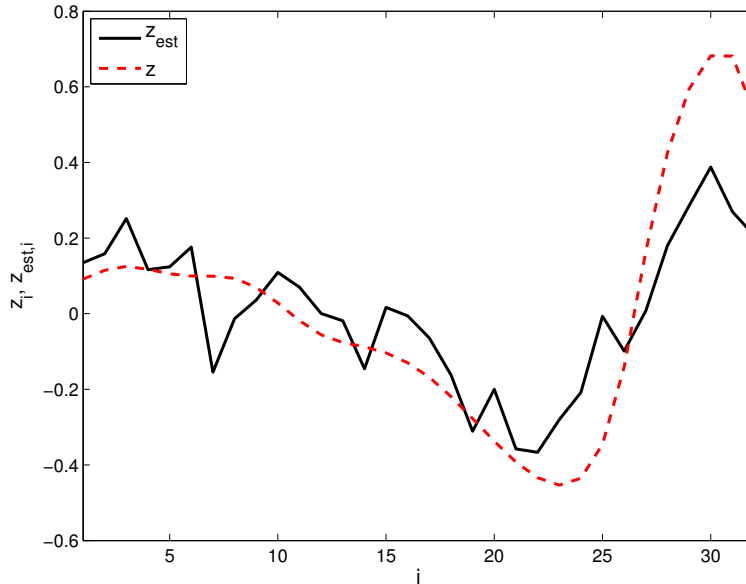
Figure 7: The retrieved signal z_{est} using CPRL versus the original audio signal z.

8 Conclusion and Discussion

A novel method for the compressive phase retrieval problem has been presented. The method takes the form of an SDP problem and provides the means to use compressive sensing in applications where only squared magnitude measurements are available. The convex formulation gives it an edge over previous presented approaches and numerical illustrations show state of the art performance.

One of the future directions is improving the speed of the standard SDP solver, i.e., interior-point methods, currently used for the CPRL algorithm. The authors have previously introduced efficient numerical acceleration techniques for ℓ_1-min and Sparse PCA problems. We believe similar techniques also apply to CPR. Such accelerated CPR solvers would facilitate exploring a broad range of high-dimensional CPR applications in optics, medical imaging, and computer vision, just to name a few.

23
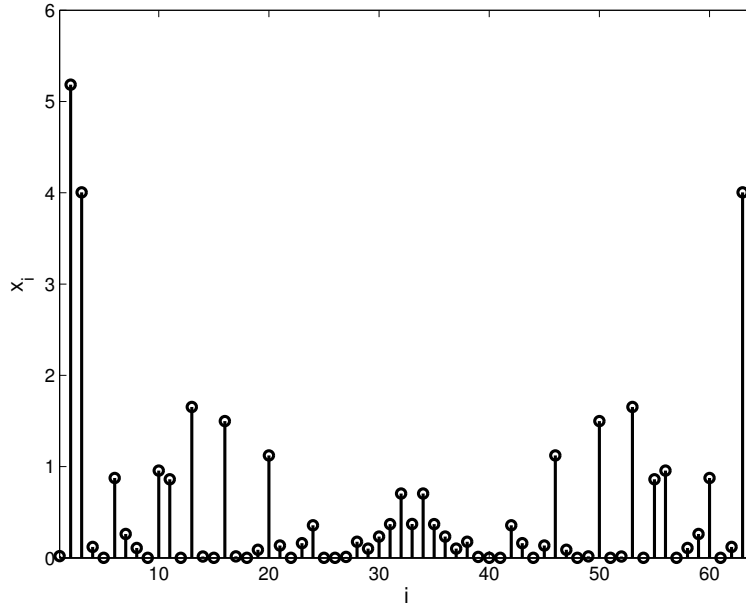
Figure 8: The magnitude of $\boldsymbol{x}$ retrieved using CPRL. The audio signal $\boldsymbol{z}_{est}$ is obtained by $\boldsymbol{z}_{est} = F_{inv}\boldsymbol{x}$.

References

R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20:345–356, 2006.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1): 183–202, 2009.

R. Berinde, A.C. Gilbert, P. Indyk, H. Karloff, and M.J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 798–805, September 2008.

A. Bruckstein, D. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51 (1):34–81, 2009.

E. Candès. The restricted isometry property and its implications for com-

24

pressed sensing. *Comptes Rendus Mathematique*, 346(9–10):589–592, 2008.

E. Candès and M. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, March 2008.

E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52:489–509, February 2006.

E. Candès, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. Technical Report arXiv:1109.0573, Stanford University, September 2011a.

E. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? *Journal of the ACM*, 58(3), 2011b.

E. Candès, T. Strohmer, and V. Voroninski. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. Technical Report arXiv:1109.4499, Stanford University, September 2011c.

A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. Technical report, Stanford University, 2010.

S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

J. Dainty and J. Fienup. Phase retrieval and image reconstruction for astronomy. In *Image Recovery: Theory and Applications*. Academic Press, New York, 1987.

A. d'Aspremont, L. El Ghaoui, M. Jordan, and G. Lanckriet. A direct formulation for Sparse PCA using semidefinite programming. *SIAM Review*, 49(3):434–448, 2007.

D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l-minimization. *PNAS*, 100(5):2197–2202, March 2003.

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

25

J. Fienup. Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. *Journal of Optical Society of America A*, 4(1):118–123, 1987.

R. Gerchberg and W. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

R. Gonsalves. Phase retrieval from modulus data. *Journal of Optical Society of America*, 66(9):961–964, 1976.

M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. `http://cvxr.com/cvx`, August 2010.

A. Hoerl and R. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

D. Kohler and L. Mandel. Source reconstruction from the modulus of the correlation function: a practical approach to the phase problem of optical coherence theory. *Journal of the Optical Society of America*, 63(2):126–134, 1973.

I. Loris. On the performance of algorithms for the minimization of ℓ_1-penalized functionals. *Inverse Problems*, 25:1–16, 2009.

S. Marchesini. Phase retrieval and saddle-point optimization. *Journal of the Optical Society of America A*, 24(10):3289–3296, 2007.

R. Millane. Phase retrieval in crystallography and optics. *Journal of the Optical Society of America A*, 7:394–411, 1990.

M. Moravec, J. Romberg, and R. Baraniuk. Compressive phase retrieval. In *SPIE International Symposium on Optical Science and Technology*, 2007.

B. Recht, M. Fazel, and P Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52 (3):471–501, 2010.

R. Tibsharani. Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society B (Methodological)*, 58(1):267–288, 1996.

J. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, October 2004.

26

A. Walther. The question of phase retrieval in optics. *Optica Acta*, 10:41–49, 1963.

A. Yang, A. Ganesh, Y. Ma, and S. Sastry. Fast ℓ_1-minimization algorithms and an application in robust face recognition: A review. In *ICIP*, 2010.

27