Sparsity-based recovery of three-photon quantum states from two-fold correlations: supplementary material

DIKLA OREN¹, YOAV SHECHTMAN^{1,2}, MAOR MUTZAFI¹, YONINA C. ELДАР³,
MORDECHAI SEGEV^{1,*}

¹Physics Department and Solid State Institute, Technion, 32000 Haifa, Israel

²Department of Chemistry, Stanford University, 375 North-South Mall, Stanford, California 94305, USA

³Electrical Engineering Department, Technion, 32000 Haifa, Israel

*Corresponding author: msegev@tx.technion.ac.il

Published 26 February 2016

This document provides supplementary information to “Sparsity-based recovery of three-photon quantum states from two-fold correlations,” <http://dx.doi.org/10.1364/optica.3.000226>. © 2016 Optical Society of America
<http://dx.doi.org/10.1364/optica.3.000226.s001>

Matrix Elements of the Sensing Matrix

In the paper, we cast the problem of finding the eigenvalues of the density operator from two-fold coincidences into matrix form and obtain Eq. 6. The sensing matrix appearing there is derived from the two-fold correlation measurement $\Gamma_{q,r}^{(2)} = \text{Tr}\{\rho a_q^\dagger a_r^\dagger a_r a_q\}$, where the operators are evaluated at the output of the array according to the solution of Eq. 2 in the paper.

The solution of the system of equations is of the form $a_i^\dagger(z) = \sum U_{ij}(z) a_j^\dagger(z)$ such that $U(z)$ is the propagator of the Hamiltonian in Eq. 1. Thus, the sensing matrix elements take the form:

$$M_{(i,j),k} = \sum_{m,n,l,r} U_{mi} U_{nj} U_{lj}^* U_{ri}^* \langle \psi_k | a_q^\dagger a_r^\dagger a_r a_q | \psi_k \rangle. \quad (S1)$$

We give two examples for different $\{|\psi_k\rangle\}_k$. In general, since every density matrix can be diagonalized $\rho = V D V^\dagger = \sum \lambda_j |d_j\rangle\langle d_j|$, the same relation as in the examples shown is satisfied $\Gamma_\alpha = \text{Tr}(\rho \hat{\Gamma}_\alpha) = \text{Tr}(V D V^\dagger \hat{\Gamma}_\alpha) = \sum \lambda_j \langle d_j | V^\dagger \hat{\Gamma}_\alpha V | d_j \rangle$, α being an index of a pair of waveguides. The properties of the matrix $\tilde{M}_{\alpha,j} = \langle d_j | V^\dagger \hat{\Gamma}_\alpha V | d_j \rangle$ enable the recovery of the signal.

Keeping in mind that the operator denoted by V is derived from the propagation in the array, manipulation of the coupling and propagation constants of the waveguides can lead to enhancement of the smearing, as discussed in the paper, for an arbitrary basis.

Low rank and Sparsity of the Coefficient Vector

As it turns out, the combination of having a low rank density operator and knowing the eigenbasis implies that the vector $\mathbf{p}$ is

sparse. This is because the density operator ρ is positive semidefinite, therefore diagonalizable. Thus a low rank ρ means that the number of its nonzero eigenvalues is small. At the same time, the coefficient vector $\mathbf{p}$, which consists of the eigenvalues of ρ , has only r nonzero elements, r being the rank of the original operator. Hence, the vector $\mathbf{p}$ is sparse and is amenable to recovery from partial measurements, as we show in the paper.

Recovery Methods

In our problem, the desired vector is related to the measurements in the following manner

$$\mathbf{y} = A\mathbf{x} + \epsilon \quad (S2)$$

$$\mathbf{y}, \epsilon \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n.$$

Here m is the number of measurements, n is the number of basis vectors, $m < n$ and ϵ represents additive noise.

When considering the dimensions of the objects in question, in particular the fact that $m < n$, we get that $A \in \mathbb{R}^{m \times n}$ has a nontrivial null space. This is the manifestation of using lower order coincidences. In this scenario, the input vector $\mathbf{x}$ cannot be uniquely recovered, unless some underlying structure in the sought signal $\mathbf{x}$ is exploited.

Since $\mathbf{x}$ is derived from a density matrix representing the state, it obeys the following constraints:

$$0 \leq x_i \leq 1 \quad i = 1, \dots, n \quad (S3)$$

$$\sum_i x_i = 1.$$

We further assume that the desired state is close to a pure state, or a slightly degraded pure state. The manifestation of that in the vector as derived from the state in the paper would be having only a small fraction of nonzero elements in x . When taking into consideration depolarization, this means that x has a small fraction of nonzero significant elements, whereas the others are very small in comparison. States admitting this kind of structure are called sparse and compressible, respectively.

The recovery of a sparse vector can be formulated in the following optimization problem [1]:

$$\begin{aligned} & \min_x \|x\|_0 \\ & \text{s. t. } \|Ax - y\|_2 < \mu \\ & 0 \leq x_i \leq 1 \\ & \sum_i x_i = 1. \end{aligned} \quad (S4)$$

Here $\|\cdot\|_0$ denotes the l_0 norm, which is the number of nonzero elements in the vector. The first constraint describes the compatibility of the signal and the measurements, where μ is related to the 2-norm of the expected measurement noise. The other two constraints are derived from the physical state.

Here we present two methods for solving the optimization problem (S4). The first one uses the sparse nature of the vector implicitly, whereas the second one does it explicitly.

The first approach relies on solving the following constrained least-squares problem

$$\begin{aligned} & \min_x \|Ax - y\|_2 \\ & \text{s. t. } 0 \leq x_i \leq 1 \\ & \sum_i x_i = 1. \end{aligned} \quad (S5)$$

Even though the sparse nature of the solution is not being used explicitly, it has been shown that under certain constraints, if there exists a nonnegative sparse solution to a linear system, then this solution is unique [2]. Thus, if a nonnegative solution to such a system is found - no matter through which method - then because of the uniqueness this solution must be sparse. In the example described here, we solve the optimization problem through least-squares, and find that we are able to recover the original state at high fidelity only when the solution is sparse. Figure S1 shows the fidelity, defined in the paper, of the reconstructed state (with respect to the original state) as a function of the sparsity of the signal, which is the number of nonzero elements in x . The fidelity decreases as the sparsity increases, which indicates that the sparsity indeed plays a role in the recovery, even if only implicitly.

The second method was developed to address the sparsity specifically. Since the problem (S4) is combinatorically hard to solve, a common solution is to replace the l_0 norm with a convex surrogate function which promotes sparse solutions, most commonly the l_1 norm. However, our sought vector is a probability vector, which means that every feasible solution has l_1 norm of unity. One way to overcome this is by employing a weighted [3] l_1 criterion. Another approach is to consider iterative methods seeking the correct support step by step. Following this latter path, we developed an iterative method based on Orthogonal Matching Pursuit (OMP) [4] adapted to our context.

The algorithm has two main iteration steps: in the first one the support, which is the set of indices on which the solution is nonzero,

Figure S1: The fidelity of the recovery for the least-squares based method (blue) and the method taking advantage of the sparsity explicitly (magenta). The fidelity is averaged over 200 realizations of the signal in a system of 15 waveguides, where the quantum state is subject to depolarization noise of 2% and the measurements are subject to Gaussian noise of 30 dB.

constantly grows by one element. In the second step the size of the support is kept constant, however its elements are changed. In both stages, the minimization problem being solved iteratively on a changing support is

$$\begin{aligned} & \min_{x_S} \|A_S x_S - y\|_2 \\ & \text{s. t. } 0 \leq x_S^i \leq 1 \\ & \sum_i x_S^i = 1. \end{aligned} \quad (S6)$$

Here, the subscript S indicates indices from the current support. The method is described in detail in Algorithm 1.

We begin by describing the first part, which ends when the proposed solution ceases to change significantly. This step is basically the OMP algorithm [4] with two main modifications. The first is the addition of the constraints. The second modification, inspired by the algorithm in [5], is related to the rule governing the increase in the size of the support. Due to the constraints, the residual $r = y - Ax$ is no longer orthogonal to the columns of A already in the support, as in OMP. Thus, when searching for the best candidate column to be added to the support, the columns not in the support are projected onto the subspace orthogonal to the one spanned by the columns in the support.

In each iteration we try to add a column of A to the support, multiplied by a positive scalar. The optimal scalar is achieved by minimizing the objective

$$\begin{aligned} c_i^* &= \arg \min_c \|(I - P)a_i c - r\|_2 \\ &= \frac{[(I - P)a_i]^\dagger r}{\|(I - P)a_i\|_2^2}. \end{aligned} \quad (S7)$$

Here, P is the projection onto the subspace spanned by the columns in the support, thus $I - P$ is the projection onto the orthogonal subspace, and a_i is the i th column of the matrix A . Using this scalar, one for each column not currently part of the support, we calculate the change in the objective for each column added:

$$\alpha_j = \frac{1}{\|(I - P)a_j\|_2^2} \left([(I - P)a_j]^\dagger r \right)^2. \quad (S8)$$

Next, the best column according to these values is added to the support and we solve the constrained least-squares problem on the updated support.

In the second stage, we keep the support size fixed however the support itself changes. In the spirit of [6], we try to find a better

support by removing one element from the current solution and adding the best possible element somewhere else not currently in the support. Due to the constraints, this dictates that the new element has exactly the size of the one we have removed. The suggested change in support results in the following value of the objective

$$\begin{aligned} f(x - x_i e_i + x_i e_j) &= \|A(x - x_i e_i + x_i e_j) - b\|_2^2 \quad (S9) \\ &= \|Ax - b - x_i a_i + x_i a_j\|_2^2 \\ &= \|r - x_i(a_i - a_j)\|_2^2 \\ &= \|r\|_2^2 + x_i(a_i - a_j)^\dagger [x_i(a_i - a_j) - 2r]. \end{aligned}$$

This is an improvement over the current solution (objective value $\|r\|_2^2$) if $x_i(a_i - a_j)^\dagger [x_i(a_i - a_j) - 2r] \leq 0$.

Therefore, at each iteration we find the best pair (i_0, j_0) such that $i_0 \in \mathcal{S}, j_0 \notin \mathcal{S}$ and $e = x_{i_0}(a_{i_0} - a_{j_0})^\dagger [x_{i_0}(a_{i_0} - a_{j_0}) - 2r]$ is minimal. For that pair we check whether $e < 0$ (and also that $|e|$ is not too small, in order to avoid circular changes). In the case of $e < 0$, we add j_0 to the support, remove i_0 from it and solve the constrained least-squares problem on the current support. Otherwise, no improvement is achieved by changing the support, and we stop.

As evident in Fig. S1, both methods achieve similar performance, with the technique using sparsity explicitly demonstrating slightly better results under these conditions.

Algorithm S1. Sparse Probability Vector Recovery

Input: $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, \epsilon_0 \in \mathbb{R}$

Goal: Recover a sparse vector $x \in \mathbb{R}^n$ satisfying $Ax \sim b, \|x\|_1 = 1$ and $0 \leq x_i \leq 1$.

Output: $x \in \mathbb{R}^n$

Initialization: $x = 0, r = b, \mathcal{S} = \emptyset, P = 0$ ($P \in \mathbb{R}^{m \times m}$)

While $\frac{\|x - x_{prev}\|_2}{\|x\|_2} > \epsilon_0$

$$I = \{i \mid [(I - P)a_i]^\dagger r \geq 0\}$$

$$\alpha_j = \frac{1}{\|(I - P)a_j\|_2^2} \left([(I - P)a_j]^\dagger r \right)^2$$

$$j_0 = \arg \max_j \alpha_j \text{ such that } j_0 \in I, j_0 \notin \mathcal{S}$$

$$\mathcal{S} \rightarrow \mathcal{S} \cup \{j_0\}$$

$$P = A_{\mathcal{S}}(A_{\mathcal{S}}^T A_{\mathcal{S}})^{-1} A_{\mathcal{S}}^T$$

$$\text{Solve } \min_{x_{\mathcal{S}}} \|A_{\mathcal{S}} x_{\mathcal{S}} - b\|_2^2 \text{ s.t. } \|x_{\mathcal{S}}\|_1 = 1, 0 \leq x_{\mathcal{S}}^i \leq 1$$

$$\text{to obtain } x_r. \quad (S10)$$

$$x_{prev} = x, x = x_r, r = b - A_{\mathcal{S}} x_{\mathcal{S}} \quad (S11)$$

end

$$f_{ij} = x_i(a_i - a_j)^\dagger (a_i - a_j - 2r) \text{ for } i \in \mathcal{S}, j \notin \mathcal{S}$$

$$i_0, j_0 = \arg \min_{i,j} f_{ij}$$

While $f_{i_0, j_0} < 0, |f_{i_0, j_0}| > THRESH$

$$\mathcal{S} \rightarrow (\mathcal{S} \setminus \{i_0\}) \cup \{j_0\}$$

Solve (S10)

Repeat (S11)

$$f_{ij} = x_i(a_i - a_j)^\dagger (a_i - a_j - 2r) \text{ for } i \in \mathcal{S}, j \notin \mathcal{S}$$

$$i_0, j_0 = \arg \min_{i,j} f_{ij}$$

end

Addressing Degeneracies

In the section presenting the sparsity-based recovery of entangled states, we noted that degeneracies arise when the state is sparse in certain bases. The degeneracy occurs when dealing with an entangled basis of the form described in the paper.

This happens only when the difference between the waveguides divided between the two parties is an odd number, for example, $|\psi\rangle = \frac{1}{\sqrt{2}} [|2_4 1_7\rangle + |1_4 2_7\rangle]$, $|\psi^\perp\rangle = \frac{1}{\sqrt{2}} [|2_4 1_7\rangle - |1_4 2_7\rangle]$. When the difference is odd, the shift invariance of the waveguide system (up to reflections from the boundaries, which do not break this invariance strongly enough), we end up with two identical columns in the matrix M in Eq. 6 in the paper. This degeneracy is demonstrated in Fig. S2. In the non-degenerate case, the recovery probability is high (blue), whereas in the degenerate case it drops when we reach sparsity of 2 (green). This degeneracy can be removed by using a more complex coupler between the modes. As demonstrated in Fig. S2, when we replace the coupled waveguide array with a random linear optical network, represented by a random unitary matrix, the recovery probability in the degenerate scenario becomes high again (magenta) as in the non-degenerate case.

Figure S2: Recovery probability of an entangled state in a basis that includes entangled states but does not lead to degeneracies (blue), in a basis which leads to degeneracies (green), and in a physical setting that removes the degeneracies – by using a random linear optical network as a coupler instead of the waveguide array (magenta).

Recovering N-photon States from N-1 Coincidence Measurements

In the paper, we propose a new approach for recovering three-photon states from two-fold coincidences, demonstrate the method in examples and examine its performance. However, in principle, this idea can be generalized to the recovery of N -photon states from $N - 1$ correlation measurements. Obviously, the $N - 1$ correlation measurements do not fully characterize the N -photon state, hence the measurements are related to the state via an underdetermined system of linear equations similar to Eq. 6 in the paper.

The dimensions of the objects in question (the signal and the measurement vectors) vary similarly with the number of waveguides. For example, for a three-photon state and two-fold correlations, the ratio of the number of measurements to the length of the coefficient vector, which is less than unity for an underdetermined system, is $\frac{3}{N_w + 2}$, whereas for a four-photon state and three-fold correlations it is $\frac{4}{N_w + 3}$. As such, the methodology of sparsity-based recovery in this case is virtually identical to the problems we solve in the paper: the only difference is in the value of

the coefficients in the ratio, but the ratio grows like $\frac{a}{N_W + b}$, where a, b are some integers.

The ability to recover the original vector depends on the properties of the matrix M appearing in Eq. 6 as well. The matrix elements of the $N - 1$ correlation measurement are of the form $M_{ij} = \langle \psi_j | \underbrace{a^\dagger \cdots a^\dagger}_{N-1} \underbrace{a \cdots a}_{N-1} | \psi_i \rangle$, whereas for two-fold correlation they are of the form $M_{ij} = \langle \psi_j | a^\dagger a^\dagger a a | \psi_i \rangle$, with the creation and annihilation operators evaluated at the output facet of the array in the same manner $a_k^\dagger(z) = \sum U_{kl} a_k^\dagger(z)$.

It is therefore likely that the sparsity-based approach will work for other values of N , although further studies are required for investigating the performance as a function of N .

Photon Number Mismatch

In this section, we address the question of photon number mismatch in the input state, where the total number of photons is not known exactly. Up to this point, the sought state was assumed to have a fixed (known) photon number, namely, three. However, as some photon sources may yield a state with a varying number of photons, it is worthwhile to relax this assumption. We investigate this path by simulating input states with 0.9 probability to have three photons and 0.1 probability to have two or four photons. As shown in Fig. S3, when taking a random linear coupler (as explained in Fig. S2), the recovery fidelity is higher than 0.9 for an initial state which is adequately sparse. Thus, our approach is able to handle states with some uncertainty in the photon number.

Figure S3: The recovery fidelity in a scenario of 15 waveguides, measurement noise of 30 dB and a random linear coupler between the input and output facets. Here, the fidelity is averaged over 100 realizations of the input state. The initial state consisted of 3 photons with probability 0.9, and of 2 or 4 photons with probability 0.1.

References

1. Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications* (Cambridge University Press, 2012).
2. A. Bruckstein, M. Elad, and M. Zibulevsky, "On the Uniqueness of Nonnegative Sparse Solutions to Underdetermined Systems of Equations," *IEEE Transactions on Information Theory* **54**, 4813–4820 (2008).
3. E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing Sparsity by Reweighted ℓ_1 Minimization," *J Fourier Anal Appl* **14**, 877–905 (2008).
4. Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993* (1993), pp. 40–44 vol.1.
5. B. Natarajan, "Sparse Approximate Solutions to Linear Systems," *SIAM J. Comput.* **24**, 227–234 (1995).
6. A. Beck and Y. C. Eldar, "Sparsity Constrained Nonlinear Optimization: Optimality Conditions and Algorithms," *SIAM J. Optim.* **23**, 1480–1509 (2013).