PHASE RETRIEVAL FOR SPARSE SIGNALS USING RANK MINIMIZATION

Kishore Jaganathan, Samet Oymak and Babak Hassibi

Department of Electrical Engineering, Caltech, Pasadena - 91125

ABSTRACT

Signal recovery from the amplitudes of the Fourier transform, or equivalently from the autocorrelation function is a classical problem. Due to the absence of phase information, signal recovery requires some form of additional prior information. In this paper, the prior information we assume is sparsity. We develop a convex optimization based framework to retrieve the signal support from the support of the autocorrelation, and propose an iterative algorithm which terminates in a signal with the least sparsity satisfying the autocorrelation constraints. Numerical results suggest that unique recovery up to a global sign change, time shift and/or time reversal is possible with a very high probability for sufficiently sparse signals.

Index Terms— Phase retrieval, sparse signals, rank minimization, convex optimization

1. INTRODUCTION

The problem of loss of phase information arises in many areas of engineering and applied physics, including x-ray crystallography, astronomical imaging, electron microscopy, channel estimation and particle scattering during the measuring process. Retrieving phase information from the measured magnitude data is known as phase retrieval. Over the last few decades, this problem has generated a lot of interest and a wide range of algorithms have been proposed (e.g., [1, 2]). A comprehensive survey can be found in [3, 4].

Phase retrieval is equivalent to recovering the signal from its autocorrelation. For the case of a one dimensional signal, the mapping from the signal to the autocorrelation is not one-one, and hence unique recovery is not possible in general. For a given Fourier transform magnitude data, every possible phase corresponds to a different signal. To overcome this, the signal is generally forced to satisfy certain constraints to limit the number of possible signals. A convex optimization based approach has been proposed recently in [5], where multiple measurements are used in an attempt to resolve the phase ambiguity.

This work was supported in part by the National Science Foundation under grants CCF-0729203, CNS-0932428 and CCF-1018927, by the Office of Naval Research under the MURI grant N00014-08-1-0747, and by Caltech's Lee Center for Advanced Networking.

Considerable research has been done towards extracting information about the support of the signal from the support of the autocorrelation [6, 7]. The support information can be instrumental in recovering the signal from its autocorrelation by reducing the complexity or providing useful constraints.

In this work, we assume that the signal is real and sparse, i.e., the number of locations where the signal is non-zero is much less compared to the length of the signal. In many applications mentioned above, sparsity is a reasonable assumption. For example, in astronomical imaging, the number of locations where the signal is non-zero is much less compared to the size. Sparsity greatly restricts the number of valid phase combinations, and research has been done to exploit this feature recently. An iterative algorithm has been proposed in [8] which guarantees sparse spectral factorization under certain sufficient conditions. We see in [8] that for uniform sparsity patterns, one can always find multiple signals satisfying the autocorrelation constraints, no matter how sparse they are.

In this paper, we propose an iterative algorithm which recovers a sparse signal up to a sign change, time reversal and time shift from the autocorrelation of the signal. Since sign change, time reversal and time shifts preserve sparsity and do not change the autocorrelation, we cannot differentiate between them. We formulate the phase retrieval problem as a constrained rank minimization problem (which in general is NP-hard). Various heuristics have been proposed in recent years (e.g., [9, 10]) to solve rank minimization problems. Here, we propose a novel algorithm for rank minimization under the constraints required of phase retrieval problem.

The algorithm we propose for sparse signal recovery is shown to converge to the sparsest solution satisfying the autocorrelation constraints. Simulations suggest that the recovery is unique up to a certain sparsity with a very high probability if the support is chosen randomly. Also, simple modifications can be made to account for measurement noise.

The paper is organized as follows. Section 2 of the paper describes the problem formulation. An iterative algorithm is proposed in Section 3 to solve the phase retrieval problem for sparse signals. In Section 4, we present the results of numerical simulations.

2. PROBLEM FORMULATION

Let $\mathbf{x} = (x_0, x_1, ...x_{n-1})^T$ be a real-valued signal of length n. We will construct a new signal $\hat{\mathbf{x}}$ of length $m = 2n$ from $\mathbf{x}$ by appending n zeros in the end so that circular indexing can be used for autocorrelation. Let $\mathbf{a} = (a_0, a_1, ...a_{m-1})^T$ be the autocorrelation of $\hat{\mathbf{x}}$, defined as

$$a_k = \sum_{i=0}^{m-1} \hat{x}_i \hat{x}_{i+k} \qquad 0 \le k \le m-1 \tag{1}$$

Let $\mathbf{y} = (y_0, y_1, ...y_{m-1})^T$ be the Fourier transform of $\hat{\mathbf{x}}$, i.e.,

$$\mathbf{y} = \mathbf{F}\hat{\mathbf{x}} \tag{2}$$

where $\mathbf{F}$ is the $m \times m$ DFT matrix. We have access to $\mathbf{Y} = (|y_0|^2, |y_1|^2, ...|y_{m-1}|^2)^T$, from which $\mathbf{a}$ can be obtained by taking the inverse Fourier transform.

Phase retrieval problem can be formulated as finding an $\hat{\mathbf{x}}$ which satisfies (1). This is clearly not a convex optimization problem as the autocorrelation constraints are non-convex. We can relax the constraints into a set of convex constraints by embedding the problem in a higher dimensional space by trying to find $\mathbf{X} = \hat{\mathbf{x}}\hat{\mathbf{x}}^T$. In particular, the problem is equivalent to

$$\begin{aligned} &\text{find} && \mathbf{X} = \hat{\mathbf{x}}\hat{\mathbf{x}}^T \\ &\text{subject to} && |y_k|^2 = trace(\mathbf{M_k X}) \qquad k = 0, ..., m-1 \end{aligned}$$

where $\mathbf{M_k}$ is the $m \times m$ matrix defined by $[\mathbf{M_k}]_{u,v} = \exp(-\frac{i2\pi k(u-v)}{m}), \forall\, u, v \in \{0, 1, ...m-1\}$.

We need to find a rank one solution $\mathbf{X}$ in the given convex set. This problem can be formulated as

$$\begin{aligned} &\text{minimize} && rank(\mathbf{X}) && (3) \\ &\text{subject to} && \mathbf{Y_k} = trace(\mathbf{M_k X}), \quad k = 0, ...m-1 && (4) \\ & && \mathbf{X} \succcurlyeq 0 \end{aligned}$$

3. RECONSTRUCTION ALGORITHM

3.1. Rank Minimization

Rank Minimization Problem (RMP) is very well studied by a lot of researchers ([9, 10, 12, 13]). In general, the RMP can be expressed as

$$\text{minimize} \quad rank(\mathbf{X}) \quad \text{subject to} \quad \mathbf{X} \in \mathcal{C}$$

where $\mathbf{X}$ is the optimization variable and $\mathcal{C}$ is a convex set. This is a non-convex problem as rank is a non-convex function. It has been shown in [9] that trace minimization is the best convex relaxation of rank minimization for positive semidefinite matrices. This relaxation is not very useful in the phase retrieval setup (3) though, as $trace(\mathbf{X})$ corresponds to the energy of the signal $\mathbf{x}$, which is of a fixed value a_0. A heuristic (5) proposed in [10] can be used instead to solve for low rank solutions.

$$\text{minimize} \quad \log\det(\mathbf{X} + \epsilon\mathbf{I}) \quad \text{subject to} \quad \mathbf{X} \in \mathcal{C} \tag{5}$$

This heuristic tries to minimize a concave function in a convex domain, which can be done using gradient descent approach. Since the gradient involves $\mathbf{X}^{-1}$, a matrix inversion has to be done at every step of the iteration and hence this method is computationally expensive. In this work, we introduce a new and specialized heuristic (6) for the phase retrieval setup. Observe that in (3), $\mathcal{C}$ is a subset of the set of positive semidefinite matrices with constant trace a_0.

$$\text{maximize} \quad ||\mathbf{X}||_F^2 \quad \text{subject to} \quad \mathbf{X} \in \mathcal{C} \tag{6}$$

It should be emphasized that, similar to (5), program (6) is non-convex as we maximize a convex function ($\|\cdot\|_F$ norm) in a convex domain. Gradient ascent approach, similar to the one used in [10] can be used. The first order Taylor expansion of $||\mathbf{X}||_F^2$ about $\mathbf{X}_k$ is

$$||\mathbf{X}||_F^2 = ||\mathbf{X}_k||_F^2 + 2\, trace(\mathbf{X}_k^T(\mathbf{X} - \mathbf{X}_k)) \tag{7}$$

For gradient ascent, we replace the convex function with a linear function having the same gradient as the convex function at the solution of the previous iteration. The problem is now reduced to

$$\text{maximize} \quad trace(\mathbf{X_k^T X}) \quad \text{subject to} \quad \mathbf{X} \in \mathcal{C} \tag{8}$$

where $\mathbf{X}_0$ is chosen randomly with all terms in it being non-negative and $\mathbf{X}_k$ for $k > 0$ can be chosen using

$$\mathbf{X}_k = \arg\max_{\mathbf{X} \in \mathcal{C}} \quad trace(\mathbf{X_{k-1}^T X}) \tag{9}$$

This is much faster than the log-determinant heuristic used in [10], which requires a matrix inversion to obtain the gradient. The algorithm will surely converge to a maximum (local or global) as the function is convex, because of which in each iteration its value will increase by an amount at least as much as the linearized objective function.

Lemma 3.1.1. *Suppose $\mathcal{C}$ is a convex set of positive semidefinite matrices with a fixed trace. If there exists a rank 1 matrix $X_\star \in \mathcal{C}$, then any matrix $X \in \mathcal{C}$ is a global optimum of (6) if and only if it is rank 1.*

Proof. For a positive semidefinite matrix with eigenvalues $\{\lambda_i\}$, we have $\lambda_i \ge 0,\ \forall\, 1 \le i \le m$ and

$$trace(\mathbf{X}) = \sum_{i=1}^{m} \lambda_i = c \quad \text{and} \quad ||\mathbf{X}||_F^2 = \sum_{i=1}^{m} \lambda_i^2$$

Then, $\|\mathbf{X}\|_F^2 = \sum_i \lambda_i^2 \leq (\sum_i \lambda_i)^2 = c^2$ and equality is achieved iff all cross terms $\lambda_i \lambda_j$, $i \neq j$, are 0 i.e. exactly one eigenvalue is nonzero or equivalently the rank of the matrix is 1 (assuming $\mathbf{X} \neq 0$). Finally, by assumption, there is a rank 1 solution. Hence c^2 can be achieved in (6). As a result, all solutions achieving the same objective have to be rank 1. □

Once the iterative algorithm for (6) converges, we can check the rank of the solution, which will tell us if we have converged to a local or a global maximum. The process can be iterated from different starting points $\mathbf{X}_0$ till we find a global maximum, which by Lemma 3.1.1 will be a rank 1 matrix. The signal $\mathbf{x}$ can be extracted from it using a simple decomposition.

3.2. Support Estimation

In this work, we aim to find the sparsest solution satisfying the autocorrelation constraints, which is an NP-hard problem. Although, ℓ_1 minimization is a popular and powerful method for sparse recovery [11], due to the special structure of measurements $\{\mathbf{M}_k\}_{i=0}^{m-1}$ (which are induced by the DFT matrix), it is not hard to construct examples for which ℓ_1 minimization fails.

For instance, let $\{\hat{\mathbf{x}}_i\}_{i=1}^{i=2m-1}$ be the set of signals constructed from $\hat{\mathbf{x}}$ by using circular time-shifting and time-reversing operators. Since these operations do not change the autocorrelation, any matrix $\mathbf{X}_i = \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^T$ will satisfy the constraints (4). Note that the constraints (4) are linear and hence any convex combination of such matrices will satisfy the constraints. As a result, although it won't be discussed here, one can construct trivial examples where there exists a convex combination with rank more than 1 and an ℓ_1 norm lesser than that of $\mathbf{X}$.

To enforce sparsity, we shall attempt to find the support of the signal. We will assume that the support of the signal is a subset of the support of the autocorrelation. This is same as assuming there is no cancellation of support in the autocorrelation, which is a very weak requirement and holds with probability one if the coefficients of the signal are chosen randomly from a non-degenerate distribution. We can formulate the support estimation problem as an RMP. Suppose $\mathbf{S}$ and $\mathbf{A}$ are $m \times m$ matrices defined as

$$S_{ij} = \begin{cases} 1 & \text{if } \hat{x}_i \,\&\, \hat{x}_j \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad A_{ij} = \begin{cases} 1 & \text{if } a_i \,\&\, a_j \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

for $0 \leq i, j \leq m-1$. If $\mathbf{u}$ is the indicator function of $\hat{\mathbf{x}}$, i.e., $u_i = 1$ if $\hat{x}_i \neq 0$ and $u_i = 0$ otherwise, then $\mathbf{S} = \mathbf{u}\mathbf{u}^T$. If s is the sparsity of the signal, the support estimation problem becomes

$$\begin{aligned} \text{minimize} \quad & rank(\mathbf{S}) \\ \text{subject to} \quad & trace(\mathbf{S}) = s \\ & \sum_i \sum_j S_{ij} = s^2 \\ & \sum_i S_{i,i+k} > 0 \text{ iff } a_k \neq 0 \\ & 0 \leq S_{ij} \leq 1 \quad 0 \leq i, j \leq m-1 \\ & \mathbf{S} \in \mathbf{A}, \mathbf{S} \succcurlyeq 0 \end{aligned} \tag{10}$$

Trace, sum, and lower-upper bound constraints on $\mathbf{S}$ fix sparsity as seen in Lemma 3.2. $\mathbf{S}$ is a subset of $\mathbf{A}$ and is positive semi definite by construction. We also note that $\sum_i S_{i,i+k}$ is zero if a_k is zero as we assume that there is no cancellation, and when a_k is non-zero, there has to be at least one non-zero entry in $\sum_i S_{i,i+k}$ so that the signal can contribute to a_k.

Lemma 3.2.1. *Any rank 1 solution $S_\star$ of the optimization problem corresponds to a valid support of a signal with sparsity s.*

Proof. Let $\mathbf{S}_\star = \mathbf{u}\mathbf{u}^T$. The constraints ensure that

$$\sum_i u_i^2 = s \quad \text{and} \quad \sum_i \sum_j u_i u_j = s^2 \Leftrightarrow \sum_i u_i = s$$

Since $0 \leq u_i \leq 1$, we have $u_i^2 \leq u_i$. If $0 < u_i < 1$ for some i, then $u_i^2 < u_i$ giving us $\sum_i u_i^2 < \sum_i u_i$ which is a contradiction. Hence for the two equations to hold, exactly s of the entries in $\mathbf{u}$ will be 1 and the remaining 0, which will hence give us a valid support of a signal whose autocorrelation will have the same support. □

Note that we assume knowledge of the sparsity of the signal. This condition can be easily relaxed by starting with s as 0 and increasing by 1 till one obtains a feasible solution.

Remark 3.2.1. *Numerical simulations suggest that there are cases where two distinct and sparse supports have the same autocorrelation support. However, based on simulations with randomly generated supports, we can say that for sufficiently sparse signals, the probability of such an event is very low.*

3.3. Sparse Phase Retrieval Algorithm

The support estimation algorithm converges to any possible support whose autocorrelation is given by the autocorrelation of the signal of interest. It is possible that there might not exist a signal on the estimated support which gives the measured autocorrelation, if the estimated support is not the same as the support of the signal that generated the autocorrelation. To overcome this, we combine the rank minimization problem of support and signal estimation. It can be seen that any solution

Algorithm 1 Phase Retrieval Algorithm

Input: The autocorrelation $\mathbf{a}$ of the zero-padded signal.
Output: The sparse signal $\hat{\mathbf{x}}$ which has the autocorrelation $\mathbf{a}$
repeat

- Initialize $k \leftarrow 0$
  repeat
- Start with random matrices $\mathbf{W}_k$ and $\mathbf{V}_k$ by picking their entries uniformly between [0,1]
- Solve the optimization problem

$$\begin{aligned} \text{maximize} \quad & trace(\mathbf{W_k'X} + \mathbf{V_k'S}) \\ \text{subject to} \quad & \mathbf{Y_k} = trace(\mathbf{M_k X}) \qquad 0 \le k \le m-1 \\ & trace(\mathbf{S}) = s \\ & \sum_i \sum_j S_{ij} = s^2 \\ & \sum_i S_{i,i+k} > 0 \text{ iff } a_k > 0 \\ & 0 \le S_{ij} \le 1 \quad 0 \le i,j \le m-1 \\ & \mathbf{X} \in \mathbf{S}, \mathbf{S} \in \mathbf{A} \\ & \mathbf{X}, \mathbf{S} \succcurlyeq 0 \end{aligned}$$

- $k \leftarrow k+1$
- $\mathbf{W}_k = \mathbf{X}_\star$, $\mathbf{V}_k = \mathbf{S}_\star$, where $\mathbf{X}_\star$ and $\mathbf{S}_\star$ is the solution to the optimization problem
  until $||\mathbf{W}_{k+1} - \mathbf{W}_k||_F^2 < \epsilon$ and $||\mathbf{V}_{k+1} - \mathbf{V}_k||_F^2 < \epsilon$

until $rank(\mathbf{X}_\star) = 1$ and $rank(\mathbf{S}_\star) = 1$ or a maximum number of iterations

to the combined optimization problem is a global maximum if and only if it is rank 1 for both the problems, hence giving us the sparsest solution which satisfies the autocorrelation constraints. Algorithm 1 summarizes the steps of the algorithm.
Noisy case: In practice, it is not possible to measure autocorrelation without noise. The algorithm can be made robust to noise by replacing $\mathbf{Y_k} = trace(\mathbf{M_k X})$ with $||\mathbf{Y_k} - trace(\mathbf{M_k X})||_2 < \delta$ for all $0 \le k \le m-1$, where δ is a function of the noise variance.

4. SIMULATION RESULTS

Simulations were performed for various sparsities and for signal length $n = 16$. For each sparsity, the support was chosen uniformly at random. The signal values in the support were drawn from an i.i.d Gaussian distribution. Figure 1 shows the success percentage of signal recovery as a function of sparsity, hence illustrates the performance of Algorithm 1. We see that when the support is randomly chosen, the recovery is unique with high probability for sufficiently sparse signals.

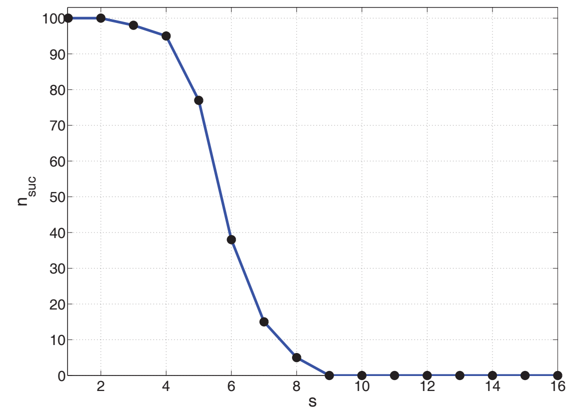

Fig. 1. Percentage of successes n_{suc} vs the signal sparsity s for $n = 16$.

5. REFERENCES

[1] R. W. Gerchberg and W. O. Saxton. "A practical algorithm for the determination of the phase from image and diffraction plane pictures". Optik 35, 237 (1972).

[2] R. A. Gonsalves. "Phase retrieval from modulus data". J. Opt. Soc. Am. 66, 961–964 (1976)

[3] J. R. Fienup, "Phase retrieval algorithms: a comparison". Appl. Opt. 21, 2758–2769 (1982).

[4] A.H. Sayed and T. Kailath. "A survey of spectral factorization methods". Numerical Linear Algebra with Applications, 8: 467–496 (2001).

[5] E. J. Candes, Y. Eldar, T. Strohmer, and V. Voroninski. "Phase retrieval via matrix completion". arXiv:1109.0573v2

[6] J. R. Fienup, T. R. Crimmins, and W. Holsztynski. "Reconstruction of the support of an object from the support of its autocorrelation", J. Opt. Soc. Am. 72, 610–624 (1982)

[7] B. J. Brames. "Efficient method of support reduction". Optics Communications, Volume 64, Issue 4, p. 333–337

[8] Y.M. Lu and M. Vetterli. "Sparse spectral factorization: Unicity and reconstruction algorithms". Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on , vol., no., pp. 5976–5979, 22-27 May 2011

[9] B. Recht, M. Fazel, and P. Parrilo. "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization". SIAM Review, Vol 52, no 3, pages 471–501, 2010

[10] M. Fazel, H. Hindi, and S. Boyd. "Log-det Heuristic for Matrix Rank Minimization with Applications to Hankel and Euclidean Distance Matrices". Proc. American Control Conference, Denver, Colorado, June 2003.

[11] E. J. Candes and T. Tao. "Decoding by linear programming". IEEE Trans. Inform. Theory, 51 4203–4215.

[12] E. J. Candes and B. Recht. "Exact matrix completion via convex optimization". Found. of Comput. Math., 9 717–772.

[13] E. J. Candes and T. Tao. "The power of convex relaxation: Near-optimal matrix completion". IEEE Trans. Inform. Theory, 56(5), 2053–2080.