Dynamic Estimation of Causal Influences in Sparsely-Interacting Neuronal Ensembles

Alireza Sheikhattar
Department of Electrical & Computer Engineering
University of Maryland
Email: arsha89@umd.edu

Behtash Babadi
Department of Electrical & Computer Engineering
Institute for Systems Research
University of Maryland
Email: behtash@umd.edu

***Abstract*—In this paper, we consider a neuronal ensemble under spontaneous activity where each neuron modulates the activity of the others through its spiking history. Assuming that the cross-history dependence parameters of the ensemble are sparse and time-varying, we perform adaptive system identification using sparse point process filters. We then provide a novel filtering and smoothing algorithm for estimating the Granger causality with high temporal resolution and with recursively computed statistical confidence intervals. We provide simulation studies which reveal significant performance gains obtained by our proposed technique in describing the causal influences in neuronal ensemble activity.**

***Index Terms*—Granger causality, filtering and smoothing, non-central chi-squared distribution, deviance statistics, sparsity.**

I. Introduction

Granger causality (GC), in its general form, is a well-established statistical measure for causality inference in time series analysis, which belongs to the general class of time series inference (TSI) methods [1] used in a wide range of applications. More recently, GC has become increasingly popular in the context of neural signal processing, to infer causal interactions from simultaneously recorded neural time series data, which could be any continuous-valued data such as electroencephalographic (EEG), magnetoencephalographic (MEG), functional Magnetic Resonance (fMRI) and local field potential (LFP) recordings; or any discrete data such as single-unit or multi-unit recordings.

Originally proposed by Wiener [2] based on temporal predictability, the basic idea of GC was later adapted into practical form by Granger [3] in the context of multivariate Auto-Regressive (MAR) modeling of stochastic processes. The rationale behind GC analysis is easily intelligible: Given two data time series $\{X_t, Y_t\}_{t=1}^{T}$, if including the history of Y_t can improve the prediction of X_{t+1}, it is implied that the history of Y_t contains unique information about X_t which is not already captured by its current covariates. In this case, it is said that Y_t has a causal influence on X_t.

Several methods have been proposed to apply GC analysis to spike data recordings from neuronal ensembles. However, very few of them take into account the discrete nature of spiking data. Particularly, in [4], the network likelihood model is used for the analysis of functional interactions among multiple spike trains from ensembles of interacting neurons. A likelihood-based variant of GC measure is proposed in [5], based on the point process framework which is directly applicable to discrete neural spiking activity of neuronal ensembles. In addition, most of the existing analyses involving GC provide static estimates throughout the entire data duration, and hence cannot capture the dynamics of causal influences over time.

In this paper, we propose a novel time-varying measure of GC based on the forgetting factor mechanism of the RLS algorithm suitable for adaptive settings. Inspired by the natural parsimony of neurophysiological time-constants in neuronal ensemble dynamics, such as those in sensory neurons with sharp frequency tuning, we assume that the cross-history dependence of the neurons in the ensemble can be described by sparse vectors. We then employ the RLS-weighted data log-likelihoods proposed in our earlier work [6] in order to design a low-complexity adaptive filter for recursive estimation of the proposed causality measure. The sparse adaptive filters in [6] provide robust and accurate estimates of the network parameters as compared to existing adaptive filters.

We next develop a statistical inference framework for the proposed GC measure based on the analysis of deviance, identifying the potential causal interactions between neurons. We derive a non-central chi-squared filtering and smoothing algorithm to estimate and track the dynamics of the causal influences among neurons in an ensemble through the non-centrality parameter associated with the chi-squared distributed deviance data. Moreover, we compute the statistical confidence regions for these estimates across time. Finally, we provide simulation studies in order to assess the identification and tracking capabilities of our proposed algorithms, which reveal remakable performance in both detecting the existing causal links and avoiding false alarms, while capturing the dynamics of the causal interactions among neurons.

The outline of the paper is organized as follows. We will introduce the notational conventions and preliminaries of the point process framework in section II. In Section III, we will present the main algorithm including the time-varying RLS-weighted GC measure, a fully recursive algorithm for computation of this new causality measure, and a non-central chi-squared filtering and smoothing algorithm. Section IV presents the simulation studies, and section V concludes the paper.

978-1-4673-9457-4/16/$31.00 ©2016 IEEE

II. Preliminaries and Notations

Throughout the paper, we use bold-face symbols and letters to denote vectors and matrices. We employ the point process framework to model the neural spiking data, which enables us to exploit the discrete nature of spiking data. A point process is a stochastic sequence of discrete events occurring at random points in continuous time, well-suited for neuronal data modeling where the action potentials are distributed randomly in time [7]. For ease of computation, our analysis in this paper is based on a discrete version of point processes, in which the observation interval $\mathcal{T}$ is discretized by bins of length Δ, resulting in a binary sequence $\{n_k^{(c)}\}_{k=1}^K$ of spike trains for neurons indexed by $c = 1, \ldots, C$. A point process can be fully characterized by its conditional intensity function (CIF) denoted by λ_k, representing the neuron's instantaneous firing rate at bin k. Given a sufficiently small Δ, the binary spiking sequence can be modeled by a conditionally independent Bernoulli process with success probability of $\lambda_k \Delta$.

We adopt a dynamic Generalized Linear Model (GLM) with logistic link function for the CIF, give by:

$$\lambda_k^{(c)} \Delta := \text{logit}^{-1}\left(\boldsymbol{\omega}_k^{(c)'} \mathbf{x}_k^{(c)}\right), \tag{1}$$

where $\text{logit}^{-1}(a) := \frac{1}{1+\exp(-a)}$ is the logistic function, and $\boldsymbol{\omega}_k^{(c)} := [\mu_k^{(c)}, \boldsymbol{\theta}_k^{(c)'}]'$ denotes the time-varying parameter vector of size $M^{(c)}$ at time k, characterizing the underlying neural encoding process of neuron (c) while capturing the dynamicity of tuning features, composed of a scalar baseline firing rate $\mu_k^{(c)}$ and the sparse modulation vector $\boldsymbol{\theta}_k^{(c)}$ tuning the effects of all possible covariates collected in the covariate vector $\mathbf{x}_k^{(c)} := \left[1, x_k^{(c)}, x_{k-1}^{(c)}, \cdots, x_{k-M+2}^{(c)}\right]'$. The effective covariates in spiking statistics include the spiking history, concurrent ensemble activity of other neurons, and extrinsic stimuli. In this paper, we consider only the intrinsic neuronal covariates from the ensemble spiking history $\left\{\{n_i^{(c)}\}_{i=1}^{k-1}\right\}_{c=1}^C$ for the GLM models at bin k, with the final aim of inspecting causal interactions among the neurons in the ensemble.

Recently in [6], we modified the data log-likelihoods to the adaptive setting, by utilizing the forgetting factor mechanism of the well-known class of Recursive Least Squares (RLS) algorithms, in which the effective log-likelihood up to time k, denoted by $\ell_k^\beta(\boldsymbol{\omega})$ is defined regressively in time as:

$$\begin{aligned}\ell_k^\beta(\boldsymbol{\omega}) &:= \sum_{i=1}^k \beta^{k-i} \ell_i(\boldsymbol{\omega}) \\ &= \sum_{i=1}^k \beta^{k-i}\big(n_i \mathbf{x}_i'\boldsymbol{\omega} - \log(1+\exp(\mathbf{x}_i'\boldsymbol{\omega}))\big),\end{aligned} \tag{2}$$

for some general tuning vector $\boldsymbol{\omega}$, and forgetting factor $0 < \beta \leq 1$, and $\ell_i(\boldsymbol{\omega})$ denotes the data log-likelihood at bin i with a logit-linked CIF obeying Bernoulli statistics.

We further assume that the parameter vectors $\boldsymbol{\theta}_k^{(c)}$ for $c = 1, \cdots, C$ and $k = 1, \cdots, K$ are sparse, resulting in a sparsely-interacting neuronal ensemble. Throughout the rest of paper, all the parameter vector estimates $\widehat{\boldsymbol{\omega}}_k$ for $k = 1, 2, \cdots, K$ are obtained using the sparse adaptive point process filter ℓ_1-PPF$_1$ introduced in [6]. The ℓ_1-PPF$_1$ estimates sparse time-varying parameter vectors from point process observations in an online fashion by recursively maximizing a sequence of ℓ_1-regularized log-likelihoods via a proximal algorithm:

$$\widehat{\boldsymbol{\omega}}_k = \underset{\boldsymbol{\omega}_k}{\text{argmax}} \left\{\ell_k^\beta(\boldsymbol{\omega}_k) - \gamma \|\boldsymbol{\omega}_k\|_1\right\}. \tag{3}$$

Statistical confidence regions for the estimated $\widehat{\boldsymbol{\omega}}_k$ components can also be obtained using a recursive node-wise regression procedure [6], [8] via the SPARLS algorithm [9].

III. Dynamic Estimation of Granger Causality

In this section, we first introduce a novel time-varying GC measure for adaptive identification of dynamic causal interactions among neurons in an ensemble. Next, we present a recursive algorithm for computation of the new GC measure in a online fashion. Finally, we develop a non-central chi-squared filtering and smoothing algorithm for estimation of the non-centrality parameter from the deviance data, as well as a method for constructing statistical confidence regions for the estimated GC measures across time.

A. RLS-weighted Granger Causality Measure

Consider simultaneous spike recordings from an ensemble of C neurons indexed by $c = 1, 2, \cdots, C$, denoted by $\{\{n_k^{(c)}\}_{k=1}^K\}_{c=1}^C$ over the time bins $k = 1, \cdots, K$. Each neuron (c) is associated with a sparse modulation parameter vector $\boldsymbol{\omega}_k^{(c)} = [\mu_k^{(c)}, \boldsymbol{\omega}_k^{(c,1)'}, \boldsymbol{\omega}_k^{(c,2)'}, \cdots, \boldsymbol{\omega}_k^{(c,C)'}]'$ consisting of a scalar baseline firing parameter $\mu_k^{(c)}$ and a collection of sparse modulation vectors $\{\boldsymbol{\omega}_k^{(c,\tilde{c})}\}_{\tilde{c}=1}^C$ accounting for the influences of the ensemble spiking activity on neuron (c). More specifically, each $\boldsymbol{\omega}_k^{(c,\tilde{c})}$ reflects the dependency of the spiking statistics of neuron (c) on the spiking history of neuron $(\tilde{c})$. We refer to this model, where the history of all the neurons are taken into account, as the *full model*. The $\boldsymbol{\omega}_k^{(c,\setminus\tilde{c})}$ represents the parameter vector of the *reduced model* which excludes only the effect of neuron $(\tilde{c})$ on (c).

A likelihood-based GC measure has been proposed in [5] for the point process framework, which is directly applicable to discrete spiking data. This measure, in its general form, is established based on the log-likelihood ratio statistics. Let us consider neuron (c) as the target neuron whose activity $\mathbf{n}^{(c)} := [n_1^{(c)}, n_2^{(c)}, \cdots, n_K^{(c)}]'$ is subject to causality analysis. The log-likelihood ratio statistic associated with the causal influence of neuron $(\tilde{c})$ on neuron (c) can be defined as:

$$\mathcal{F}_{(\tilde{c} \mapsto c)} := \log \frac{\mathcal{L}(\widehat{\boldsymbol{\omega}}^{(c)} | \, \mathbf{n}^{(c)}, \mathcal{H}^{(c)})}{\mathcal{L}(\widehat{\boldsymbol{\omega}}^{(c,\setminus\tilde{c})} | \, \mathbf{n}^{(c)}, \mathcal{H}^{(c,\setminus\tilde{c})})}, \tag{4}$$

where $\mathcal{L}(\widehat{\boldsymbol{\omega}} | \mathbf{n}, \mathcal{H})$ denotes the likelihood of estimated parameter vector $\widehat{\boldsymbol{\omega}}$ given the observation sequence $\mathbf{n}$ and the history of the covariates included in the model denoted by $\mathcal{H}$. Based on this formulation, GC from neuron $(\tilde{c})$ to neuron (c) can be

measured as the reduction in the point process log-likelihood of neuron (c) for the reduced model compared to the full model, obtained by excluding the history covariates associated with neuron $(\tilde{c})$. Note that this form of GC with the mutual set of conditional covariates (the spiking history of all other neurons in ensemble) is referred to as *conditional Granger causality*, which allows us to effectively distinguish between the direct and indirect causal interactions between a set of multiple simultaneously acquired data time series.

The Granger causality, in its classical form, is expressed in terms of linear multi-variate auto-regressive (MAR) models [3], [10]. Unlike the classical GC measure, specialized for models with linear Gaussian statistics, the GC measure in Eq. (4) benefits from the likelihood-based inference covering a large range of complex statistical models. Moreover, the MAR-based GC measure requires the data time series to be stationary, whereas most neural data recordings (such as single-unit, fMRI, MEG, EEG) exhibit a degree of non-stationarity with underlying parameters changing in time.

In order to address this issue, we introduce a novel time-varying GC measure to capture the dynamics of causal interactions among neurons. To this end, we leverage the RLS-weighted log-likelihood functions in Eq. (2) to induce adaptivity into the measure. Replacing the standard data log-likelihoods in the likelihood-based GC in Eq. (4) by their adaptive counterparts in Eq. (2), we define the new time-varying GC measure from neuron $(\tilde{c})$ to neuron (c) at time bin k as follows:

$$\mathcal{F}^{\beta}_{k,(\tilde{c}\mapsto c)} := s_k(\widehat{\boldsymbol{\omega}}_k^{(c,\tilde{c})})\big(\ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k^{(c)}) - \ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k^{(c,\setminus\tilde{c})})\big), \quad (5)$$

where $s_k(\widehat{\boldsymbol{\omega}}_k^{(c,\tilde{c})}) := \mathrm{sign}(\sum_l \widehat{\omega}_{k,l}^{(c,\tilde{c})})$ reflects the effective sign of estimated coefficients $\widehat{\boldsymbol{\omega}}_k^{(c,\tilde{c})}$, which determines the excitatory or inhibitory nature of underlying causal interaction, and the rest is simply the log-likelihood ratio statistics of the reduced and full models at time bin k.

B. Recursive Computation of Granger Causality

The RLS-inspired exponential weighting of data log-likelihoods in Eq. (2) paves the way for recursive computation of the log-likelihood ratio statistics in Eq. (5). With the final aim of online estimation of dynamic causal interactions among neurons, we develop a low-complexity recursive update rules for the adaptive log-likelihoods $\ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k)$, and consequently for the proposed GC measure $\mathcal{F}^{\beta}_{k,(\tilde{c}\mapsto c)}$.

We exploit the smoothness of the point process log-likelihood function, and approximate each scalar-valued log-likelihood function $\ell_i(\widehat{\boldsymbol{\omega}}_k)$ using a second order Taylor's series expansion around the causal counterpart $\widehat{\boldsymbol{\omega}}_i$, where we retain the first three terms of expansion as follows:

$$\ell_i(\widehat{\boldsymbol{\omega}}_k) \approx \ell_i(\widehat{\boldsymbol{\omega}}_i) + (\widehat{\boldsymbol{\omega}}_k - \widehat{\boldsymbol{\omega}}_i)'\dot{\boldsymbol{\ell}}_i(\widehat{\boldsymbol{\omega}}_i) + \tfrac{1}{2}(\widehat{\boldsymbol{\omega}}_k - \widehat{\boldsymbol{\omega}}_i)'\ddot{\boldsymbol{\ell}}_i(\widehat{\boldsymbol{\omega}}_i)(\widehat{\boldsymbol{\omega}}_k - \widehat{\boldsymbol{\omega}}_i), \quad (6)$$

where $\dot{\boldsymbol{\ell}}_i(\cdot)$ and $\ddot{\boldsymbol{\ell}}_i(\cdot)$ denote the gradient vector and Hessian matrix with respect to $\boldsymbol{\omega}$ which can be computed from Eq. (2) for the logit-linked GLM point process model as follows:

$$\dot{\boldsymbol{\ell}}_i(\widehat{\boldsymbol{\omega}}_i) = \varepsilon_i \mathbf{x}_i, \quad (7)$$

$$\ddot{\boldsymbol{\ell}}_i(\widehat{\boldsymbol{\omega}}_i) = -\kappa_i \mathbf{x}_i \mathbf{x}_i', \quad (8)$$

where $\varepsilon_i := n_i - \lambda_i(\widehat{\boldsymbol{\omega}}_i)\Delta$ denotes the point process innovation term at time step i, and the coefficient $\kappa_i := \lambda_i(\widehat{\boldsymbol{\omega}}_i)\Delta(1 - \lambda_i(\widehat{\boldsymbol{\omega}}_i)\Delta)$ is obtained from the second-order derivative of logistic link function. Substituting the quadratic Taylor's approximation of Eq. (6) into Eq. (2) and rearranging terms will lead us to the following recursive formulation for the adaptive log-likelihoods at time bin k:

$$\ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k) = a_k + \widehat{\boldsymbol{\omega}}_k'\mathbf{b}_k + \frac{1}{2}\widehat{\boldsymbol{\omega}}_k'\mathbf{B}_k\widehat{\boldsymbol{\omega}}_k, \quad (9)$$

where

$$a_k = \sum_{i=1}^{k} \beta^{k-i}\big(\ell_i(\widehat{\boldsymbol{\omega}}_i) - \varepsilon_i \mathbf{x}_i'\widehat{\boldsymbol{\omega}}_i - \frac{1}{2}\kappa_i(\mathbf{x}_i'\widehat{\boldsymbol{\omega}}_i)^2\big),$$

$$\mathbf{b}_k = \sum_{i=1}^{k} \beta^{k-i}\big(\varepsilon_i + \kappa_i \mathbf{x}_i'\widehat{\boldsymbol{\omega}}_i\big)\mathbf{x}_i, \text{ and } \mathbf{B}_k = -\sum_{i=1}^{k} \beta^{k-i}\kappa_i \mathbf{x}_i\mathbf{x}_i'.$$

It is easy to see that all a_k, $\mathbf{b}_k$ and $\mathbf{B}_k$ follow recursive update rules at time step k as follows:

$$a_k = \beta a_{k-1} + \ell_k(\widehat{\boldsymbol{\omega}}_k) - \varepsilon_k \mathbf{x}_k'\widehat{\boldsymbol{\omega}}_k - \frac{1}{2}\kappa_k(\mathbf{x}_k'\widehat{\boldsymbol{\omega}}_k)^2,$$

$$\mathbf{b}_k = \beta\mathbf{b}_{k-1} + \big(\varepsilon_k + \kappa_k \mathbf{x}_k'\widehat{\boldsymbol{\omega}}_k\big)\mathbf{x}_k, \text{ and}$$

$$\mathbf{B}_k = \beta\mathbf{B}_{k-1} - \kappa_k \mathbf{x}_k\mathbf{x}_k'.$$

By performing the recursive computation of Eq. (9) for both the full model and the reduced model, a fully recursive update procedure for the new GC measure in Eq. (5) will be achieved, which enables us to track the causal interactions between neurons in an online fashion. This procedure is summarized in Algorithm 1.

Algorithm 1 Recursive update rule for $\ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k)$

Inputs: $\mathbf{x}_k$, n_k, $\widehat{\boldsymbol{\omega}}_k$, a_{k-1}, $\mathbf{b}_{k-1}$, and $\mathbf{B}_{k-1}$.

1: $\lambda_k\Delta = \mathrm{logit}^{-1}(\mathbf{x}_k'\widehat{\boldsymbol{\omega}}_k)$
2: $\kappa_k = \lambda_k\Delta(1 - \lambda_k\Delta)$
3: $\varepsilon_k = n_k - \lambda_k\Delta$
4: $a_k = \beta\, a_{k-1} + \ell_k(\widehat{\boldsymbol{\omega}}_k) - \varepsilon_k \mathbf{x}_k'\widehat{\boldsymbol{\omega}}_k - \frac{1}{2}\kappa_k(\mathbf{x}_k'\widehat{\boldsymbol{\omega}}_k)^2$
5: $\mathbf{b}_k = \beta\, \mathbf{b}_{k-1} + \big(\varepsilon_k + \kappa_k \mathbf{x}_k'\widehat{\boldsymbol{\omega}}_k\big)\mathbf{x}_k$
6: $\mathbf{B}_k = \beta\, \mathbf{B}_{k-1} - \kappa_k \mathbf{x}_k\mathbf{x}_k'$

Output: $\ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k) = a_k + \widehat{\boldsymbol{\omega}}_k'\mathbf{b}_k + \frac{1}{2}\widehat{\boldsymbol{\omega}}_k'\mathbf{B}_k\widehat{\boldsymbol{\omega}}_k$

C. Non-central χ^2 Filtering and Smoothing Algorithm

The relative strength of a potential causal interaction can be inferred from the estimated GC values. However, the nonzero values of GC measure $\mathcal{F}^{\beta}_{k,(\tilde{c}\mapsto c)}$ do not necessarily imply the existence of a causal influence from neuron $(\tilde{c})$ to neuron (c), due to the noisy and biased nature of the estimates. Hence, it is essential to investigate the statistical significance of the computed GC measures, in order to infer the possible causal interactions between associated neurons with statistical precision.

We investigate the statistical significance of the recursively computed GC measures using the analysis of deviance statistics. The deviance statistic is a likelihood-based measure widely used for statistical inference of generalized linear models, which generalizes the concept of residual sum of squares in the ordinary linear regression models. The *deviance* of a specific fitted model is defined as twice the log-likelihood ratio of the fitted model to the saturated model,

$$D(\widehat{\boldsymbol{\omega}}^{\max};\widehat{\boldsymbol{\omega}}) := 2\big(\ell(\widehat{\boldsymbol{\omega}}^{\max}) - \ell(\widehat{\boldsymbol{\omega}})\big), \tag{10}$$

where $\widehat{\boldsymbol{\omega}}$ and $\widehat{\boldsymbol{\omega}}^{\max}$ denote the estimated parameter vector associated with the fitted and saturated model, respectively. In the GLM setting, the saturated model refers to the over-fitted model with maximum number of parameters equal to the number of observations, where the log-likelihood function reaches its maximum value. It is clear that the likelihood-based GC in Eq. (4) can be written as the deviance difference of the full ($\widehat{\boldsymbol{\omega}}^{(F)}$) and reduced ($\widehat{\boldsymbol{\omega}}^{(R)}$) models, $D(\widehat{\boldsymbol{\omega}}^{(F)};\widehat{\boldsymbol{\omega}}^{(R)}) = 2\big(\ell(\widehat{\boldsymbol{\omega}}^{(F)}) - \ell(\widehat{\boldsymbol{\omega}}^{(R)})\big)$. The deviance difference has been proven as an effective statistical measure for assessing the discrepancy between any two fitted models, in particular in the case of nested GLMs, where the reduced model is a special case of the more complex full model [11].

Our approach to the statistical inference is based on the asymptotic distribution of the deviance statistic, which has been well studied in the literature [12]–[15]. It has been shown that the deviance difference statistic asymptotically follows a non-central chi-squared distribution with M^d degrees of freedom and the non-centrality parameter ν under certain regularity conditions, where $M^d := M^F - M^R$ is the dimensionality difference of the two nested models. This asymptotic result reduces to a central chi-squared distribution when there is no causal link between two neurons. These asymptotic results can be extended to the adaptively modified version of the deviance statistic as follows:

$$D_k^{\beta}(\widehat{\boldsymbol{\omega}}^{(F)};\widehat{\boldsymbol{\omega}}^{(R)}) := 2(1+\beta)\left(\ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k^{(F)}) - \ell_k^{\beta}(\widehat{\boldsymbol{\omega}}_k^{(R)})\right) \xrightarrow{d} \chi^2(M^d,\nu_k),$$

as β approaches unity; $\nu_k \geq 0$ is the non-centrality parameter at time step k, and D_k^{β} is the deviance statistic conformed to the forgetting factor-based weighting of log-likelihoods and the ℓ_1-regularized ML estimates [16]. The non-centrality parameter ν_k takes a key role in the statistical inference procedure, as the small values of ν_k imply that there exists no significant source of deviance between the reduced and full models. On the contrary, significant values of ν_k can be inferred as a significant deviance between the more complex full model and the reduced model. In this case, we deduce that the history components of neuron $(\tilde{c})$ improves the prediction of the activity of target neuron (c).

Since ν_k can only take non-negative values, we choose $\nu_k = \exp(z_k)$ to be exponentially linked to a state variables z_k. Let z_k follow a first order AR(1) state dynamics

$$z_k = \rho z_{k-1} + e_k, \tag{11}$$

where $0 < \rho \leq 1$ is a scaling factor and $e_k \sim \mathcal{N}(0,\sigma^2)$ is a zero-mean i.i.d. Gaussian random variable with variance σ_e^2. Our algorithm comprises of three main steps: the forward recursive non-central chi-squared filter, the backward fixed interval smoother (FIS), and an optional Expectation Maximization (EM) algorithm to estimate the optimal $\widehat{\sigma}_e^2$ given the observed set of deviance data.

For the filtering algorithm, we exploit the unimodal property of non-central chi-squared distribution and make a recursive Gaussian approximation to the posterior probability density function $p(z_k|D_{1:k}^{\beta})$, where the posterior modes and variances are computed in recursive form [7]. Let $z_{k|l}$ and $\sigma_{k|l}^2$ denote the respective mode and variance of the state variable z_k given the deviance samples up to and including time l, $\{D_i^{\beta}\}_{i=1}^{l}$. Using the Bayes rule and substituting the non-central chi-squared density function into the log-posterior, we get:

$$z_{k|k} := \underset{z_k}{\operatorname{argmax}} \left\{ -\frac{(D_k^{\beta} + \exp(z_k))}{2} + \frac{\alpha}{2}(\log D_k^{\beta} - z_k) + \log I_{\alpha}(\zeta_k) - \frac{(z_k - z_{k|k-1})^2}{2\sigma_{k|k-1}^2} \right\}, \tag{12}$$

where $\zeta_k := \sqrt{D_k^{\beta}\exp(z_k)}$, and $\alpha := M_d/2 - 1$ denotes the order of modified Bessel function of the first kind denoted by $I_{\alpha}(.)$, $z_{k|k-1} = \rho z_{k-1|k-1}$ and $\sigma_{k|k-1}^2 = \rho^2\sigma_{k-1|k-1}^2 + \sigma_e^2$. Note that in Eq. (12) a Gaussian approximation is applied to the density $p(z_k|D_{1:k-1}^{\beta}) \sim \mathcal{N}(z_{k|k-1},\sigma_{k|k-1}^2)$. Hence, the posterior mode $z_{k|k}$ can be computed as the solution to the following nonlinear equation:

$$z_k = z_{k|k-1} + \frac{\sigma_{k|k-1}^2}{2}\big(\zeta_k\, r_{\alpha}(\zeta_k) - \exp(z_k)\big), \tag{13}$$

where the function $r_{\alpha}(\zeta) := I_{\alpha+1}(\zeta)/I_{\alpha}(\zeta)$ is the ratio of modified Bessel functions of first kind with order difference of one. This nonlinear equation can be solved numerically using well-known iterative methods such as the Newton's method.

Given $z_{k|k}$, the posterior variance $\sigma_{k|k}^2$ can be computed as the negative inverse of the second order derivative of posterior density function:

$$\sigma_{k|k}^2 = \left(\left(\sigma_{k|k-1}^2\right)^{-1} + \frac{\exp(z_{k|k})}{2} - \frac{\zeta_{k|k}^2}{4}\left(1 - \frac{I_{\alpha-1}(\zeta_{k|k})I_{\alpha+1}(\zeta_{k|k})}{I_{\alpha}(\zeta_{k|k})^2}\right)\right)^{-1},$$

where $\zeta_{k|k} := \sqrt{D_k^{\beta}\exp(z_{k|k})}$. We used the recurrence relation, $I_{\alpha-1}(\zeta) = I_{\alpha+1}(\zeta) + (2\alpha/\zeta)I_{\alpha}(\zeta)$ to simplify the update rule for posterior variance.

Given filtered outputs $z_{k|k}$ and $\sigma_{k|k}^2$ obtained from the forward filtering algorithm, we perform backward smoothing based on the fixed interval smoothing algorithm of [7], computing the smoothed posterior modes $z_{k|K}$ and variances $\sigma_{k|K}^2$ for $k = 1,\ldots,K$ in a backward fashion as follows:

$$\begin{cases} z_{k-1|K} &= z_{k-1|k-1} + s_k(z_{k|K} - z_{k|k-1}) \\ \sigma_{k-1|K}^2 &= \sigma_{k-1|k-1}^2 + s_k^2(\sigma_{k|K}^2 - \sigma_{k|k-1}^2) \end{cases}, \tag{14}$$

where $s_k := \rho\sigma^2_{k-1|k-1}/\sigma^2_{k|k-1}$ is the backward smoothing gain factor. It should be noted that, unlike the forward filtering, the backward smoothing step is appropriate for off-line applications, as it refines the preceding filtered estimates $z_{k|k}$ using the future estimated states $i > k$.

Statistical confidence regions for both the filtered estimates $\widehat{z}_k^{\text{filtered}} \sim \mathcal{N}(z_{k|k}, \sigma^2_{k|k})$ and smoothed estimates $\widehat{z}_k^{\text{smoothed}} \sim \mathcal{N}(z_{k|K}, \sigma^2_{k|K})$ can be easily computed at each time step k. Consequently, the computation of confidence regions for $\widehat{\nu}_k^{\text{filtered}} = \exp\big(\widehat{z}_k^{\text{filtered}}\big)$ at a significance level η is straightforward as $\mathcal{CR}^{\nu}_k = \big[\exp\big(z_{k|k} \pm \Phi^{-1}(1-\eta/2)\sigma_{k|k}\big)\big]$, given the monotonicity of the exponential link function. Similarly, we can compute the confidence regions for the smoothed estimate $\widehat{\nu}_k^{\text{smoothed}} = \exp\big(\widehat{z}_k^{\text{smoothed}}\big)$.

Finally, we estimate the optimal variance $\widehat{\sigma}^2_e$ tuned to the complete set of deviance data $\{D^{\beta}_k\}_{k=1}^K$, via the Expectation Maximization (EM) algorithm [17]. We take $z_{1:K}$ as the set of latent variables for the EM algorithm. The E-step for $(l+1)$-th iteration of EM computes:

$$\mathbb{E}_{\mathbf{z}}\Big[\log p\big(D^{\beta}_{1:K}, z_{1:K}|\sigma^2_e\big)\big|D^{\beta}_{1:K}, \widehat{\sigma}^{2,(l)}_e\Big] = -\frac{K}{2}\log(\sigma^2_e) - \frac{1}{2\sigma^2_e}\sum_{k=1}^{K}\Big\{\Big(\sigma^2_{k|K}+z^2_{k|K}\Big)+\rho^2\Big(\sigma^2_{k-1|K}+z^2_{k-1|K}\Big) - 2\rho\Big(\sigma^2_{k-1,k|K} + z_{k-1|K}z_{k|K}\Big)\Big\}, \quad (15)$$

where $\mathbb{E}_{\mathbf{z}}\big[.|D^{\beta}_{1:K}, \widehat{\sigma}^{2,(l)}_e\big]$ denotes the expectation operator with respect to the latent variables given the complete set of deviance data $D^{\beta}_{1:K}$ and the current estimate of the parameter $\widehat{\sigma}^{2,(l)}_e$. It is noteworthy that calculation of the E-step involves computation of the smoothed means and variances $\mathbb{E}_{\mathbf{z}}\big[z^2_k \,|\, D^{\beta}_{1:K}, \widehat{\sigma}^{2,(l)}_e\big] = \sigma^2_{k|K}+z^2_{k|K}$, which is readily available from the non-central chi-squared smoothing algorithm (14), and covariance terms $\mathbb{E}_{\mathbf{z}}\big[z_{k-1}z_k|D^{\beta}_{1:K}, \widehat{\sigma}^{2,(l)}_e\big] = \sigma^2_{k-1,k|K} + z_{k-1|K}z_{k|K}$, which can be computed using a state-space covariance smoothing algorithm [18] as $\sigma^2_{k-1,k|K} = s_k\sigma^2_{k|K}$. The M-step can be solved easily for $\widehat{\sigma}^{2,(l+1)}_e$ by maximizing (15), which results in:

$$\widehat{\sigma}^{2,(l+1)}_e = \frac{1}{K}\sum_{k=1}^{K}\Big\{\Big(\sigma^2_{k|K}+z^2_{k|K}\Big)+\rho^2\Big(\sigma^2_{k-1|K}+z^2_{k-1|K}\Big) - 2\rho\Big(\sigma^2_{k-1,k|K} + z_{k-1|K}z_{k|K}\Big)\Big\}. \quad (16)$$

IV. Application to Simulated Spiking Data

We consider two simulated spike trains synthetically generated based on a network of two neurons as depicted in Figure 1–A (top row), in which the inhibitory and excitatory interactions are represented by black and white circles, respectively. As illustrated, a dynamic scenario is examined in which causal interactions evolve with time. Both neurons have self-inhibitory effects, and neuron (1) is influenced by neuron (2) through an excitatory interaction.

For the point process simulation setting, we consider a total observation period of $\mathcal{T} = 180s$ discretized to $K = 180000$ time bins of length $\Delta = 1\ ms$. The binary spike trains $\{\{n^{(c)}_k\}_{k=1}^K\}_{c=1,2}$ were generated for both neurons based on the dynamic CIF with Bernoulli spiking statistics. The firing probability $\lambda^{(c)}_k\Delta$ of each neuron is fitted by a dynamic GLM model with logistic link function and intrinsic covariates $\{n^{(c)}_i\}_{i=1}^{k-1}$ drawn from its own spiking history and the history of the other neuron, in the case that a causal link exists. The baseline firing parameter is selected to be the same for both neurons, $\mu_k = -2.09$ to set the average spiking rate to $\lambda_{avg}\Delta \approx 0.1 \ll 1$. The sparse modulation parameter vector $\boldsymbol{\omega}^{(1)}_k = [\mu_k, \boldsymbol{\omega}^{(1,1)'}_k, \boldsymbol{\omega}^{(1,2)'}_k]'$ of length $M^{(1)} = 101$ is associated with neuron (1), where $\boldsymbol{\omega}^{(1,1)}_k$ and $\boldsymbol{\omega}^{(1,2)}_k$ are sparse vectors of length 50 accounting for the self-inhibitory effect and excitatory causal link $(2 \mapsto 1)$, respectively, with respective support sets $\text{Supp}(\boldsymbol{\omega}^{(1,1)}_k) = \{1, 3, 15\}$ and $\text{Supp}(\boldsymbol{\omega}^{(1,2)}_k) = \{5, 10\}$ of size $s^{(1,1)} = 3$ and $s^{(1,2)} = 2$, and respective initial values of $\{-2, +1, +0.5\}$ and $\{-0.5, +1\}$. The parameter vector associated with neuron (2) is set to $\boldsymbol{\omega}^{(2)}_k = [\mu_k, \boldsymbol{\omega}^{(2,2)'}_k]'$ of length $M^{(2)} = 51$, with a sparse self-inhibitory vector $\boldsymbol{\omega}^{(2,2)}_k$ of size 50 with support set $\text{Supp}(\boldsymbol{\omega}^{(2,2)}_k) = \{1, 5, 10\}$ of size $s^{(2,2)} = 3$ and respective values of $\{-2, -1, +1\}$, and no causal influence from neuron (1). The norm of all sparse parameter vectors $\boldsymbol{\omega}^{(1,1)}_k, \boldsymbol{\omega}^{(1,2)}_k$ and $\boldsymbol{\omega}^{(2,2)}_k$ is normalized to 1. We assume the self-inhibitory effects to be static for both neurons (all the coefficients of $\boldsymbol{\omega}^{(c,c)}_k$ remain constant along time), and a dynamic causal influence evolving in time. We enforce a time evolution for the in-support coefficients of $\boldsymbol{\omega}^{(1,2)}_k$, where they remain constant for the first minute, followed by a linear decline to zero within the second minute, and remaining at zero in the final $1min$ segment of the simulation. We use the sparse adaptive filter ℓ_1-PPF$_1$ to estimate the sparse parameter vectors $\widehat{\boldsymbol{\omega}}_k$ at every time step k for both the full and reduced models. For the ℓ_1-PPF$_1$ filtering algorithm, a forgetting factor of $\beta = 0.9999$ is selected sufficiently close to one, and $L = 1$ number of iterations. The regularization parameter is tuned to $\gamma = 7$, obtained by the two-fold even-odd cross validation [6].

Figure 1–A (top row) depicts the three states of evolution for two-neuron network over the time, as the excitatory link $(2 \mapsto 1)$ gradually weakens, until it completely vanishes. In Figure 1–B, the simulated spiking data is plotted (blue vertical lines) for both neurons within windows of $0.2s$ selected from the end-points of all three segments of simulation. The subsequent four rows show the detected time-course of changes associated with dynamic causal link $(2 \mapsto 1)$ (Fig. 1–C and 1–D) and the non-existing link $(1 \mapsto 2)$ (Fig. 1–E and 1–F), obtained using the proposed non-central χ^2 filtering and smoothing algorithm tuned for two different smoothing settings: 1) the optimal smoothing factor is tuned to the data as $\widehat{\sigma}^2_e = 2 \times 10^{-5}$ using the EM algorithm (Figs 1–C and 1–E), and 2) the hand-tuned smooth setting where a small $\widehat{\sigma}^2_e = 10^{-9}$ is chosen (Figs 1–D and 1–F). Each panel (C through F) shows the forward-filtered estimates (blue) $\widehat{\nu}_k = \exp(z_{k|k})$ and the further backward-

smoothed estimates (red) $\widehat{\nu}_k = \exp(z_{k|K})$ as well as the raw estimate (black) $\widehat{\nu}_k \approx D_k^\beta - M^d$ obtained from the raw deviance data. The 95% confidence regions has been plotted along with the both filtered and smoothed estimates. For the χ^2 filtering and smoothing algorithm, we have $M^d = 50$, and the initials are chosen as $z_{0|0} = 10$ and $\sigma^2_{0|0} = 1$, and the scaling factor as $\rho = 1$. A careful inspection of Figure 1 reveals the striking performance of the proposed algorithm in detection and tracking both causal influences: in Fig. 1–E and 1–F, $\widehat{\nu}_k$ is estimated as zero, avoiding the false alarm detection, which implies that there is no causal influence from neuron (1) on neuron (2) throughout the simulation period. In Fig. 1–C and 1–D, the dynamic causal link $(2 \mapsto 1)$ is correctly detected and the underlying changes in the causal influence is precisely tracked across the time, as it implies that the excitatory causal influence from neuron (2) to neuron (1) is weakening in the second segment, and completely vanishing throughout the final segment of the experiment.

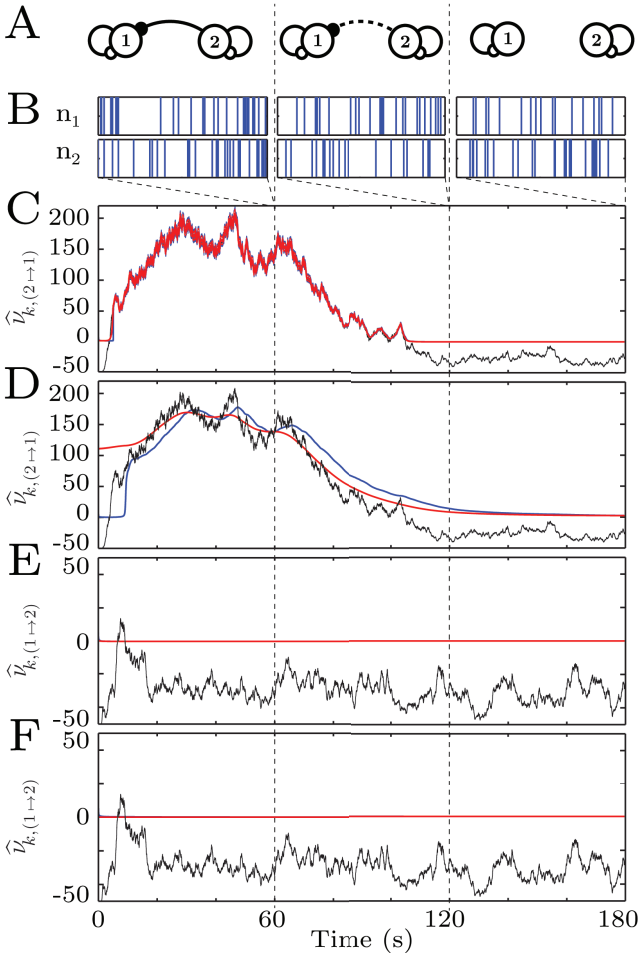

Fig. 1. The estimated non-centrality parameter across time representing the time-course of causal changes in between the two neurons using the proposed filtering and smoothing algorithm for two different settings. The filtered and smoothed estimates are shown by blue and red traces along with the raw estimate $D_k^\beta - M^d$ shown by gray traces, respectively. The colored hulls around the estimates show the 95% confidence regions.

V. Conclusion

In this paper, we considered recursive estimation of the time-varying Granger causality in an ensemble of neurons interacting through their cross-history interdependence. We introduced a novel measure of causality using the forgetting factor mechanism of RLS-type algorithms. We constructed recursive update rules for the computation of the causality measure, and obtained non-central chi-squared filtering and smoothing algorithms to denoise the estimated causality measure and to povide recursively computed statistical confidence regions. Application of our filtering and smoothing algorithms to simulated data shows that they are capable of identifying the dynamics of causal interactions with high accuracy.

References

[1] S. L. Bressler and A. K. Seth, "Wiener–granger causality: a well established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.

[2] N. Wiener, "The theory of prediction," *Modern mathematics for engineers*, vol. 1, pp. 125–139, 1956.

[3] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[4] M. Okatan, M. A. Wilson, and E. N. Brown, "Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity," *Neural computation*, vol. 17, no. 9, pp. 1927–1961, 2005.

[5] S. Kim, D. Putrino, S. Ghosh, and E. N. Brown, "A granger causality measure for point process models of ensemble neural spiking activity," *PLoS Comput Biol*, vol. 7, no. 3, p. e1001110, 2011.

[6] A. Sheikhattar, J. B. Fritz, S. A. Shamma, and B. Babadi, "Recursive sparse point process regression with application to spectrotemporal receptive field plasticity analysis," *IEEE Trans. on Signal Processing, in press. preprint: arXiv preprint arXiv:1507.04727*, 2016.

[7] A. Smith and E. N. Brown, "Estimating a state-space model from point process observations," *Neural Computation*, vol. 15, no. 5, pp. 965–991, 2003.

[8] A. Sheikhattar, J. B. Fritz, S. A. Shamma, and B. Babadi, "Adaptive sparse logistic regression with application to neuronal plasticity analysis," in *2015 Asilomar Conference on Signals, Systems, and Computers, Nov. 811, Pacific Grove, CA*, 2015.

[9] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm," *Signal Processing, IEEE Trans. on*, vol. 58, no. 8, pp. 4013–4025, 2010.

[10] J. F. Geweke, "Measures of conditional linear dependence and feedback between time series," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 907–915, 1984.

[11] A. J. Dobson and A. Barnett, *An introduction to generalized linear models*. CRC press, 2008.

[12] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.

[13] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical society*, vol. 54, no. 3, pp. 426–482, 1943.

[14] R. R. Davidson and W. E. Lever, "The limiting distribution of the likelihood ratio statistic under a class of local alternatives," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 209–224, 1970.

[15] H. Peers, "Likelihood ratio and associated test criteria," *Biometrika*, vol. 58, no. 3, pp. 577–587, 1971.

[16] A. Sheikhattar, J. B. Fritz, S. A. Shamma, and B. Babadi, "Dynamic analysis of causal influences in neuronal ensembles," *preprint*, 2016.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[18] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the em algorithm," *Journal of time series analysis*, vol. 3, no. 4, pp. 253–264, 1982.