Experimental compressive phase space tomography

Tian, Lei

https://hdl.handle.net/2144/23575

Boston University
Experimental compressive phase space tomography

Lei Tian,1,∗ Justin Lee,2 Se Baek Oh,1,4 and George Barbastathis1,3

1Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
2Department of Health Science and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
3Singapore-MIT Alliance for Research and Technology (SMART) Centre, 3 Science Drive 2, Singapore 117543, Singapore
4Currently with KLA-Tencor, 1 Technology Drive, Milpitas, CA 95035, USA

∗lei_tian@mit.edu

Abstract: Phase space tomography estimates correlation functions entirely from snapshots in the evolution of the wave function along a time or space variable. In contrast, traditional interferometric methods require measurement of multiple two-point correlations. However, as in every tomographic formulation, undersampling poses a severe limitation. Here we present the first, to our knowledge, experimental demonstration of compressive reconstruction of the classical optical correlation function, i.e. the mutual intensity function. Our compressive algorithm makes explicit use of the physically justifiable assumption of a low-entropy source (or state.) Since the source was directly accessible in our classical experiment, we were able to compare the compressive estimate of the mutual intensity to an independent ground-truth estimate from the van Cittert–Zernike theorem and verify substantial quantitative improvements in the reconstruction.

© 2012 Optical Society of America

OCIS codes: (030.0030) Coherence and statistical optics; (050.5082) Phase space in wave optics; (070.0070) Fourier optics and signal processing; (100.6950) Tomographic image processing.

1. Introduction

Correlation functions provide complete characterization of wave fields in several branches of physics, e.g. the mutual intensity of stationary quasi–monochromatic partially coherent light [1], and the density matrix of conservative quantum systems (i.e., those with a time–independent Hamiltonian) [2]. Classical mutual intensity expresses the joint statistics between two points on a wavefront, and it is traditionally measured using interferometry: two sheared versions of a field are overlapped in a Young, Mach–Zehnder, or rotational shear [3,4] arrange-
ment, and two–point ensemble statistics are estimated as time averages by a slow detector under the assumption of ergodicity [1, 5].

As an alternative to interferometry, phase space tomography (PST) is an elegant method to measure correlation functions. In classical optics, PST involves measuring the intensity under spatial propagation [6–8] or time evolution [9]. In quantum mechanics, analogous techniques apply [10–13]. However, the large dimensionality of the unknown state makes tomography difficult. In order to recover the correlation matrix corresponding to just \( n \) points in space, a standard implementation would require at least \( n^2 \) data points.

Compressed sensing [14–16] exploits sparsity priors to recover missing data with high confidence from a few measurements derived from a linear operator. Here, sparsity means that the unknown vector contains only a small number of nonzero entries in some specified basis. Low–rank matrix recovery (LRMR) [17, 18] is a generalization of compressed sensing from vectors to matrices: one attempts to reconstruct a high–fidelity and low–rank description of the unknown matrix from very few linear measurements.

In this paper, we present the first, to our knowledge, experimental measurement and verification of the correlation function of a classical partially coherent field using LRMR. It is worth noting that LRMR came about in the context of compressive quantum state tomography (QST) [19], which utilizes different physics to attain the same end goal of reconstructing the quantum state. In PST, one performs tomographic projection measurements, rotating the Wigner space between successive projections by evolving the wave function [6, 7]. This is directly analogous to the classical optical experiment we are presenting here, where we perform intensity measurements (i.e., tomographic projections in Wigner space) and utilize propagation along the optical axis to rotate the Wigner space between projections. The difference lies in the fact that in QST the state is recovered via successive applications of the Pauli dimensionality–reducing operator, and there is no need to evolve the state. Nevertheless, both approaches lead to the same Hermitian LRMR problem, as long as the assumption of a quasi–pure unknown state is satisfied. In [20], it was shown that estimation of a low–rank matrix of dimension \( n \) and rank \( r \) requires only \( O(rn \ln n) \) to \( O(rn \ln^2 n) \) data points. A similar LRMR method was also used to recover the complex amplitude of an unknown object under known illumination [21–23]. Since the complex amplitude of the object is time–invariant, a rank–one solution was assumed in these works.

The low–rank assumption for classical partially coherent light anticipates a source composed of a small number of mutually incoherent effective sources, i.e. coherent modes [24], to describe measurements. This is essentially equivalent to the low entropy assumption [19], i.e. a nearly pure quantum state in the quantum analogue. This assumption is valid for lasers, synchrotron and table–top X–ray sources [25], and Köhler illumination in optical microscopes [1]. An additional requirement for LRMR to succeed is that measurements are “incoherent” with respect to the eigenvectors of the matrix, i.e. the measured energy is approximately evenly spread between modes [20, 21]. Diffraction certainly mixes the coherent modes of the source rapidly, so we expect LRMR to perform well for classical PST. The same expectation for QST has already been established [19].

2. Theory and method

The two–point correlation function of a stationary quasi–monochromatic partially spatially coherent field is the mutual intensity function [1]

\[
J(x_1, x_2) = \langle g^* (x_1) g(x_2) \rangle ,
\]

where \( \langle \cdot \rangle \) denotes the expectation value over a statistical ensemble of realizations of the field \( g(x) \).
The measurable quantity of the classical field, \( i.e. \) the intensity, after propagation by distance \( z \), is \([1]\)

\[
I(x_0; z) = \int dx_1 dx_2 J(x_1, x_2) \exp \left( -\frac{i\pi}{\lambda z} (x_1^2 - x_2^2) \right) \exp \left( i2\pi \frac{x_1 - x_2}{\lambda z} x_0 \right). \tag{2}
\]

This can be expressed in operator form as

\[
I = \text{tr}(P_{x_0} J), \tag{3}
\]

where \( P \) denotes the free–space propagation operator that combines both the quadratic phase and Fourier transform operations in Eq. (2), \( \text{tr}(\cdot) \) computes the trace, and \( x_0 \) denotes the lateral coordinate at the observation plane. By changing variables \( x = (x_1 + x_2)/2, \ x' = x_1 - x_2 \) and Fourier transforming the mutual intensity with respect to \( x \) we obtain the Ambiguity Function (AF) \([26–28]\)

\[
\mathcal{A}(u', x') = \int J(x + \frac{x'}{2}, x - \frac{x'}{2}) \exp \left(-i2\pi u' x\right) dx. \tag{4}
\]

Equation (2) can be written as \([6–8, 26, 28]\),

\[
\tilde{I}(u'; z) = \mathcal{A}(u', \lambda z u'), \tag{5}
\]

where \( \tilde{I} \) is the Fourier transform of the vector of measured intensities with respect to \( x_0 \). Thus, radial slices of the AF may be obtained from Fourier transforming the vectors of intensities measured at corresponding propagation distances, and from the AF the mutual intensity can be recovered by an additional inverse Fourier transform, subject to sufficient sampling.

To formulate a linear model for compressive PST, the measured intensity data is first arranged in Ambiguity space. The mutual intensity is defined as the “sparse” unknown to solve for. To relate the unknowns (mutual intensity) to the measurements (AF), the center–difference coordinate–transform is first applied, expressed as a linear transformation \( T \) upon the mutual intensity \( J \), followed by Fourier transform \( F \), and adding measurement noise \( e \) as

\[
\mathcal{A} = F \cdot T \cdot J + e. \tag{6}
\]

The mutual intensity propagation operator is unitary and Hermitian, since it preserves energy. We use eigenvalue decomposition to determine the basis where the measurement is sparse. The resulting basis, \( i.e. \) the set of eigenvectors, is also known as coherent modes in optical coherence theory, whereas the whole process is known as coherent mode decomposition \([24]\). The goal of the LRMR method is to minimize the number of coherent modes to describe measurements. By doing LRMR, we impose two physically meaningful priors: (1) existence of the coherent modes \([24]\), and (2) sparse representation of the partially coherent field in terms of coherent modes.

Mathematically, if we define all the eigenvalues \( \lambda_i \) and the estimated mutual intensity as \( \hat{J} \), the method can be written as

\[
\text{minimize } \text{rank}(\hat{J}) \]

subject to \( \mathcal{A} = F \cdot T \cdot J \),

\[
\lambda_i \geq 0, \text{ and } \sum_i \lambda_i = 1. \tag{7}
\]

Direct rank minimization is NP–hard; however, it can be accomplished by solving instead a proxy problem: convex minimization of the “nuclear norm” (\( \ell_1 \) norm) of the matrix \( J \) \([17, 29]\).
The corresponding problem is stated as

\[
\begin{align*}
\text{minimize} & \quad \| \hat{J} \|_* \\
\text{subject to} & \quad \mathcal{A} = \mathcal{F} \cdot T \cdot \hat{J}, \\
& \quad \lambda_i \geq 0, \text{ and } \sum_i \lambda_i = 1,
\end{align*}
\]

where the nuclear norm is the sum of the singular values \( \sigma_i = |\lambda_i|, \| \hat{J} \|_* = \sum_i \sigma_i \). This problem is convex and a number of numerical solvers can be applied to solve it. In our implementation, we used the singular value thresholding (SVT) method [30]. The output estimate after each iteration of SVT typically has a sub-normalized total energy, i.e. \( \sum \lambda_i < 1 \); we compensated for this by renormalizing at the end of each iteration [19].

3. Numerical simulations

![Fig. 1.](image)

(a) Input mutual intensity of a GSMS with parameters \( \sigma_I = 17 \) and \( \sigma_c = 13 \), (b) data point locations in the Ambiguity space, mutual intensities estimated by (c) FBP and (d) LRMR methods.

First, we demonstrate the LRMR method with a numerical example using a 1D Gaussian-Schell model source (GSMS). Both the intensity distribution and the degree of coherence of GSMS follow a Gaussian distribution [31]

\[
J(x_1, x_2) = |I(x_1)|^{1/2} |I(x_2)|^{1/2} \mu(x_1 - x_2),
\]

(9)
Fig. 2. The first nine coherent modes of the mutual intensity in Fig. 1(a). (a) Theoretical modes, and (b) LRMR estimates.

Fig. 3. Eigenvalues of the mutual intensity in Fig. 1(a). (a) Theoretical values, (b) FBP estimates, (c) LRMR estimates, and (d) absolute errors in the LRMR estimates versus mode index.

and

\[ I(x) = \exp \left( \frac{x^2}{2\sigma^2} \right), \]
\[ \mu(x_1 - x_2) = \exp \left( \frac{(x_1 - x_2)^2}{2\sigma^2} \right). \]  

\[ (10) \]
where $\sigma_d$ determines the spatial extent of the source, and $\sigma_c$ is proportional to the coherence length and determines the number of coherent modes in the input source. The eigenvalues of GSMS are never zero (analytical solution given in [31]). We defined the number of modes (rank of the source) $r$ as the first $r$ modes containing the 99% of the total energy.

One example is shown in Fig. 1(a). The parameters in this example are $\sigma_d = 17$ and $\sigma_c = 13$ (rank $r = 6$). Intensities are calculated at 40 different axial distances and the coverage in Ambiguity space is shown in Fig. 1(b). We simulate the case where data from both the near field and the far field are missing due to the finite range of camera scanning motion allowed in the actual experiment. The missing cone around the $u'$-axis is due to missing data from near field, while the data missing from far field results in the missing cone around the $x'$-axis. Both cones have an apex angle of 20 degrees.

For comparison, the data is first processed using the traditional filtered-backprojection (FBP) method [32]. Applying the Fourier-slice theorem to Eq. (5) implies that the 1D Fourier transform of a radial slice in the Ambiguity space (an intensity measurement) is related to a projection in the AF's 2D Fourier space (the Wigner space [33,34]). The Wigner distribution function (WDF) is related to the mutual intensity by

$$W(x, u) = \int J\left(x + \frac{x'}{2}, x - \frac{x'}{2}\right) \exp(-i2\pi xu') dx'.$$

To implement the FBP method, each intensity projection is first filtered by a Ram–Lak kernel apodized by a Hamming window; the estimated WDF is obtained by back-projecting all the filtered intensities, and then an inverse Fourier transform is applied to produce the mutual intensity. Figure 1(c) shows the reconstructed mutual intensity following this procedure. Three types of artifacts can be seen in this reconstruction. First, the reconstructed mutual intensity has lower values along the diagonal of the matrix due to the missing cones. However, this is unphysical because a correlation function should always have maximum value at zero separation. Second, the estimated degree of coherence is lower than the original field. The third artifact is the high frequency noise around the diagonal of the matrix, which is due to undersampling between the radial slices. All these artifacts have been greatly suppressed or completely removed by LRMR, whose reconstruction result is shown in Fig. 1(d). The disappearance of the correlation peak along the diagonal (i.e., the intensity) when we use FBP for the reconstruction can be best explained with the help of Figure 1(b). Going from the Ambiguity space to the mutual intensity space involves Fourier transforming along horizontal lines, parallel to the $u'$ axis. The diagonal in particular corresponds to the line $x' = 0$. It can be easily seen that, due to the missing cone, pretty much all the data are missing from that line, except near the origin; thus resulting in a low-pass filtering effect. The fact that the compressive reconstruction method manages to restore the physically correct values of the correlation along the diagonal corroborates that the missing cone is successfully retrieved in our LRMR reconstruction. The FBP reconstruction may also be compared quantitatively to the compressive reconstruction in terms of the global degree of coherence parameter $\bar{\mu} = \sqrt{\sum_i \lambda_i^2 / \sum_i |\lambda_i|}$ [35, 36], which was found as 0.150 and 0.617, respectively; the true state has $\bar{\mu} = 0.618$.

The coherent modes for a GSMS are Hermite–Gaussian sources [31]. The theoretical and LRMR estimated first nine coherent modes in this example are shown in Fig. 2(a) and 2(b), respectively. The theoretical eigenvalues are shown in Fig. 3(a). The FBP and LRMR estimated eigenvalues are compared in Fig. 3(b) and 3(c), respectively. The FBP estimates have several negative values, which does not satisfy the positive energy constraint. The absolute errors in LRMR estimates are plotted in Fig. 3(d).

Next, we study the noise performance of the LRMR method with a numerical example. In this simulation, the dimension of the input GSMS is $256 \times 256$ with parameters $\sigma_d = 36$ and $\sigma_c = 18$.
(rank $r = 9$). We generate noisy data with different signal-to-noise ratio (SNR) from both an additive random Gaussian noise model and a Poisson noise model. However, we emphasize that the reconstruction algorithm does not make use of the noise statistics. For each SNR level, we repeat the simulation 100 times with different random noise terms, and then record the average relative mean-square-error (MSE) from the LRMR reconstruction. The ratio between the number of samples taken from the intensity measurements and the rank $r$ of the input mutual intensity matrix determines the oversampling rate [22]. This rate is plotted versus relative MSE for different SNR cases in Fig. 4. For good performance, the required oversampling rate is at least 5–6 (the theoretical oversampling rate is on the order of $\ln(256) = 5.5$ according to [20]). Furthermore, the LRMR method is robust to noise in the sense that the reconstruction degrades gracefully as the SNR decreases.

4. Experimental result

The experimental arrangement is illustrated in Fig. 5. The illumination is generated by an LED with 620nm central wavelength and 20nm bandwidth. To generate partially coherent illumination, a single slit of width 355.6μm (0.014") is placed immediately after the LED and one focal length (75 mm) to the left of a cylindrical lens. One focal length to the right of the lens,
we place the second single slit of width 457.2μm \( (0.018') \), which is used as a one–dimensional (1D) object.

The goal is to retrieve the mutual intensity immediately to the right of the object from a sequence of intensity measurements at varying \( z \)–distances downstream from the object, as described in the theory. We measured the intensities at 20 \( z \)–distances, ranging from 18.2mm to 467.2mm, to the right of the object. The data are given in Fig. 6. Each 1D intensity measurement consists of 512 samples, captured by a CMOS sensor with 12μm pixel size. The dimension of the unknown mutual intensity matrix to be recovered is \( 512 \times 512 \). Since only intensities at positive \( z \), i.e. downstream from the object, are accessible, we can only fill up the top right and bottom left quadrants of Ambiguity space. The other two quadrants are filled symmetrically, i.e. assuming that if the field propagating to the right of the object were phase conjugated with respect to the axial variable \( z \), it would yield the correct field to the left of the object, i.e. negative \( z \) [8, 13]. Under this assumption, a total of 40 radial slices are sampled in Ambiguity space, as shown in Fig. 7. The apex angle of the missing cone around the \( u' \)–axis is approximately 17.4 degrees, and the one around the \( x' \)–axis is approximately 28.6 degrees. The number of
measurements is only 7.8% of the total number of entries in the unknown mutual intensity matrix.

The reconstructions from the FBP and LRMR methods are compared in Fig. 8(a) and 8(b), respectively. The FBP reconstruction suffers from the same artifacts detailed in the numerical simulations section. All these artifacts are greatly suppressed or completely removed in the LRMR reconstruction. In the real part of the reconstruction, the width of the square at the center is approximately $456\mu m$ (38 pixels), which agrees with the actual width of the slit. The imaginary part is orders of magnitude smaller than the real part.

FBP estimated eigenvalues contain several negative values, and are shown in Fig. 9(a). This does not satisfy the positive energy constraint. LRMR estimated eigenvalues are compared in Fig. 9(b), and all eigenvalues are positive.

We further validated our compressive estimates by measuring the field intensity immediately to the right of the illumination slit [Fig. 10(a)]. Assuming that the illumination is spatially incoherent (a good assumption in the LED case), the mutual intensity of the field immediately to
the left of the object is the Fourier transform of the measured intensity, according to the van Cittert–Zernike theorem [1, 5]. This calculated mutual intensity, based on the measurement of Fig. 10(a) and screened by the object slit, is shown in Fig. 10(b). The eigenvalues computed by coherent mode decomposition are shown in Fig. 10(c) and are in good agreement with the LRMR estimates, as compared in Fig. 10(d). It is seen that 99% of the energy is contained in the first 13 modes, which confirms our low–rank assumption. The FBP reconstruction may also be compared quantitatively to the compressive reconstruction in terms of the global degree of coherence parameters, which were experimentally found as 0.12 and 0.46, respectively; whereas the estimate yielded by the van Cittert–Zernike theorem is 0.49. The first nine eigenvectors of each individual mode are shown and compared in Fig. 11. Small errors in the compressive estimate are because the missing cone is still not perfectly compensated by the compressive approach, and because of other experimental imperfections.

In this classical experiment, we have the benefit that direct observation of the one–dimensional object is available; thus, we were able to carry out quantitative analysis of the accuracy of the compressive estimate. In the quantum analogue of measuring a complete quantum state, direct observation would have of course not been possible, but the accuracy attained through the compressive estimate should be comparable, provided the low entropy assumption
5. Discussion

In conclusion, we experimentally demonstrated compressive reconstruction of the mutual intensity function of a classical partially coherent source using phase space tomography. By exploiting the physically justifiable assumption of a quasi–pure source, both measurement and post–processing dimensionality are greatly reduced. We used the van Cittert–Zernike theorem to estimate the true mutual intensity function as a way to cross–validate the compressive reconstruction, and found indeed good agreement.

Here we followed a much simplified version of the approach described in [21] which showed that the complex operators describing the measurements should be uniformly distributed in the $n$–dimensional unit sphere, whereas we simply utilized free space propagation. The phase masks described in [21] to implement optimal sampling are outside the scope of the present
Acknowledgment

The authors thank Baile Zhang, Jonathan Petruccelli, and Yi Liu for helpful discussions. Financial support was provided by Singapore’s National Research Foundation through the Center for Environmental Sensing and Modeling (CENSAM) and BioSystems and bioMechanics (BioSyM) independent research groups of the Singapore-MIT Alliance for Research and Technology (SMART) Centre, by the US National Institutes of Health and by the Chevron–MIT University Partnership Program.