Sparse Phase Retrieval Using Partial Nested Fourier Samplers

Heng Qiao  Piya Pal

University of Maryland
qiaoheng@umd.edu

November 14, 2015
Overview

1 Problem Formulations and Related Work
   - Problem Formulations
   - Related Work
Overview

1. Problem Formulations and Related Work
   - Problem Formulations
   - Related Work

2. PNFS and Iterative Recovery Algorithm
   - Partial Nested Fourier Sampler (PNFS)
Overview

1. Problem Formulations and Related Work
   - Problem Formulations
   - Related Work

2. PNFS and Iterative Recovery Algorithm
   - Partial Nested Fourier Sampler (PNFS)

3. Sparse Phase Retrieval with Prior Knowledge
   - Cancellation based Approach
   - Number of Measurements via Convex Program
Overview

1. Problem Formulations and Related Work
   - Problem Formulations
   - Related Work

2. PNFS and Iterative Recovery Algorithm
   - Partial Nested Fourier Sampler (PNFS)

3. Sparse Phase Retrieval with Prior Knowledge
   - Cancellation based Approach
   - Number of Measurements via Convex Program

4. Numerical Results
Phase Retrieval: Recovery of a signal given the magnitude of its measurements.

Applications:
- X-ray crystallography: recover Bragg peaks from missing-phase data
- Diffraction imaging, optics, astronomical imaging, microscopy
- Acoustics, blind channel estimation, interferometry, quantum information
Consider data \( \mathbf{x} \in \mathbb{C}^N \), sampler set \( \mathcal{F} = \{ \mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_M \} \) and measurements \( \mathbf{y} = [y_1, y_2, \cdots, y_M]^T \in \mathbb{R}_+^M \)

\[
y_i = |\langle \mathbf{x}, \mathbf{f}_i \rangle| \]
Consider data $\mathbf{x} \in \mathbb{C}^N$, sampler set $\mathcal{F} = \{f_1, f_2, \cdots, f_M\}$ and measurements $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T \in \mathbb{R}^M$. 

$$y_i = |\langle \mathbf{x}, f_i \rangle|$$

$$\Leftrightarrow y_i^2 = f_i^H \mathbf{x} \mathbf{x}^H f_i$$
Mathematical Formulation

Consider data $\mathbf{x} \in \mathbb{C}^N$, sampler set $\mathcal{F} = \{f_1, f_2, \cdots, f_M\}$ and measurements $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T \in \mathbb{R}_+^M$

\[ y_i = |\langle \mathbf{x}, f_i \rangle| \]

\[ \Leftrightarrow y_i^2 = f_i^H \mathbf{x} \mathbf{x}^H f_i \]

\[ \Leftrightarrow y_i^2 = \left( f_i^T \otimes f_i^H \right) \text{Vec} \left( \mathbf{x} \mathbf{x}^H \right) \]
Consider data $\mathbf{x} \in \mathbb{C}^N$, sampler set $\mathcal{F} = \{f_1, f_2, \cdots, f_M\}$ and measurements $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T \in \mathbb{R}_+^M$

$$y_i = |\langle \mathbf{x}, f_i \rangle|$$
$$\Leftrightarrow y_i^2 = f_i^H \mathbf{x} \mathbf{x}^H f_i$$
$$\Leftrightarrow y_i^2 = (f_i^T \otimes f_i^H) \text{Vec} \left( \mathbf{x} \mathbf{x}^H \right)$$

$\mathcal{F}$ can consist of either Fourier or general samplers [1].
Following is the summary of contemporary results on sufficient $M$ [1]:

1. For general data $\mathbf{x}$ and samplers $\mathcal{F}$, $M = 4N - 4$ is sufficient.
Following is the summary of contemporary results on sufficient $M$ [1]:

1. For general data $\mathbf{x}$ and samplers $\mathcal{F}$, $M = 4N - 4$ is sufficient.
2. For $s$-sparse data $\mathbf{x}$:
Prior Art and Results

Following is the summary of contemporary results on sufficient $M$ [1]:

1. For general data $x$ and samplers $\mathcal{F}$, $M = 4N - 4$ is sufficient.
2. For $s$-sparse data $x$:
   - If $\mathcal{F}$ consists of DFT samplers, $M \geq s^2 - s + 1$ with Collision Free Condition [2].
   - If $\mathcal{F}$ consists of random samplers, $M = O(s \log N)$ is sufficient via convex program.
Prior Art and Results

Following is the summary of contemporary results on sufficient $M$ [1]:

1. For general data $\mathbf{x}$ and samplers $\mathcal{F}$, $M = 4N - 4$ is sufficient.
2. For $s$-sparse data $\mathbf{x}$:
   - If $\mathcal{F}$ consists of DFT samplers, $M \geq s^2 - s + 1$ with Collision Free Condition [2].
   - If $\mathcal{F}$ consists of random samplers, $M = O(s \log N)$ is sufficient via convex program.
If the samplers in $\mathcal{F}$ are drawn from DFT of proper dimension, phase retrieval can be formulated as recovering data from its autocorrelation $r_x \in \mathbb{C}^{2N-1}$ defined as

$$[r_x]_l = \min\{N,N-l\} \sum_{k=\max\{1,1-l\}}^{\min\{N,N-l\}} x_k \bar{x}_{k+l}$$

$0 \leq |l| \leq N - 1$

The pair-wise products are coupled together which hides the sparse support of $x$. To avoid this, Collision Free Condition is proposed [2].

**Definition** (Collision-Free Condition) [2] A sparse vector $x$ has collision-free property if for pairs of distinct entries $(p, q), (m, n)$ in the support of $x$, $p - q \neq m - n$ unless $(p, q) = (m, n)$. 

Heng Qiao, Piya Pal (UMD)  
Phase Retrieval with PNFS  
November 14, 2015 6 / 22
Objectives of this paper

- $F$ consists of Fourier samplers.
- The sufficient measurement number $M$ should be $O(s \log N)$ with convex program.
- The Collision Free Condition on the sparse support should be relaxed.
Partial Nested Fourier Sampler

PNFS is a generalization of DFT-based sampler which with nested index array instead of consecutive one.

**Definition**

*(Partial Nested Fourier Sampler:)* We define a Partial Nested Fourier Sampler (PNFS) as

\[ f_i = \alpha \left[ z_i^1, z_i^2, \ldots, z_i^{N-1}, z_i^{2N-2} \right]^T \]

where \( \alpha = (4N - 5)^{-1/4} \) and \( z_i = e^{j2\pi (i-1)/(4N-5)} \).
Decoupling Effect of Nested Index Set

The nested index set $\mathcal{N} = \{1, 2, \cdots, N - 1, 2N - 2\}$ can resolve the coupling difficulty by exploiting the second-order difference set.

**Example**

Consider $N = 3$ and two different index set $\mathcal{N}_1 = \{0, 1, 2\}$ and $\mathcal{N}_2 = \{0, 1, 3\}$. $\mathcal{N}_2$ is a nested index set.

For $\mathcal{N}_1$, ignoring the negative part, we have

$$\{z^0_i : x_1 \bar{x}_1, x_2 \bar{x}_2, x_3 \bar{x}_3\} \ {z^1_i : x_1 \bar{x}_2, x_2 \bar{x}_3\} \ {z^2_i : x_1 \bar{x}_3\}$$

For $\mathcal{N}_2$ we have

$$\{z^0_i : x_1 \bar{x}_1, x_2 \bar{x}_2, x_3 \bar{x}_3\} \ {z^1_i : x_1 \bar{x}_2\} \ {z^2_i : x_2 \bar{x}_3\} \ {z^3_i : x_1 \bar{x}_3\}$$
Decoupling Effect of Nested Index Set

The nested index set $\mathcal{N} = \{1, 2, \cdots, N - 1, 2N - 2\}$ can resolve the coupling difficulty by exploiting the second-order difference set.

**Example**

Consider $N = 3$ and two different index set $\mathcal{N}_1 = \{0, 1, 2\}$ and $\mathcal{N}_2 = \{0, 1, 3\}$. $\mathcal{N}_2$ is a nested index set.

For $\mathcal{N}_1$, ignoring the negative part, we have

$$\{z_i^0 : x_1 \bar{x}_1, x_2 \bar{x}_2, x_3 \bar{x}_3\} \{z_i^1 : x_1 \bar{x}_2, x_2 \bar{x}_3\} \{z_i^2 : x_1 \bar{x}_3\}$$

For $\mathcal{N}_2$ we have

$$\{z_i^0 : x_1 \bar{x}_1, x_2 \bar{x}_2, x_3 \bar{x}_3\} \{z_i^1 : x_1 \bar{x}_2\} \{z_i^2 : x_2 \bar{x}_3\} \{z_i^3 : x_1 \bar{x}_3\}$$

**Advantage:** The sparse support is revealed in vectorized measurements model.
Measurement Structure with PNFS

Plugging the PNFS sampler $f_i$ into vectorized measurement model, we have

$$
y_i^2 = \frac{1}{\sqrt{4N-5}} \left[ z_i^{-(2N-3)}, \ldots, z_i^{-1}, 1, z_i^1, \ldots, z_i^{2N-3} \right] \tilde{x}
$$

where $\tilde{x} \in \mathbb{C}^{4N-5}$ is the corresponding rearranged version of $\text{Vec}(xx^H)$ with following form

$$
[\tilde{x}]_m = \begin{cases} 
\sum_{k=1}^{N} |x_k|^2 & m = 0 \\
\sum_{k=1}^{N-1-m} x_k \bar{x}_{k+m} & m = 1, 2, \ldots, N - 2 \\
x_{2N-2-m} \bar{x}_N & N - 1 \leq m \leq 2N - 3 \\
[\tilde{x}]_{-m} & m < 0
\end{cases}
$$
Permuted Version of PNFS

The support of $\mathbf{x}$ is easily identified in $\tilde{\mathbf{x}}$ if $x_N$ is nonzero. If no prior knowledge available, we will need column-permuted version of PNFS defined as

$$f_i^{(l)} = \frac{1}{\sqrt{4N-5}} \left[ z_i^1, z_i^2, \cdots, z_i^{N-1}, z_i^{2N-2} \right] \Pi^{(l)}$$

(3)

$\Pi^{(l)}$ is a permuting matrix such that the vector $\mathbf{x}^{(l)} = \Pi^{(l)} \mathbf{x}$ satisfies $[\mathbf{x}^{(l)}]_l = x_N, [\mathbf{x}^{(l)}]_N = x_l, [\mathbf{x}^{(l)}]_i = x_i, i \neq l, N.$

For each $l$, we collect $\tilde{M}$ phaseless measurements $y_i^{(l)}, i = 1, 2, \cdots, \tilde{M}$ using the permuted PNFS vector (3) and obtain

$$\tilde{\mathbf{y}}^{(l)} = \mathbf{Z} \tilde{\mathbf{x}}^{(l)}$$

(4)

where $[\tilde{\mathbf{y}}^{(l)}]_i = (y_i^{(l)})^2, [\mathbf{Z}]_{i,m} = \frac{1}{\sqrt{4N-5}} e^{j2\pi \frac{(i-1)m}{4N-5}}$, $1 \leq i \leq \tilde{M}, -2N + 3 \leq m \leq 2N - 3.$

Objective: If $\mathbf{x}$ is non-zero, we will finally find $\Pi^{(l)}$ such that $[\mathbf{x}^{(l)}]_N \neq 0.$
Iterative Algorithm

**Input:** data $\mathbf{x}$  \hspace{1cm} **Output:** estimation $\mathbf{x}^\#$

**Initialization:** $l = N$

**Loop:**

1. **Step S1:** Using the permuted PNFS vectors (3), obtain $4N - 5$ phaseless measurements

   $$y_i^{(l)} = | \langle \mathbf{x}, \mathbf{f}_i^{(l)} \rangle |, \; i = 1, 2, \cdots, 4N - 5$$

   Recover $\tilde{\mathbf{x}}^{(l)} = \mathbf{Z}^{-1} \tilde{\mathbf{y}}^{(l)}$

2. **Step S2:** If $[\tilde{\mathbf{x}}^{(l)}]_m = 0, \forall |m| \geq N - 1$, declare $x_l = 0$. Assign $l \rightarrow l - 1$ and go back to Step S1.

   If $[\tilde{\mathbf{x}}^{(l)}]_m \neq 0$ for some $m$ with $|m| \geq N - 1$, proceed to the recovery stage.
Iterative Algorithm: Continued

Recovery:

1. Choose $m^* \in \{1, 2, \ldots, N-2\}$ such that $[\tilde{x}^{(l)}]_{m^*} \neq 0$ and compute

$$|x_N^{(l)}| = \sqrt{[\tilde{x}^{(l)}]_{m^*}/\beta}$$

and

$$\beta = \sum_{k=1}^{N-1-m^*} [\tilde{x}^{(l)}]_{2N-2-k} [\tilde{x}^{(l)}]_{2N-2-k-m^*}$$

2. Obtain estimate $x^\#$ as

$$[x^\#]_q = \begin{cases} 
\left( \frac{[\tilde{x}^{(l)}]_{2N-2-q}}{|x_N^{(l)}|} \right) & q \neq \{l, N\} \\
|x_N^{(l)}| & q = l \\
\frac{[\tilde{x}^{(l)}]_{2N-2-l}}{|x_N^{(l)}|} & q = N 
\end{cases}$$
The complexity of the algorithm mainly depends on the number of trials to find $[x^{(l)}]_N \neq 0$.

**Theorem**

Let $x \in \mathbb{C}^N$ be $s$-sparse with $s \geq 3$. The estimate $x^\#$ produced by the iterative algorithm described in Table 1 is equal to $x$ (in the sense of $\mathbb{C} \setminus \mathbb{T}$) if the total number of phaseless measurements $M$ equals $4N - 5$ for the best case and $(N - s + 1)(4N - 5)$ for the worst case.

**Corollary**

If $x$ is not sparse (i.e. $s = N$), the number of measurements needed for recovering $x$ is $M = 4N - 5$. 
Sketch of Proof

The main idea in the proof is to show the existence of $m^*$ such that $[\tilde{x}^{(l)}]_{m^*} \neq 0$. Denote $\tilde{x} = [x_1, x_2, \cdots, x_{N-1}]^T$ and let $r_{\tilde{x}} \in \mathbb{C}^{2N-3}$ be the autocorrelation vector of $\tilde{x}$. Suppose $m^*$ does not exist, implying $[\tilde{x}]_m = 0$ for $1 \leq |m| \leq N - 2$. Hence, $[r_{\tilde{x}}]_n = \gamma \delta(n)$ where $\gamma = [\tilde{x}]_0 - |x_N|^2$ and $\delta(n)$ is Kronecker delta. This means that $\hat{r}_{\tilde{x}}(e^{j\omega}) = \sum_{n=-N+2}^{N-2} [r_{\tilde{x}}]_n e^{-j\omega n}$ is an all-pass filter. However, $\hat{r}_{\tilde{x}}(e^{j\omega}) = |\hat{x}(e^{j\omega})|^2$ where $\hat{x}(e^{j\omega}) = \sum_{n=-N+2}^{N-2} [\tilde{x}]_n e^{-j\omega n}$. This implies $\hat{x}(e^{j\omega})$ is also an all-pass filter. Since $\hat{x}(e^{j\omega})$ is an FIR filter, this is not possible unless we have

$$[\tilde{x}]_n = \lambda \delta(n - n_0) \quad (5)$$

for some $n_0$ satisfying $1 \leq n_0 \leq N-1$ and $\lambda$ is a constant. However, since $s \geq 3$, $\tilde{x}$ has at least two non zero entries which contradicts (5). Therefore, the existence of $m^*$ is guaranteed.
Sketch of Proof

The main idea in the proof is to show the existence of $m^*$ such that $[\tilde{x}(l)]_{m^*} \neq 0$. Denote $\mathbf{x} = [x_1, x_2, \cdots, x_{N-1}]^T$ and let $r_{\mathbf{x}} \in \mathbb{C}^{2N-3}$ be the autocorrelation vector of $\mathbf{x}$. Suppose $m^*$ does not exist, implying $[\tilde{x}]_m = 0$ for $1 \leq |m| \leq N - 2$. Hence, $[r_{\mathbf{x}}]_n = \gamma \delta(n)$ where $\gamma = [\tilde{x}]_0 - |x_N|^2$ and $\delta(n)$ is Kronecker delta. This means that $\hat{r}_{\mathbf{x}}(e^{j\omega}) = \sum_{n=-N+2}^{N-2} [r_{\mathbf{x}}]_n e^{-j\omega n}$ is an all-pass filter. However, $\hat{r}_{\mathbf{x}}(e^{j\omega}) = |\hat{x}(e^{j\omega})|^2$ where $\hat{x}(e^{j\omega}) = \sum_{n=-N+2}^{N-2} [\mathbf{x}]_n e^{-j\omega n}$. This implies $\hat{x}(e^{j\omega})$ is also an all-pass filter.

Since $\hat{x}(e^{j\omega})$ is an FIR filter, this is not possible unless we have

$$[\tilde{x}]_n = \lambda \delta(n - n_0)$$

for some $n_0$ satisfying $1 \leq n_0 \leq N - 1$ and $\lambda$ is a constant. However, since $s \geq 3$, $\mathbf{x}$ has at least two non-zero entries which contradicts (5).

Therefore, the existence of $m^*$ is guaranteed.

Observation: PNFS hits the lower bound $4N - 5$ if $\mathbf{x}$ has no zero entries.
Cancellation of Measurements

If we have some prior knowledge of $x$ that $x_N$ is nonzero, PNFS can achieve better bound for sparse phase retrieval. This is based on the idea of cancellation via two sets of measurements, $\tilde{y}, \tilde{y}' \in \mathbb{C}^{\tilde{M}}$ as

$$[	ilde{y}]_i = |\langle x, f_i \rangle|^2$$
$$[	ilde{y}']_i = |\langle x, f'_i \rangle|^2$$

where $f_i$ denotes the PNFS vector (as in Def. 3) and $f'_i$ is defined as

$$f'_i = \frac{1}{\sqrt{4N - 5}} \left[ z_1^i, z_2^i, \ldots, z_{N-1}^i, 0 \right]$$

where $z_i = e^{j2\pi(i-1)/(4N-5)}$. 

\[ z_i = e^{j2\pi(i-1)/(4N-5)}. \]
Denoting $\hat{y} = \tilde{y} - \tilde{y}'$, we have

$$\hat{y} = Z\hat{x}$$

where

$$\left[\hat{x}\right]_m = \begin{cases} 
|x_N|^2 & m = 0 \\
0 & m = 1, 2, \cdots, N - 2 \\
x_{2N-2-m}\tilde{x}_N & m = N - 1, \cdots, 2N - 3 \\
[\hat{x}]_{-m} & m < 0
\end{cases}$$

and $Z \in \mathbb{C}^{\tilde{M},4N-5}$ defined as in (4). Notice that $\tilde{x}$ has sparsity $2s - 1$ and support of $x$ (except the $N$th entry) is identical to that of the subvector of $\tilde{x}$ indexed by $m = N - 1, \cdots, 2N - 3$. 
The power of cancellation is revealing the sparse support of $\mathbf{x}$ and then convex program is applicable. We can recover $\hat{\mathbf{x}}$ by solving the $l_1$ minimization:

$$\min_{\theta} \|\theta\|_1 \quad \text{subject to} \quad \hat{\mathbf{y}} = \mathbf{Z}\theta \quad (P1)$$

The vector $\mathbf{x}$ can then be recovered from the solution of $(P1)$. 

**Theorem**

Let $\mathbf{x} \in \mathbb{C}^N$ be a sparse vector with $s$ non-zero elements and $\mathbf{x}^N \neq 0$. Suppose we construct the difference measurement vector $\hat{\mathbf{y}}$ as in (9) using $\tilde{M}$ pairs of sampling vectors $\{f_{ik}, f'_{ik}\}_{k=1}^{\tilde{M}}$ where indices $i_k$ are selected uniformly at random between 1 and $4N - 5$. Then $\mathbf{x}$ can be recovered (in sense of $C^T$) by solving $(P1)$ if $\tilde{M} = C_2 s \log N$ for some constant $C$. 

Heng Qiao, Piya Pal (UMD) 
Phase Retrieval with PNFS 
November 14, 2015
The power of cancellation is revealing the sparse support of \( x \) and then convex program is applicable. We can recover \( \hat{x} \) by solving the \( l_1 \) minimization:

\[
\min_{\theta} \|\theta\|_1 \quad \text{subject to} \quad \hat{y} = Z\theta \quad (P1)
\]

The vector \( x \) can then be recovered from the solution of (P1).

**Theorem**

Let \( x \in \mathbb{C}^N \) be a sparse vector with \( s \) non zero elements and \( x_N \neq 0 \). Suppose we construct the difference measurement vector \( \hat{y} \) as in (9) using \( \tilde{M} \) pairs of sampling vectors \( \{ f_{ik}, f'_{ik} \}_{k=1}^{\tilde{M}} \) where indices \( i_k \) are selected uniformly at random between 1 and \( 4N - 5 \). Then \( x \) can be recovered (in sense of \( \mathbb{C} \setminus \mathbb{T} \)) by solving (P1) if \( \tilde{M} = C s \log N \) for some constant \( C \).
The probability of “no-collision” as a function of sparsity $s$. The ambient dimension is $N = 10000$ and the result is averaged over 2000 runs.
Validation of the Theorem 2

The global phase ambiguity is $\rho = x_N / x_N^\#$. Using $\rho$ we can compute the entry-wise estimation error as $|x_i - \rho x_i^\#|$ for $1 \leq i \leq N$.

**Figure:** The phase transition plot for Theorem 2. $M = 2\tilde{M}$ is the total number of measurements needed and $N = 150$. The red line represents $3s\log N$. The color bar denotes probability of success from 0 to 1. The white cells denote successful recoveries (i.e. $|x_i - \rho x_i^\#| \leq 10^{-6}$ for all entries) and black cells denote failures. The results are averaged over 100 runs.
The PNFS are general Fourier samplers and can be easily implemented via DFT plus coded diffraction [3].

The recovery algorithm is deterministic for general case and hits lower bound for nowhere vanishing data $x$.

If prior knowledge available, $O(s \log N)$ is possible with cancellation process and convex program.
