Granger Causality for Compressively Sensed Sparse Signals

Aditi Kathpalia
Department of Complex Systems,
Institute of Computer Science of the Czech Academy of Sciences,
Prague, Czech Republic

Nithin Nagaraj
Consciousness Studies Programme,
National Institute of Advanced Studies,
Bengaluru, India
(Dated: October 21, 2022)

Compressed sensing is a scheme that allows for sparse signals to be acquired, transmitted and stored using far fewer measurements than done by conventional means employing Nyquist sampling theorem. Since many naturally occurring signals are sparse (in some domain), compressed sensing has rapidly seen popularity in a number of applied physics and engineering applications, particularly in designing signal and image acquisition strategies, e.g., magnetic resonance imaging, quantum state tomography, scanning tunneling microscopy, analog to digital conversion technologies. Contemporaneously, causal inference has become an important tool for the analysis and understanding of processes and their interactions in many disciplines of science, especially those dealing with complex systems. Direct causal analysis for compressively sensed data is required to avoid the task of reconstructing the compressed data. Also, for some sparse signals, such as for sparse temporal data, it may be difficult to discover causal relations directly using available data-driven/ model-free causality estimation techniques. In this work, we provide a mathematical proof that structured compressed sensing matrices, specifically Circulant and Toeplitz, preserve causal relationships in the compressed signal domain, as measured by Granger Causality. We then verify this theorem on a number of bivariate and multivariate coupled sparse signal simulations which are compressed using these matrices. We also demonstrate a real world application of network causal connectivity estimation from sparse neural spike train recordings from rat prefrontal cortex.

I. INTRODUCTION

The use of compressed (or compressive) sensing or sampling for acquisition of signals has become very popular in the last two decades. The basic idea behind the technique is the acquisition and compression of data at the same time. Since most real-world signals are sparse in some domain, they can be acquired with only a few measurements. These measurements contain sufficient information about the signals, making their perfect reconstruction possible (assuming sparsity) [1, 2].

Compressed sensing has found applications in designing magnetic resonance imaging techniques [3, 4], cameras [5], analog to digital conversion techniques [2], ghost imaging [6, 7], nuclear magnetic resonance spectroscopy [8, 9], quantum state tomography [10, 11], radio interferometry based black hole observation [12, 13] and scanning tunneling microscopy [14]. Since large number of signals are being acquired and stored in the compressed domain, it becomes imperative to analyze the properties of these signals and study their interdependence in the compressed domain itself, avoiding the cumbersome task of their reconstruction. Hence, methods have been proposed to perform many important computations in the domain of signal processing and learning on compressed low dimensional data directly. These computations include tasks such as regression [15], classification [16–19], signal detection [20], nearest neighbor finding [21] and manifold learning [22].

Causality analysis techniques study interactions between different variables of a complex system and provides richer information than mere linear correlation based analysis. This information is required by many disciplines of science in order to study specific systems based on cause-effect relationships between observables and eventually control their required parts, where need be [23, 24]. Neuroscience [25–26], econometrics [32–35], earth sciences [27–31], engineering [36] etc., have employed different time series causality estimation techniques to study interactions in systems such as the human brain, stock prices, earth’s climate.

Causality estimation for time series data began with the pioneering work of Granger [37]. Granger Causality (GC) continues to be the most widely used method for causality estimation till date. GC is inspired from Wiener’s idea of causal prediction, according to which “a time series $X$ causes a time series $Z$, if the past values of $X$ contain information that help predict $Z$ above and beyond the information contained in the past values of $Z$ alone”. To estimate GC from $X$ to $Z$, $Z$ is modelled as...
a linear autoregressive process in two independent ways:

\[
Z(t) = \sum_{\tau=1}^{\infty} (a_{\tau} Z(t-\tau)) + \sum_{\tau=1}^{\infty} (c_{\tau} X(t-\tau)) + \varepsilon_c,
\]

where \(t\) denotes any time instance, \(a_{\tau}, b_{\tau}, c_{\tau}\) are coefficients at a time lag of \(\tau\) and \(\varepsilon_c, \varepsilon\) are error terms (gaussian distributed) in the two models. To determine if \(X\) causes \(Z\) or not, logarithm of the ratio of the prediction error variances is computed:

\[
F_{X \rightarrow Y} = \ln \frac{\text{var}(\varepsilon)}{\text{var}(\varepsilon_c)},
\]

This measure is called the GC F-statistic. If the model represented by Eq. \((1)\) is a better model for \(Z(t)\) than Eq. \((2)\), then \(\text{var}(\varepsilon_c) < \text{var}(\varepsilon)\) and \(F_{X \rightarrow Y}\) will be greater than 0, suggesting that \(X\) Granger Causes \(Z\). In spite of using a simplistic autoregressive model, GC is applicable (in principle) in a wide range of covariance stationary processes [38–39].

A number of attempts have been made to generalize GC to the nonlinear case, such as, by using, an estimator based on correlation integral [53], a non-parametric regression approach [40], local linear predictors [41], mutual nearest neighbors [42–43], kernel estimators [44], just to name a few. Other causality estimation techniques have also been proposed based on the Wiener’s idea of causation. These include Transfer Entropy [45], Conditional Mutual Information [46], Compression-Complexity Causality [47], and Direct Transfer Function [48]. All these techniques are largely model-free/non-parametric and do not require an explicit assumption of the underlying causal-model. Other than these, there are other techniques which are model-based, which require an explicit assumption of the underlying physical model generating the data. Owing to these assumptions, these methods are often specific to certain systems in specific application domains. Some examples include Dynamic Causal Modeling [49], which was developed for neuroscience and Structural Equation Modeling [50], which has found applications in medicine, engineering and social sciences.

In today’s world of data-driven science, both compressed sensing and causality estimation have gained impressive ground across disciplines, and play an important role in experimental [10, 51–54] and computational [55–57] sciences as well as the study of complex systems and networks [25, 28, 58–61]. By saving measurement time and cost, compressed sensing has enabled experiments to be conducted more efficiently. Causality estimation, on the other hand, has supplied tools to study interactions in systems and unravel their underlying dynamics. The motivation for this study lies at the intersection of these two disciplines. Specifically, we wish to extend causal estimation to the compressed domain. It is often the case that causal network discovery is required to be done on high dimensional data having large number of variables. Application of causality estimation techniques to compressed signals will prevent the computationally expensive task of reconstructing signals acquired using compressive sensing. This would enable a fast and reliable causal discovery, helping in compressed computation, which has gained importance in recent times. In the following subsections, we discuss the technique of compressed sensing and the requirement for causal inference in the compressive domain in some detail.

### A. Compressed Sensing

Compressed sensing is a relatively recent signal processing technique that helps in the acquisition, transmission, storage as well as modification of large amounts of data in an efficient manner. It is well-known that when a signal is sampled at the Nyquist rate, perfect recovery of the original signal is guaranteed. However, the theory of compressed sensing claims that signal recovery is possible by using far fewer measurements under certain conditions. There are two conditions required for this: (1) the input signal be sparse when expressed in a proper basis, and (2) the basis in which it is sensed/acquired be incoherent with the above mentioned sparsifying basis [2, 62]. Sparsity refers to the case in which most entries of a discrete-time signal/ array are zero. If a signal has ‘\(k\)’ non-zero entries, then \(k\) is much less than the total length of the signal. This means that, when expressed in the proper basis \(\psi\), which is called the sparsifying matrix, these signals have concise representations. Incoherence is basically the idea that objects that have a sparse representation in \(\psi\) should spread out in the domain in which they are acquired (or sensed). This is analogous to the property of duality – an impulse in the time domain is spread out in the frequency domain (and vice versa). In other words, incoherence means that the sampling/sensing waveforms have a very dense representation in \(\psi\).

Let \(y\) (an \(m \times 1\) dimension vector) be a set of measurements acquired for a sparse signal \(z\) (of dimension \(n \times 1\)). \(z\) could be itself sparse or could be expressed in the form \(\psi \alpha\), where \(\psi\) is the sparsifying matrix and \(\alpha\) contains the sparse coefficients. \(y\) is called the compressed signal or compressive measurement vector. Then, sparse signal recovery can be seen as a compressed sensing problem as follows:

\[
y = \phi z,
\]

where \(\phi\) is an \(m \times n\) sensing matrix. The input vector \(z\) is in \(\mathbb{R}^n\) and is mapped to the output vector \(y\), which is in \(\mathbb{R}^m\). The sensing matrix, \(\phi\), has a rank \(\leq m\) (and \(m \ll n\)).

\(\phi\) must satisfy, \(\phi z_1 \neq \phi z_2\) for all \(k\)-sparse vectors \(z_1 \neq z_2\). Here, \(k\)-sparse vectors are those which contain
a maximum number of $k$ non-zero entries. This property allows the mapping to be invertible on all $k$-sparse vectors, ensuring that no two output vectors are mapped to the exact same sparse vector during recovery. For this to be the case, it is required that $\phi$ has at least $2k$ rows, i.e. $m \geq 2k$.

**Restricted Isometry Property (RIP)** is a key notion in compressed sensing that determines the efficiency and robustness of the sensing matrix $\phi$ to capture information about sparse signals, such that the aim of recovering them is satisfactorily fulfilled. For each integer, $k = 1, 2, \ldots$, let $\delta_k$ be the RIP/isometry constant of matrix $\phi$ associated with that $k$. Then, $\delta_k$ is defined as the smallest positive value such that

\[
(1 - \delta_k)\|z\|_2^2 \leq \|\phi z\|_2^2 \leq (1 + \delta_k)\|z\|_2^2,
\]

holds for all $k$-sparse vectors $z$. $\phi$ is said to fulfill an RIP of order $k$, if $\delta_k$ is not too close to one. This implies that for $k$-sparse signals, the euclidean length is roughly preserved, and so these $k$-sparse vectors cannot be in the null space of $\phi$. Consequently, this ensures that their reconstruction is possible.

Even when $\phi$ satisfies the above conditions, Eq. 4 still remains under-determined with infinitely many solutions. The best (sparse) solution can be recovered from the compressed measurement vector $y$, using a variety of algorithms, which employ, for example, optimization techniques (such as $l_1$ minimization [1, 65]) or greedy approach (such as Orthogonal Matching Pursuit [66]).

**B. Causal Inference in the Compressive Domain**

As compressed sensing schemes are being widely employed for the acquisition of signals, their transmission and storage, it is important to know whether causal relationships between different variables in the compressed domain can be inferred directly from the compressed data, that is without reconstructing the signals. Also, it is necessary to know, under what circumstances and for what kind of sensing matrices are causal relationships preserved in the compressed domain.

Further, sometimes, in case of sparse data, available techniques for model-free causality estimation may not perform adequately when applied directly to the original data (uncompressed). For example, Granger Causality cannot be computed directly for point processes and only spectral GC estimated from fourier transformed stationary point processes is meaningful [67]. Another way that has been explored for such cases is converting the point process data to continuous domain through the use of smoothing kernels [68] or low pass filters [48, 69] and then applying GC or one of its extensions. Other than this, GC in combination with machine learning algorithms has also been used to learn causality for certain kinds of point processes [70]. Transfer entropy (in its original discrete time form), when applied to point processes or event-based data, has also been shown to have a number of drawbacks and hence continuous time Transfer Entropy has been developed and proposed as an alternative [71, 72]. However, there are no works validating the use of model-free causal estimation techniques, generally on sparse processes or compressed representations of sparse processes. There are some specialized model-based methods that have been proposed to be applied to specific kinds of sparse processes. For example, point process generalized linear model (GLM) based techniques [73–75], MVAR nonlinear Poisson model [76], dynamic Bayesian network [80] and Cox model [81] have been proposed for effective connectivity analysis of neural spike train data.

In this work, we discuss, if and how the use of Granger causality can be done in order to discover causal relations for sparse data by using them directly as acquired in the compressive domain or transforming them into compressed domain. Specifically we show that structured compressed sensing matrices, circulant and toepitz, when used to directly sense or (posteriorly) compress sparse signals preserve GC between these signals in the compressed domain. To the best of our knowledge, both our problem statement and the solution proposed are novel and no work has previously been done in this direction.

When the connections in the network are sparse, there exist techniques such as Lasso Granger [82], its several variations [83, 84], CaSPiAN [85] etc., that employ a combination of compressed sensing techniques (such as variable (model) selection, for example, Lasso) and causality estimation methods (such as GC). These are methods to infer network connectivity and are of use for study of systems such as Genetic Regulatory Networks [86]. These methods are not to be confused with the work that we present in this paper since none of these methods attempt to infer causal relationships in the compressed domain.

This paper is organized as follows. In Section II we provide a mathematical proof that circulant and toepitz structured sensing matrices preserve GC for compressively sensed signals. In Section III we demonstrate the performance of these sensing matrices by discovering causality from signals sensed using these matrices. This is done for the case of bivariate and multivariate simulations of sparse signals. We also apply the matrices to estimate GC from compressed counterparts of neural spike trains recorded from rat prefrontal cortex. We discuss our results and conclude in Section IV.

II. SPECIAL SENSING MATRICES THAT PRESERVE GRANGER CAUSALITY

As discussed before, a sensing matrix, $\phi$, on multiplication with an $n \times 1$ sparse signal, $z$, yields a compressed signal, $y$, of size $m \times 1$. This is according to the equation, $y = \phi z$. Gaussian random values in $\phi$ prove to be one of the most effective ways in helping recover $z$.
from $y$. However, one of the drawbacks of using a random matrix is that hardly any properties of the original signal are preserved after compression. Hence, it was intuitive to explore matrices which possess some kind of structure and also satisfy the essential properties for sparse signal recovery, so that no information is lost from the signals.

Circulant and Toeplitz matrices have been shown to be effective sensing matrices for recovery of sparse signals. In fact, Toeplitz and partial random circulant matrices have both been shown to satisfy the RIP property. In this section we provide complete mathematical proofs showing that signals sensed using either of these matrices are effective at preserving causality as measured by GC.

A. Circulant Sensing Matrices

Let $C$ be a circulant matrix of order $n \times n$, composed of random entries $a_0, a_1, a_2, \ldots, a_{n-1}$, where $a_i \sim \mathcal{N}(0, 1)$ (Gaussian distribution with zero mean and unit variance):

$$C = \begin{bmatrix} a_0 & a_{n-1} & a_{n-2} & \cdots & a_1 \\ a_1 & a_0 & a_{n-1} & \cdots & a_2 \\ a_2 & a_1 & a_0 & \cdots & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{bmatrix}_{n \times n}. \quad (6)$$

It is known that a circulant matrix is diagonalizable by a Discrete Fourier Transform (DFT) matrix. Let $F_n$ denote the DFT matrix:

$$F_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)^2} \end{bmatrix}_{n \times n}. \quad (7)$$

So,

$$C_n = (F_n^{-1})^{T}DF_n, \quad (8)$$

where, $D$ is a diagonal matrix formed by the entries $D_{k,k} = G_k$, where $G_k$ is the $k^{th}$ coefficient of the DFT of $\hat{a} = [a_0, a_1, \ldots, a_{n-1}]$. Let, $z$ (of dimension $n \times 1$), when multiplied by $C$ return:

$$y = Cz \quad (9)$$

From Eq. 8,

$$Cz = (F_n^{-1})^{T}DF_nz, = F_n^{-1}D(F_nz). \quad (10)$$

So, if $Z(f_i)$ denotes the Fourier coefficient of $z$ at frequency $f_i$,

$$Cz = F_n^{-1} \begin{bmatrix} G_0 & 0 & 0 & \cdots & 0 \\ G_1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & G_{n-1} \end{bmatrix} \begin{bmatrix} Z(f_0) \\ Z(f_1) \\ \vdots \\ Z(f_{n-1}) \end{bmatrix}, \quad (11)$$

Multiplying eqn. 9 with a projection matrix $P$ of order $m \times n$ (with $m < n$) on both sides.

$$Py = PCz, \quad (12)$$

where,

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{m \times n}. \quad (13)$$

This is equivalent to multiplying $z$ with an $m \times n$ compressed sensing matrix $PC = C_{m \times n}$, with circulant structure. From equations 12 and 11

$$Py = PF_n^{-1} \begin{bmatrix} G_0Z(f_0) \\ G_1Z(f_1) \\ \vdots \\ G_{n-1}Z(f_{n-1}) \end{bmatrix}, \quad (14)$$

where,

$$F_n^{-1} = \begin{bmatrix} 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)^2} \end{bmatrix}_{m \times n}. \quad (15)$$

Let $Py = y'$, where $y'$ is an $m \times 1$ vector. $y'$ is essentially a compressively sensed vector. Multiplying 14 on both sides with Fourier transform matrix of order $m \times m$,
\[ F_{m \gamma} = F_m F_m^{-1} \begin{bmatrix} G_0 Z(f_0) \\ G_1 Z(f_1) \\ \vdots \\ G_{n-1} Z(f_{n-1}) \end{bmatrix} = P_{m \times n} \begin{bmatrix} G_0 Z(f_0) \\ G_1 Z(f_1) \\ \vdots \\ G_{n-1} Z(f_{n-1}) \end{bmatrix}, \] (16)
yields,

\[ \begin{bmatrix} Y'(f_0) \\ Y'(f_1) \\ \vdots \\ Y'(f_{m-1}) \\ S_{y'y'}(f_0) \\ S_{y'y'}(f_1) \\ \vdots \\ S_{y'y'}(f_{n-1}) \end{bmatrix} = \begin{bmatrix} G_0 Z(f_0) \\ G_1 Z(f_1) \\ \vdots \\ G_{m-1} Z(f_{m-1}) \\ G_0^2 S_{zz}(f_0) \\ G_1^2 S_{zz}(f_1) \\ \vdots \\ G_{m-1}^2 S_{zz}(f_{m-1}) \end{bmatrix}, \] (17)

Here, \( Y'(f_0), Y'(f_1), \ldots \) are the Fourier coefficients of \( y' \). Squaring the equation on both sides, gives a relation between the spectral power coefficients of \( y' \) and \( z \). \( S_{y'y'} \) and \( S_{zz} \) represent the spectral coefficients of \( y' \) and \( z \), respectively.

For two processes \( z_1 \) and \( z_2 \), which may be coupled and can be modeled like autoregressive processes,

\[
z_1(t) = \sum_{j=1}^{\infty} b_{11,j} z_1(t-j) + \sum_{j=1}^{\infty} b_{12,j} z_2(t-j) + \varepsilon_1(t),
\]
\[
z_2(t) = \sum_{j=1}^{\infty} b_{21,j} z_1(t-j) + \sum_{j=1}^{\infty} b_{22,j} z_2(t-j) + \varepsilon_2(t)
\] (18)

Rewriting the above equation in terms of the lag operator,

\[
\begin{bmatrix} b_{11}(L) & b_{12}(L) \\ b_{21}(L) & b_{22}(L) \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},
\] (19)

where, \( b_{ij}(L) = \sum_{k=0}^{\infty} b_{ij,k} L^k \), with \( b_{ij,0} = \delta_{ij} \) or the Kronecker delta function. The covariance matrix of noise terms is:

\[
\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},
\] (20)

where, \( \Sigma_{11} = \text{var}(\varepsilon_1), \Sigma_{22} = \text{var}(\varepsilon_2) \) and \( \Sigma_{12} = \Sigma_{21} = \text{cov}(\varepsilon_1, \varepsilon_2) \).

Taking Fourier transform of Eq. (19) on both sides:

\[
\begin{bmatrix} B_{11}(f) & B_{12}(f) \\ B_{21}(f) & B_{22}(f) \end{bmatrix} \begin{bmatrix} Z_1(f) \\ Z_2(f) \end{bmatrix} = \begin{bmatrix} E_1(f) \\ E_2(f) \end{bmatrix},
\] (21)

In terms of transfer function matrix \( H(f) = [B_{ij}(f)]^{-1}, \) Eq. (21) can be expressed as:

\[
\begin{bmatrix} Z_1(f) \\ Z_2(f) \end{bmatrix} = \begin{bmatrix} H_{11}(f) & H_{12}(f) \\ H_{21}(f) & H_{22}(f) \end{bmatrix} \begin{bmatrix} E_1(f) \\ E_2(f) \end{bmatrix}.
\] (22)

The spectral density matrix for the above is given by:

\[
S(f) = H(f) \Sigma H^*(f),
\] (23)

where * denotes matrix adjoint. To check the causal influence from \( z_2 \) to \( z_1 \) using spectral Granger Causality, we need to look at the autospectrum of \( z_1 \), which is given by:

\[
S_{11}(f) = H_{11}(f)\Sigma_{11}H_{11}^*(f) + 2\Sigma_{12}\text{Re}(H_{11}H_{12}^*) + H_{12}(f)\Sigma_{22}H_{12}^*(f).
\] (24)

In the above expression, due to the cross terms, causal power contribution to \( z_1 \) is not explicit. Geweke, in his work [31], came up with a transformation that makes the causal power term and the intrinsic term explicit. To obtain this transformation for \( z_1 \), Eq. (21) is left multiplied on both sides by:

\[
\begin{bmatrix} 1 \\ -\Sigma_{12}/\Sigma_{11} \end{bmatrix}.
\]

This gives,

\[
\begin{bmatrix} B_{11}(f) & B_{12}(f) \\ B_{21}(f) & B_{22}(f) \end{bmatrix} \begin{bmatrix} Z_1(f) \\ Z_2(f) \end{bmatrix} = \begin{bmatrix} E_1(f) \\ E_2(f) \end{bmatrix},
\] (25)

where, \( \tilde{B}_{21}(f) = B_{21}(f) - \frac{\Sigma_{12}}{\Sigma_{11}}B_{11}(f), \) \( \tilde{B}_{22}(f) = B_{22}(f) - \frac{\Sigma_{12}}{\Sigma_{11}}B_{12}(f). \) Let the new transfer function be given by \( \tilde{H}(f) \), whose elements then become \( \tilde{H}_{11}(f) = H_{11}(f) + \frac{\Sigma_{12}}{\Sigma_{11}}H_{12}(f), \) \( \tilde{H}_{12}(f) = H_{12}(f), \) \( \tilde{H}_{21}(f) = H_{21}(f) + \frac{\Sigma_{12}}{\Sigma_{11}}H_{22}(f) \) and \( \tilde{H}_{22}(f) = H_{22}(f). \) \( \tilde{E}_2(f) = E_2(f) - \frac{\Sigma_{12}}{\Sigma_{11}}E_1(f) \) which results in the cov\( (E_1, \tilde{E}_2) = 0 \) and \( \tilde{S}_{22} \), the new variance of the error term of \( z_2 \), is

\[
\Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}}.
\]

As \( \Sigma_{12} \) becomes zero, the autospectrum of \( z_1(t) \) is now decomposed into two explicit parts:

\[
\tilde{S}_{11}(f) = \tilde{H}_{11}(f)\Sigma_{11}\tilde{H}_{11}^*(f) + \tilde{H}_{12}(f)\tilde{S}_{22}\tilde{H}_{12}^*(f).
\] (26)

The first term in the decomposition of \( \tilde{S}_{11}(f) \) represents an intrinsic power term of \( z_1(t) \) and the second is due to the ‘causal power’ contribution from \( z_2(t) \). Spectral Granger Causality at a particular frequency \( f \) is given by the logarithm of the ratio of total power to intrinsic power, so GC from \( z_2 \) to \( z_1 \) at \( f \):

\[
I_{z_2 \rightarrow z_1}(f) = \ln \frac{\tilde{S}_{11}(f)}{\tilde{S}_{11}(f) - (\Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}})H_{12}(f)^2}.
\] (27)

Multiplying Eq. (21) with a constant, \( G_f \), on both sides:

\[
\begin{bmatrix} B_{11}(f) & B_{12}(f) \\ B_{21}(f) & B_{22}(f) \end{bmatrix} \begin{bmatrix} G_f Z_1(f) \\ G_f Z_2(f) \end{bmatrix} = \begin{bmatrix} G_f E_1(f) \\ G_f E_2(f) \end{bmatrix}.
\] (28)
This gives (using Eq. (17)),
\[
\begin{bmatrix}
B_{11}(f) & B_{12}(f) \\
B_{21}(f) & B_{22}(f)
\end{bmatrix}
\begin{bmatrix}
Y'_1(f) \\
Y'_2(f)
\end{bmatrix}
= \begin{bmatrix}
G_1 E_1(f) \\
G_1 E_2(f)
\end{bmatrix}.
\tag{29}
\]

The above gives us a relation between the frequency components of \(y'_1\) and \(y'_2\) which can be thought of as being compressed counterparts of \(z_1\) and \(z_2\). If \(z_1\) and \(z_2\) are compressed by the same compressed sensing matrix, their coefficients corresponding to a particular frequency will be scaled by the same constant. This follows directly from Eq. (17).

Hence, the transformation of \(z_1\) to \(y'_1\) and of \(z_2\) to \(y'_2\), changes only the variance of \(z_1\), \(z_2\), basically scaling \(E_1(f)\) and \(E_2(f)\) by the same quantity. Thus the variance and covariance terms of \(E_1\) and \(E_2\) are all scaled by \(G_2^2\). To compute the causal influence from \(y'_2\) to \(y'_1\), the same procedure holds as for \(z_2\) to \(z_1\), with the only difference that both the numerator and the denominator in the logarithmic term of Eq. (27) are multiplied by \(G_2^2\) as all the error variance and covariance terms are scaled by this quantity. As this scaling factor exists both in the numerator and denominator, it gets cancelled out, yielding the same causal influence from \(Y'_2(f)\) to \(Y'_1(f)\) as exists between \(Z_2(f)\) to \(Z_1(f)\). Thus,
\[
I_{y'_2\to y'_1}(f) = I_{z_2\to z_1}(f)
\tag{30}
\]

Geweke [94] showed that for all processes of practical interest,
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} I_{z_2\to z_1}(f) \, df = F_{z_2\to z_1},
\tag{31}
\]
where, \(F_{z_2\to z_1}\) refers to the time-domain GC from \(z_2\) to \(z_1\). Now, since \(z_1\) and \(z_2\) are sparse and all information contained in them is equally well represented in the compressed sensing domain by \(y_1\) and \(y_2\) (invertability/reconstruction of all sparse vectors is guaranteed by the RIP property), all the information including causal information is present in the limited/coarser frequency components of the compressed signal. Hence,
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} I_{y'_2\to y'_1}(f) \, df = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_{y'_2\to y'_1}(f) \, df = F_{y'_2\to y'_1},
\tag{32}
\]
showing that the time-domain GC is preserved.

To find the causal influence from \(z_1\) to \(z_2\) or from \(y'_1\) to \(y'_2\) at any frequency \(f\), one can multiply by the transformation matrix,
\[
\begin{bmatrix}
1 & -\Sigma_{12}/\Sigma_{22} \\
0 & 1
\end{bmatrix},
\]
then applying this to \(x\), to delineate the intrinsic and the causal terms.

**Toeplitz Sensing Matrices**

When our compressed signal is obtained by multiplication with a toeplitz sensing matrix, the following discussion holds. Let us start by considering an \(n \times n\) toeplitz matrix, \(T_n\).

\[
T = \begin{bmatrix}
a_0 & a_{n-1} & a_{n-2} & \cdots & a_1 \\
-1 & a_0 & a_{n-1} & \cdots & a_2 \\
a_{-2} & a_{-1} & a_0 & \cdots & a_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{-(n-1)} & a_{-(n-2)} & a_{-(n-3)} & \cdots & a_0
\end{bmatrix}_{n \times n},
\tag{33}
\]

where, \(a_i \sim \mathcal{N}(0, 1)\). Multiplying \(T\) with a signal \(z\) gives,
\[
y_{n \times 1} = T_{n \times n} z_{n \times 1}
\tag{34}
\]

This multiplication can be achieved by embedding \(T_n\) into a circulant matrix, \(C_{2n}\), of dimension \(2n \times 2n\). This embedding is done in such a manner that the first \(n\) entries of \(C_{2n}z\) will be equal to \(T_nz\):
\[
\begin{bmatrix}
y \\
y_d
\end{bmatrix} = C_{2n} \begin{bmatrix}
z \\
0
\end{bmatrix} = \begin{bmatrix}
T_n & B_n \\
B_n & T_n
\end{bmatrix} \begin{bmatrix}
z \\
0
\end{bmatrix} = \begin{bmatrix}
T_{n \times n} \\
B_{n \times n}
\end{bmatrix}z_{2n \times 1},
\tag{35}
\]

Here, the signal \(z\) is appended with \(n\) zeroes, resulting in a \(2n \times 1\) vector. Once again, this circulant matrix is diagonalizable by a DFT matrix. Then, as in Eqs. (10) and (11),
\[
\begin{bmatrix}
y \\
y_d
\end{bmatrix} = F_{2n}^{-1} D F_{2n} \begin{bmatrix}
z \\
0
\end{bmatrix} = \begin{bmatrix}
G_0 Z(f_0) \\
G_1 Z(f_1) \\
G_2 Z(f_2) \\
\vdots \\
G_{2n-1} Z(f_{2n-1})
\end{bmatrix},
\tag{36}
\]

where, \(Z(f)\) is the Fourier coefficient at frequency \(f\) of \(z\). The \(G\)s here are the coefficients of the DFT of the first column in the entries of \(C_{2n}\), given by, \(\tilde{a}_{2n} = [a_0, a_{-1}, \ldots, a_{-(n-1)}, a_0, a_1, \ldots, a_{(n-1)}]\).

It is known that zero padding a (finite length) signal only increases its frequency resolution without losing any information in the frequency content of the original signal \(z\).

Now, we multiply Eq. (35) on both sides with projection matrix \(P\) of order \(m \times 2n\) (with \(m < n\),
\[
\begin{bmatrix}
y \\
y_d
\end{bmatrix} = P_{m \times 2n} \begin{bmatrix}
T_{n \times n} \\
B_{n \times n}
\end{bmatrix}z_{2n \times 1},
\tag{37}
\]

which can alternatively be written as:
\[
P_{m \times n} y = y' = P_{m \times n} T_n z = T_{m \times n} z,
\tag{38}
\]
as \(m < n\). The above equation can be seen as left multiplying \(z\) with a compressed sensing matrix \(T_{m \times n}\) with toeplitz structure. Going back to the representation of \(y\)
matrices have circulant or toeplitz structure.

Thus, we have proved that GC is preserved if sensing matrices have circulant or toeplitz structure.

III. RESULTS

A. Simulations

Sparse input signals, $z_1$ and $z_2$, were first simulated. These were then passed as an input to the same sensing matrix $C$ in order to obtain compressed signals, $y_1$ and $y_2$ respectively. The length of input signals was kept as $n = 2000$ and that of the compressed signals was kept as $m = 200$. Sparsity of the input signals, $k$, was kept as 20, except in the cases where results were specifically obtained for varying $k$.

Let the signals $z_1$ and $z_2$ be defined over a set $T = \{1, 2, 3, \ldots, N\}$, which can be considered to be a set of time points. Let $T_1 \subset T$ such that $T_1$ consists of $k$ elements chosen from $T$ in a uniformly random manner. Also, let $T_2 \subset T$ be defined as $T_2 = \{t_2 : t_2 = t_1 + 1, \forall t_1 \in T_1\}$. $T_1^c$ and $T_2^c$ denote the complement sets of $T_1$ and $T_2$ respectively ($T$ being the universal set). The signals $z_1$ and $z_2$ were generated by sparsifying a pair of autoregressive processes $Z_1$ and $Z_2$. These were generated as per the following equations:

$$Z_1(t) = \alpha Z_1(t - 1) + \varepsilon_1(t), \quad \forall t \in T,$$

$$z_1(t) = \begin{cases} 
Z_1(t), & \forall \{t \in T_1\}, \\
0, & \forall \{t \in T_1^c\}, 
\end{cases}$$

where $\varepsilon_1$ and $\varepsilon_2$ are independent Gaussian noise drawn from $\mathcal{N}(0, 0.1)$, $\alpha = 0.8$, $\beta = 0.08$ and $\gamma = 0.75$. At $k$ randomly selected time points, $z_1(t)$ takes the values of $Z_1(t)$ and is set to zero elsewhere (at the rest $N - k$ time points). $z_2(t)$ is zero whenever $z_1(t-1) = 0$. At the rest $k$ time points, $z_2(t)$ retains the values of $Z_2(t)$. Defined this way, both $z_1$ and $z_2$ are $k$-sparse signals. As can be seen from the above equations, $z_1$ has a causal influence on $z_2$ (with coupling coefficient $\gamma$, controlling the strength of causation), and there is no causal influence from $z_2$ to $z_1$.

1. Performance of Structured Matrices

Performance of circulant and toeplitz matrices was evaluated for sparse signal recovery and causality estimation, while varying the level of sparsity, $k$. 100 realizations of $z_1$ and $z_2$ with random initial values were simulated. Percentage success in signal reconstruction and correct causality detection is shown in Fig. 1. $k$ was varied from 5 to 50 in steps of 5. CVX, a MATLAB package for convex optimization [95] was used for $l_1$ minimization and subsequent sparse signal reconstruction for all results in this paper. For each realization, sparse signal reconstruction ($z_1$ and $z_2$ from $y_1$ and $y_2$ respectively) is counted as successful if the mean squared error between the original and reconstructed signal is less than $10^{-5}$ for both $z_1$ and $z_2$. The mean squared error of reconstruction for a realization of $z_1$ is estimated as:

$$MSE(z_1) = \frac{1}{N} \sum_{i=1}^{N} (z_1(i) - \hat{z}_1(i))^2,$$  \hspace{1cm} (44)

![FIG. 1. Percentage success (using 100 realizations) in sparse signal reconstruction and causality estimation for varying sparsity ($k$) in case of (a) circulant and (b) toeplitz sensing matrices. Signal recovery degrades for higher values of $k$, however, causality estimation improves.](image-url)
where, \( i \) denotes the temporal index of samples in \( z_1 \), \( N \) is the length of \( z_1 \) and \( \hat{z}_1 \) is the reconstructed signal for \( z_1 \). For each realization causality estimation is counted as being successful if the Granger causality F-statistic from \( y_1 \) to \( y_2 \) is found to be significant and from \( y_2 \) to \( y_1 \) is found to be insignificant. Multivariate Granger Causality (MVGC) toolbox [38] was used for the computation of GC F-statistic and its significance for all experiments in this work. The order of the AR processes was determined using Akaike Information Criterion with maximum number of lags to be considered for the processes set to 30. Significance testing was done using \( \chi^2 \) method with significance level, \( \alpha \), set to 0.01. Rest of the parameters of the toolbox were set to their default values [38]. These settings for the computation of GC and its significance were kept constant for all experiments in this paper.

Percentage success in causality estimation when the level of sparsity is kept constant at \( k = 20 \) but the degree of circulant and toeplitz structure is varied, is shown in Fig. 2. The degree of structure was varied by allowing only the first \( S \leq m \) rows of the sensing matrix to have circulant/toeplitz structure while the entries for the rest of \( m-S \) rows were selected randomly from \( \mathcal{N}(0,1) \). \( S \) was varied from 0 to \( m (=200) \) in steps of 20. For each \( S \), 100 realizations of \( z_1 \) and \( z_2 \) were simulated and compressed to obtain \( y_1 \) and \( y_2 \) respectively (using a different sensing matrix for each realization). Reconstruction of \( z_1 \) and \( z_2 \) was found to be successful for all the realizations for all the values of \( S \) here. Hence, the depiction of percentage success in sparse signal recovery is omitted in Fig. 2.

**FIG. 2.** Percentage success (using 100 realizations) in correct causality estimation for varying number of structured rows, \( S \), in case of circulant and Toeplitz sensing matrices. Causality estimation improves as the number of structured rows increases for both types of matrices.

2. **Varying the Coefficient of Causation**

For simulated input signals \( z_1 \) and \( z_2 \), the coupling coefficient \( \gamma \) (see Eqs. (12) and (13)) was varied from 0 to 4 in steps of 0.5. GC F-statistic was estimated from \( y_1 \) to \( y_2 \) (\( F_1 \)) and from \( y_2 \) to \( y_1 \) (\( F_2 \)) for both fully circulant and toeplitz sensing matrices. In Figs. 3(a) and 3(b), mean values of \( F_1 \) and \( F_2 \), obtained using 100 random realizations are shown for circulant and Toeplitz matrices respectively as the coefficient \( \gamma \) is varied.

**FIG. 3.** Mean Granger Causality values (using 100 realizations) from \( y_1 \rightarrow y_2 \) (\( F_1 \)) and \( y_2 \rightarrow y_1 \) (\( F_2 \)) as the coupling coefficient \( \gamma \) is varied in case of sensing by (a) circulant and (b) toeplitz matrices. With increasing \( \gamma \), strength of GC estimated in the direction of coupling between the compressed signals increases.

3. **Multivariate neural spike trains**

A simple network of simulated spike trains from 10 neurons was generated as in [77]. MATLAB scripts from [77] are openly available and were directly used for the purpose. These spike trains are based on the Generalized Linear Model (GLM). The connectivity matrix for this network is as shown in Fig. 1 (leftmost plot). Excitatory and inhibitory connections are not differentiated in the displayed connectivity matrix. 20 realizations with each of the 10 nodes generating spike trains of length 10000 time points were simulated.

Each of the spike trains were compressed using circulant and Toeplitz sensing matrices with \( m = 2000 \). \( m \) was chosen to be greater than four times the sparsity of the signal with maximum sparsity. Conditional MVGC was estimated using the MVGC toolbox with the parameters of the toolbox set as discussed before. The connectivity matrix recovered from the compressed signal using circulant and Toeplitz sensing matrices is shown in Fig. 4. It is found that both types of matrices perfectly recover all the existing causal connections in the simulated network and do not detect any false positives. Thus, the sensitivity as well as the specificity of the estimated connections is found to be 1.

**B. Real Data**

Sparse neuronal spike train data, recorded from the prefrontal cortex of adult male Sprague-Dawley rats, acquired for the study in [77], are available open source
shows the connectivity matrices obtained by from these neurons. Thus, \( m = 20114 \), and is taken to be 7000 (set to be greater than 2 times the sparsity of the signal having the maximum sparsity).

In [77], a ‘Structural Information Enhanced’ regularization method has been developed in order to aid the GLM framework to better estimate functional connectivity between neurons. This technique is mainly for large sparse spike train datasets. As discussed in the introduction, GLM is a model-based method. Simulation results in [77] indicate that when the parameter selection for GLM is done using the proposed regularized method (abbreviated as SGL), the technique performs better than existing approaches. As a result, the authors display a confidence in functional connectivity estimated in the rats’ prefrontal cortex using the above discussed data recordings. However, even for the simple network simulation of 10 neurons discussed in the previous section, the method does not perfectly estimate the network connections. For the simulation case considered in this manuscript, the best that the method does is to discover an average 7.23 correct connections (that is, true positives) out of the existing 10 connections, using a 100 realizations of the simulation. This result is given in Table 1 of [77]. Our method, on the other hand, recovers the connectivity matrix perfectly (see Fig. 4). It is found that the SGL method has high specificity but low sensitivity. In other words, most truly zero-valued connections are estimated to be zero, however, the sensitivity of detecting significant connections is not as good.

For the real data used in this section, we compare the performance of fully circulant and toeplitz sensing matrices to discover causal connections for the network with the connectivity discovered by SGL method. This was done as there is no ground truth available to compare with. Fig. 5 shows the connectivity matrices obtained by the three methods. Significant connections are displayed as black squares. Any non-zero couplings discovered by SGL are displayed as significant, while conditional GC estimation and its significance detection were done by the MVGC toolbox using the same parameters as discussed for the simulations before. The SGL results are basically a reproduction of the results for ‘Experiment 6’, displayed in Fig. 10 of [77]. True negative rate or specificity of the results using circulant and toeplitz sensing matrices with respect to the SGL results were found to be 0.998 and 0.999 respectively. True positive rate or sensitivity was found to be 0.333 for both circulant and toeplitz sensing. As discussed before, for simulations, SGL showed high specificity but relatively low sensitivity. Thus, we can expect the relative specificity of our method obtained here to be close to the true specificity. On the other hand, we would expect the true sensitivity to be much better than the relative sensitivity obtained here as our method showed a specificity and sensitivity of 1 for the simulation network, which is superior to the performance of SGL for the simulation.

IV. DISCUSSION AND CONCLUSIONS

This work discusses the task of discovering causal connections for signals in the compressed sensing domain. To the best of our knowledge, ours’ is the pioneering study to solve the problem of causality estimation between compressed measurement signals without the need to reconstruct sparse signals. Causal inference is fun-
fundamental to understanding the interaction and underlying dynamics of processes in many disciplines of science. With the widespread use of compressed sensing for acquisition, transfer and storage of signals, it becomes essential that causal analyses be performed in the compressed domain. This will save the computational cost of reconstruction of sparse signals. Furthermore, it is known that for sparse temporal data such as sparse point processes, it is not directly possible to estimate ‘model-free’ causality for time-domain data. For such data as well, causal analysis directly in compressed domain or after transforming sparse signals to compressed domain can be very useful.

We have mathematically proved that structured compressed sensing matrices, specifically circulant and toeplitz matrices preserve GC for sparse signals in the compressed domain. These signals need not necessarily be sparse in the temporal domain and may possess a sparse representation in another domain. However, the processes must be covariance stationary in the temporal domain for GC to be applicable. GC is the oldest and most widely used method for data-driven causality estimation and has been applied in a large number of studies across disciplines. Our proof opens up numerous more useful applications of GC.

To verify the proof and demonstrate its significance in practical applications, the performance of the structured sensing matrices is tested for GC based causality detection using a number of simulation experiments. For two coupled sparse processes, when the amount of signal sparsity $k$ is varied and the structure of sensing matrix is kept as circulant or toeplitz, percentage success in causality estimation increases and reaches 100% at high values of $k$ (> 30). On the other hand, percentage success in signal reconstruction deteriorates with increasing values of $k$. This can be seen from Fig. 1. When the input signal sparsity level is constant, but the number of structured rows in the sensing matrix are increased, then as seen in Fig. 2, the percentage success in causality estimation increases and approaches 100% for fully structured (circulant or toeplitz) matrix. This illustrates that these structured sensing matrices are capable of preserving time domain GC owing to the preservation of spectral GC at different frequencies. As the number of structured rows in these matrices increase from zero to higher values, causal content at more and more frequencies is preserved, resulting in an improved causality detection performance. For a fully structured matrix, spectral GC at all frequencies is preserved.

Along with preserving the direction of causation, the structured sensing matrices used here also preserve the relative strength of causation in the compressed domain. This is evident from Fig. 3 where the strength of estimated GC F-statistic in the actual causal direction increases in the compressed domain, when $\gamma$, the unidirectional coupling coefficient, is increased. In the direction in which there is no coupling, estimated causal strengths are observed to be close to zero in the compressed domain.

In the case of multivariate simulations of sparse neural spike trains, it is seen that the use of conditional GC on compressed signals helps to discover the connectivity structure of the network perfectly for both circulant and toeplitz structured sensing matrices. 100% specificity and sensitivity of the technique is observed in this case and the performance of our method is shown to be superior to a GLM model-based method of discovering causal connectivity. Model-based methods are widely used for discovering functional (causal) connectivity from sparse neural spike trains, however may have limited scope because of specific model assumptions and high computational costs.

We also demonstrate a real world application of our method on neural spike train recordings from a rat prefrontal cortex, where the rat was performing a delayed-alternation task of working memory. These results shown in Fig. 4 are promising, as many common connections are discovered by circulant and toeplitz sensing matrices and some of these overlap with the connections reported by a GLM based method, ‘SGL’ [77]. As our method shows better performance than the SGL on simulated multivariate data, the results obtained in this work can be considered to be more reliable.
Networks that are required to be analyzed based on causal connections, and for which signals are acquired in the compressed domain, the use of structured sensing matrices would help to solve the causal estimation task within the compressed domain. Preservation of granger causality by circulant and toeplitz matrices would then be the deciding criteria for design of sensors in various systems. In fact, even currently, many measurement technologies impose structure on the matrix. This is because structured matrices possess other advantages, for example, requirement of lesser number of independent random variables for matrix generation and better efficiency of recovery algorithms as the matrix admits a fast matrix–vector multiply \(^9, ^{10}\). Causal inference in the compressed domain in these cases is a useful and powerful technique.

Also, in case of naturally occurring sensors that can be approximated to be sensing the signals based on operation by a structured matrix like partial or full circulant/toeplitz, causality analysis for the signals can be easily and reliably performed. One way by which the operation performed by a sensing matrix can be deciphered is by checking the properties of the sensed signal. For example, toeplitz matrices are known to perform a moving average operation on the input signals \(^9, ^{10}\). Some neural signals are known to be compressed signals \(^7, ^8\) while most single unit neuronal signals are sparse spike trains \(^9, ^{10}\). Also, some neuroimaging modalities, such as the fMRI widely employ compressed sensing approaches for acquisition of signals \(^100, ^{102}\). On a separate note, direct granger causal analysis of fMRI data has faced criticisms, one of the reasons for this being the unreliability of results due to low temporal resolution of acquired data \(^23\). Hence, recognition and design of causality-preserving sensing matrices can prove to be extremely useful for analysing functional neural connections and brain connectivity.

It will be interesting to check if any other type of sensing matrices with binary/real-valued entries preserve GC. Further, it will be useful to test if other causality measures such as nonlinear variations of GC, Transfer Entropy \(^45\) and Compression Complexity Causality \(^47\), preserve causality under circulant, toeplitz or other types of sensing matrices. We would also like to estimate GC directly from real compressively sensed signals acquired using structured sensing matrices discussed in this work and try to cross-verify the validity of the proposed scheme in such cases.

ACKNOWLEDGMENTS

The authors would like to thank past interns at the National Institute of Advanced Studies: M.A. Abhijith, Aishwarya Nambissan, Aswathi Gopinath and M. Krishnapriya (from Amrita University) for running some preliminary simulations associated with this work. This study is supported by Czech Science Foundation, Project No. GA19-16066S and by the Czech Academy of Sciences, Praenium Academiae awarded to M. Paluš. Nithin Nagaraj gratefully acknowledges the financial support of Cognitive Science Research Initiative, Dept. of Science & Tech., Govt. of India (CSRI-DST) Grant No. DST/CSRI/2017/54(G).

[1] D. L. Donoho, Compressed sensing, IEEE Transactions on information theory 52, 1289 (2006).
[2] E. J. Candès and M. B. Wakin, An introduction to compressive sampling, IEEE signal processing magazine 25, 21 (2008).
[3] M. Lustig, D. Donoho, and J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 58, 1182 (2007).
[4] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, Compressed sensing MRI, IEEE signal processing magazine 25, 72 (2008).
[5] D. Takhar, A compressed sensing camera: New theory and an implementation using digital micromirrors, Proc. Computational Imaging IV at SPIE Electronic Imaging, San Jose, 2006 (2006).
[6] O. Katz, Y. Bromberg, and Y. Silberberg, Compressive ghost imaging, Applied Physics Letters 95, 131110 (2009).
[7] C. Zhao, W. Gong, M. Chen, E. Li, H. Wang, W. Xu, and S. Han, Ghost imaging lidar via sparsity constraints, Applied Physics Letters 101, 141123 (2012).
[8] K. Kazimierczuk and V. Y. Orekhov, Accelerated nmr spectroscopy by using compressed sensing, Angewandte Chemie International Edition 50, 5556 (2011).
[9] D. J. Holland, M. J. Bostock, L. F. Gladden, and D. Nieltsch, Fast multidimensional nmr spectroscopy using compressed sensing, Angewandte Chemie 123, 6678 (2011).
[10] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, Physical review letters 105, 150401 (2012).
[11] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators, New Journal of Physics 14, 095022 (2012).
[12] M. Honma, K. Akiyama, M. Uemura, and S. Ikeda, Super-resolution imaging with radio interferometry using sparse modeling, Publications of the Astronomical Society of Japan 66, 95 (2014).
[13] S. Ikeda, F. Tazaki, K. Akiyama, K. Hada, and M. Honma, Prec: A new method for interferometry imaging from closure phase, Publications of the Astronomical Society of Japan 68, 45 (2016).
[14] Y. Nakanishi-Ohno, M. Haze, Y. Yoshida, K. Hukushima, Y. Hasegawa, and M. Okada, Compressed sensing in scanning tunneling mi-
crosscopy/spectroscopy for observation of quasi-particle interference, Journal of the Physical Society of Japan 85, 093702 (2016).

[15] S. Zhou, J. Lafferty, and L. Wasserman, Compressed and privacy-sensitive sparse regression, IEEE Transactions on Information Theory 55, 846 (2009).

[16] A. Blum, Random projection, margins, kernels, and feature-selection, in International Statistical and Optimization Perspectives Workshop “Subspace, Latent Structure and Feature Selection” (Springer, 2005) pp. 52–68.

[17] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh, Compressive sampling for signal classification, in 2006 Fortieth Asilomar Conference on Signals, Systems and Computers (IEEE, 2006) pp. 1430–1434.

[18] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, The smashed filter for compressive classification and target recognition, in Computation Imaging V, Vol. 6498 (International Society for Optics and Photonics, 2007) p. 64980H.

[19] M. F. Duarte, M. A. Davenport, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, Multiscale random projections for compressive classification, in 2007 IEEE International Conference on Image Processing, Vol. 6 (IEEE, 2007) pp. VI-161–VI-164.

[20] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, Sparse signal detection from incoherent projections, in 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Vol. 3 (IEEE, 2006) pp. III–III.

[21] P. Indyk and R. Motwani, Approximate nearest neighbors: towards removing the curse of dimensionality, in Proceedings of the thirtieth annual ACM symposium on Theory of computing (1998) pp. 604–613.

[22] C. Hegde, M. Wakin, and R. Baraniuk, Random projections for manifold learning, Advances in neural information processing systems 20 (2007).

[23] J. Pearl and D. Mackenzie, The book of why: the new science of cause and effect (Basic Books, 2018).

[24] A. Kathpalia and N. Nagaraj, Measuring causality, Resonance 191 (2021).

[25] A. K. Seth, A. B. Barrett, and L. Barnett, Granger causality analysis in neuroscience and neuroimaging, Journal of Neuroscience 35, 3293 (2015).

[26] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, Transfer entropy: a model-free measure of effective connectivity for the neurosciences, Journal of computational neuroscience 30, 45 (2011).

[27] T. J. Mosedale, D. B. Stephenson, M. Collins, and T. C. Mills, Granger causality of coupled climate processes: Ocean feedback on the north atlantic oscillation, Journal of climate 19, 1182 (2006).

[28] G. Tirabassi, C. Masoller, and M. Barreiro, A study of the air-sea interaction in the south atlantic convergence zone through granger causality, International Journal of Climatology 35, 3440 (2015).

[29] J. Runge, S. Bathiany, E. Bolth, G. Camps-Valls, D. Coumou, E. Deyle, C. Glynoum, M. Kretschmer, M. D. Mahecha, J. Muñoz-Mari, et al., Inferring causation from time series in earth system sciences, Nature communications 10, 1 (2019).

[30] M. Paluš, Multiscale atmospheric dynamics: cross-frequency phase-amplitude coupling in the air temperature, Physical review letters 112, 078702 (2014).

[31] N. Jajcay, S. Kravtsov, G. Sugihara, A. A. Tsonis, and M. Paluš, Synchronization and causality across time scales in el niño southern oscillation, npj Climate and Atmospheric Science 1, 1 (2018).

[32] J. Geweke, Inference and causality in economic time series models, Handbook of econometrics 2, 1101 (1984).

[33] C. Hiemstra and J. D. Jones, Testing for linear and nonlinear granger causality in the stock price-volume relation, The Journal of Finance 49, 1639 (1994).

[34] S. Z. Chiou-Wei, C.-F. Chen, and Z. Zhu, Economic growth and energy consumption revisited: evidence from linear and nonlinear granger causality, Energy Economics 30, 3063 (2008).

[35] T. Dimpfl and F. J. Peter, Using transfer entropy to measure information flows between financial markets, Studies in Nonlinear Dynamics & Econometrics 17, 85 (2013).

[36] M. Bauer, J. W. Cox, M. H. Caviness, J. J. Downs, and N. F. Thornhill, Finding the direction of disturbance propagation in a chemical process using transfer entropy, IEEE transactions on control systems technology 15, 12 (2007).

[37] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica: journal of the Econometric Society, 424 (1969).

[38] L. Barnett and A. K. Seth, The MVGC multivariate granger causality toolbox: a new approach to granger-causal inference, Journal of neuroscience methods 223, 50 (2014).

[39] T. W. Anderson, The statistical analysis of time series, Vol. 19 (John Wiley & Sons, 2011).

[40] D. Bell, J. Kay, and J. Malley, A non-parametric approach to non-linear causality testing, Economics Letters 51, 7 (1996).

[41] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Analyzing multiple nonlinear time series with extended granger causality, Physics letters A 324, 26 (2004).

[42] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble, Physical Review E 54, 6708 (1996).

[43] M. Le Van Quyen, J. Martinerie, C. Adam, and F. J. Varela, Nonlinear analyses of interictal eeg map the brain interdependences in human focal epilepsy, Physica D: Nonlinear Phenomena 127, 250 (1999).

[44] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Kernel method for nonlinear granger causality, Physical Review Letters 100, 144103 (2008).

[45] T. Schreiber, Measuring information transfer, Physical Review Letters 85, 461 (2000).

[46] M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Synchronization as adjustment of information rates: Detection from bivariate time series, Physical Review E 63, 046211 (2001).

[47] A. Kathpalia and N. Nagaraj, Data based intervention approach for complexity-causality measure, PeerJ Computer Science 5 (2019).

[48] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance, Biological cybernetics 85, 145 (2001).

[49] K. Friston, L. Harrison, and W. Penny, Dynamic causal
modelling, Neuroimage 19, 1273 (2003).
[50] J. Pearl, Causality: models, reasoning and inference, Vol. 29 (Springer, 2000).
[51] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, A single-pixel terahertz imaging system based on compressed sensing, Applied Physics Letters 93, 121105 (2008).
[52] V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candès, and M. Dahan, Compressive fluorescence microscopy for biological and hyperspectral imaging, Proceedings of the National Academy of Sciences 109, E1679 (2012).
[53] A. B. Tayler, M. Benning, A. J. Sederman, D. J. Holland, and L. F. Gladden, Ultrafast magnetic-resonance-imaging velocimetry of liquid-liquid systems: Overcoming chemical-shift artifacts using compressed sensing, Physical Review E 89, 063009 (2014).
[54] Y. Ma, Y. Lee, C. Best-Popescu, and L. Gao, High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells, Proceedings of the National Academy of Sciences 118 (2021).
[55] X. Andrade, J. N. Sanders, and A. Aspuru-Guzik, Application of compressed sensing to the simulation of atomic systems, Proceedings of the National Academy of Sciences 109, 13928 (2012).
[56] C. Saïgre-Tardif, R. Faigiri, H. Zhao, L. Li, and P. del Hougne, Intelligent meta-imagers: from compressed to learned sensing, Applied Physics Reviews 9, 011314 (2022).
[57] M. Tsumori, S. Nagai, R. Harakawa, T. Sasaki, and M. Iwahashi, Contour-adaptive image restoration based on compressed sensing for ultrafast phenomena, Applied Physics Letters 117, 044102 (2020).
[58] V. J. Barranca, D. Zhou, and D. Cai, Compressive sensing reconstruction of feed-forward connectivity in pulse-coupled nonlinear networks, Physical Review E 93, 062016 (2016).
[59] A. Sheikhattar, S. Miran, J. Liu, J. B. Fritz, S. A. Shamma, P. O. Kanold, and B. Babadi, Extracting neuronal functional network dynamics via adaptive granger causality analysis, Proceedings of the National Academy of Sciences 115, E3869 (2018).
[60] M. Paluš, From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics, Contemporary physics 48, 307 (2007).
[61] J. D. Finkle, J. J. Wu, and N. Bagheri, Windowed granger causality inference strategy improves discovery of gene regulatory networks, Proceedings of the National Academy of Sciences 115, 2252 (2018).
[62] E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on information theory 52, 489 (2006).
[63] E. J. Candès et al., The restricted isometry property and its implications for compressed sensing, Comptes rendus mathematique 346, 589 (2008).
[64] \|u\|_p denotes the \(p\)-norm of vector \(u\), which is computed as \(\|u\|_p = (\sum_{i=1}^{n} |u_i|^p)^{1/p}\), where \(n\) is the length of the vector.
[65] Y. Tsaig and D. L. Donoho, Extensions of compressed sensing, Signal processing 86, 549 (2006).
[66] J. A. Tropp and A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Transactions on information theory 53, 4655 (2007).
[67] A. G. Nedungadi, G. Rangarajan, N. Jain, and M. Ding, Analyzing multiple spike trains with nonparametric granger causality, Journal of computational neuroscience 27, 55 (2009).
[68] K. Sameshima and L. A. Baccalá, Using partial directed coherence to describe neuronal ensemble interactions, Journal of neuroscience methods 94, 93 (1999).
[69] L. Zhu, Y.-C. Lai, F. C. Hoppensteadt, and J. He, Probing changes in neural interaction during adaptation, Neural Computation 15, 2359 (2003).
[70] H. Xu, M. Farajtabar, and H. Zha, Learning granger causality for hawkes processes, in International conference on machine learning (PMLR, 2016) pp. 1717–1726.
[71] R. E. Spinney, M. Prokopenko, and J. T. Lizier, Transfer entropy in continuous time, with applications to jump and neural spiking processes, Physical Review E 95, 032319 (2017).
[72] D. P. Shorten, R. E. Spinney, and J. T. Lizier, Estimating transfer entropy in continuous time between neural spike trains or other event-based data, PLoS computational biology 17, e1008054 (2021).
[73] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects, Journal of neurophysiology 93, 1074 (2005).
[74] M. Okatan, M. A. Wilson, and E. N. Brown, Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity, Neural computation 17, 1927 (2005).
[75] S. Kim, D. Putrino, S. Ghosh, and E. N. Brown, A granger causality measure for point process models of ensemble neural spiking activity, PLoS computational biology 7, e1001110 (2011).
[76] Z. Chen, D. F. Putrino, S. Ghosh, R. Barbieri, and E. N. Brown, Statistical inference for assessing functional connectivity of neuronal ensembles with sparse spiking data, IEEE transactions on neural systems and rehabilitation engineering 19, 121 (2010).
[77] C. Zhang, Y. Chai, X. Guo, M. Gao, D. Devilbiss, and Z. Zhang, Statistical learning of neuronal functional connectivity, Technometrics 58, 350 (2016).
[78] A. Casile, R. T. Faghihi, and E. N. Brown, Robust point-process granger causality analysis in presence of exogenous temporal modulations and trial-by-trial variability in spike trains, PLoS computational biology 17, e1007675 (2021).
[79] M. Krumin and S. Shoham, Multivariate autoregressive modeling and granger causality analysis of multiple spike trains, Computational intelligence and neuroscience 2010 (2010).
[80] S. Eldawlatly, Y. Zhou, R. Jin, and K. G. Oweiss, On the use of dynamic bayesian networks in reconstructing functional neuronal networks from spike train ensembles, Neural computation 22, 158 (2010).
[81] T. Berry, F. Hamilton, N. Peixoto, and T. Sauer, Detecting connectivity changes in neuronal networks, Journal of neuroscience methods 209, 388 (2012).
[82] A. Arnold, Y. Liu, and N. Abe, Temporal causal modeling with graphical granger methods, in Proceedings of the 13th ACM SIGKDD international conference on
Knowledge discovery and data mining (2007) pp. 66–75.
[83] A. Shojaie and G. Michailidis, Discovering graphical
granger causality using the truncating lasso penalty,
Bioinformatics 26, 1517 (2010).
[84] A. C. Lozano, N. Abe, Y. Liu, and S. Rosset, Grouped
graphical granger modeling for gene expression regulatory
networks discovery, Bioinformatics 25, i110 (2009).
[85] A. Emad and O. Milenkovic, Caspian: A causal
compressive sensing algorithm for discovering directed in-
teractions in gene networks, PloS one 9 (2014).
[86] K. Hlaváčková-Schindler and S. Pereverzyev, Lasso
granger causal models: some strategies and their effi-
ciency for gene expression regulatory networks, in De-
cision Making: Uncertainty, Imperfection, Deliberation
and Scalability (Springer, 2015) pp. 91–117.
[87] E. J. Candes and T. Tao, Near-optimal signal recovery
from random projections: Universal encoding strategies?,
IEEE transactions on information theory 52, 5406
(2006).
[88] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin,
A simple proof of the restricted isometry property for
random matrices, Constructive Approximation 28, 253
(2008).
[89] H. Rauhut, Circulant and toeplitz matrices in com-
pressed sensing, arXiv preprint arXiv:0902.4394 (2009).
[90] W. Yin, S. Morgan, J. Yang, and Y. Zhang, Practical
compressive sensing with toeplitz and circulant matri-
ces, in Visual Communications and Image Processing
2010, Vol. 7744 (International Society for Optics and
Photonics, 2010) p. 77440K.
[91] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright,
and R. D. Nowak, Toeplitz-structured compressed sens-
ing matrices, in 2007 IEEE/SP 14th Workshop on Sta-
tistical Signal Processing (IEEE, 2007) pp. 294–298.
[92] H. Rauhut, J. Romberg, and J. A. Tropp, Restricted
isometries for partial random circulant matrices, Ap-
plied and Computational Harmonic Analysis 32, 242
(2012).
[93] P. Davis, Circulant Matrices, Monographs and text-
books in pure and applied mathematics (Wiley, 1979).
[94] J. Geweke, Measurement of linear dependence and feed-
back between multiple time series, Journal of the Amer-
ican statistical association 77, 304 (1982).
[95] M. Grant and S. Boyd, CVX: Matlab software for dis-
ciplined convex programming, version 2.1 (2014).
[96] R. M. Gray et al., Toeplitz and circulant matrices: A re-
view, Foundations and Trends in Communications and
Information Theory 2, 155 (2006).
[97] S. Ganguli and H. Sompolinsky, Compressed sensing,
sparsity, and dimensionality in neuronal information
processing and data analysis, Annual review of neuro-
science 35, 485 (2012).
[98] N. Nagaraj and K. Sahasranand, Neural signal multi-
plexing via compressed sensing, in 2016 International
Conference on Signal Processing and Communications
(SPCOM) (IEEE, 2016) pp. 1–5.
[99] W. Gerstner and W. M. Kistler, Spiking neuron mod-
els: Single neurons, populations, plasticity (Cambridge
university press, 2002).
[100] H. Jung, J. C. Ye, and E. Y. Kim, Improved k–t blast
and k–t sense using focuss, Physics in Medicine & Biol-
ogy 52, 3201 (2007).
[101] X. Zong, J. Lee, A. J. Poplawsky, S.-G. Kim, and J. C.
Ye, Compressed sensing fMRI using gradient-recalled
echo and epi sequences, NeuroImage 92, 312 (2014).
[102] J. Tsao and S. Kozerke, Mri temporal acceleration tech-
niques, Journal of Magnetic Resonance Imaging 36, 543
(2012).