Appendices for "Dynamic Analysis of Higher-Order Coordination in Neuronal Assemblies via De-Sparsified Orthogonal Matching Pursuit"

Shoutik Mukherjee and Behtash Babadi
Department of Electrical and Computer Engineering
Institute for Systems Research
University of Maryland, College Park
College Park, MD 20742
`{smukher2, behtash}@umd.edu`

A Derivations

This appendix contains the derivations of recursive update rules for the log-likelihood and its gradient for efficient implementations of AdOMP and computation of the adaptive de-biased deviance difference (Section A.1). The appendix also includes the derivations of the de-sparsified AdOMP estimate whose asymptotic behavior is characterized in Theorem 1 (Section A.2). Additionally, this appendix summarizes the forward/backward algorithm applied to obtain smooth baseline rate parameters for the null hypothesis (Section A.3).

A.1 Recursive Computation of the Gradient and Log-Likelihood

The AdOMP algorithm uses the gradient of the log-likelihood at each iteration of the matching pursuit to determine the next addition to the parameter support set and to solve the new maximization problem with a gradient descent algorithm. We derive a recursive update to compute the gradient at the k^{th} window.

The gradient of the objective subsumes the gradients with respect to each $\boldsymbol{\omega}^{(m)}$. That is, $\nabla_{\boldsymbol{\omega}} \ell_k^{\beta}(\boldsymbol{\omega}_k) = \left[\nabla_{\boldsymbol{\omega}^{(1)}} \ell_k^{\beta}(\boldsymbol{\omega}_k), \dots, \nabla_{\boldsymbol{\omega}^{(C^*)}} \ell_k^{\beta}(\boldsymbol{\omega}_k)\right]'$, where, for each $m = 1, \dots, C^*$,

$$\nabla_{\boldsymbol{\omega}^{(m)}} \ell_k^{\beta}(\boldsymbol{\omega}_k) = \sum_{i=1}^{k} \beta^{k-i} \nabla_{\boldsymbol{\omega}^{(m)}} \ell_i(\boldsymbol{\omega}_k) = \sum_{i=1}^{k} \beta^{k-i} \boldsymbol{X}_i{}' \left[\boldsymbol{n}_i^{*(m)} - \boldsymbol{\lambda}_i^{*(m)}(\boldsymbol{\omega}_k)\Delta\right]. \tag{14}$$

The m^{th} mark's CIF over the i^{th} window is denoted by $\boldsymbol{\lambda}_i^{*(m)}(\boldsymbol{\omega}_k)\Delta = [\lambda_{1+(i-1)W}^{*}{}^{(m)}(\boldsymbol{\omega}_k)\Delta, \dots, \lambda_{iW}^{*}{}^{(m)}(\boldsymbol{\omega}_k)\Delta]'$, and the common term $(1-\beta)$ is dropped for convenience, as it does not affect the optimal solution. Observe that the gradient, evaluated at the l^{th} iteration $\hat{\boldsymbol{\omega}}_{(l),k}$, can equivalently be written as

$$\nabla_{\boldsymbol{\omega}} \ell_k^{\beta}(\hat{\boldsymbol{\omega}}_{(l),k}) = \beta \left(\nabla_{\boldsymbol{\omega}} \ell_{k-1}^{\beta}(\hat{\boldsymbol{\omega}}_{(l),k})\right) + \nabla_{\boldsymbol{\omega}} \ell_k(\hat{\boldsymbol{\omega}}_{(l),k}), \tag{15}$$

suggesting a recursive update. However, this definition requires the value of the gradient (and consequentially, the CIFs) from the previous window evaluated at $\hat{\boldsymbol{\omega}}_{(l),k}$, which is unavailable in an online setting. However, the values $\boldsymbol{\lambda}_i^{*}(\hat{\boldsymbol{\omega}}_i)\Delta$ have been evaluated at previous windows; thus, each $\boldsymbol{\lambda}_i^{*}(\hat{\boldsymbol{\omega}}_{(l),k})\Delta$ is approximated by its first-order Taylor approximation around $\hat{\boldsymbol{\omega}}_i$. In the following, we consider only the m^{th} mark, noting that each component of the gradient may be obtained by applying the procedure to each mark in parallel.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

First, the aforementioned Taylor approximation is computed for the CIFs of the m^{th} mark at window i^{th}. Defining the diagonal matrix

$$\boldsymbol{\Lambda}_i^{*(m,j)} := \begin{cases} diag\left(\boldsymbol{\lambda}_i^{*(m)}\Delta \odot (1 - \boldsymbol{\lambda}_i^{*(j)}\Delta)\right), & m = j \\ \\ diag\left(\boldsymbol{\lambda}_i^{*(m)}\Delta \odot \boldsymbol{\lambda}_i^{*(j)}\Delta\right), & m \neq j \end{cases} \tag{16}$$

with diagonal terms given by the specified point-wise products, we find the first-order Taylor approximation to be

$$\boldsymbol{\lambda}_i^{*(m)}(\hat{\boldsymbol{\omega}}_{(l),k})\Delta \approx \boldsymbol{\lambda}_i^{*(m)}(\hat{\boldsymbol{\omega}}_i)\Delta + \sum_{j=1}^{C^*} \boldsymbol{\Lambda}_i^{*(m,j)}(\hat{\boldsymbol{\omega}}_i)\ \boldsymbol{X}_i \left(\hat{\boldsymbol{\omega}}_{(l),k}^{(j)} - \hat{\boldsymbol{\omega}}_i^{(j)}\right). \tag{17}$$

Note that evaluating the approximation of the CIF for the m^{th} requires the CIFs of the other marks through $\boldsymbol{\Lambda}_i^{*(m,j)}$. Making the simplifying assumption that $(\lambda_i^{*(m)}\Delta)\cdot(\lambda_i^{*(j)}\Delta) = o(\Delta)$, $m \neq j$, the cross-terms are rendered negligible and the parallelized CIF approximation is given by

$$\boldsymbol{\lambda}_{(l),k}^{*\ (m)}\Delta := \boldsymbol{\lambda}_i^{*(m)}(\hat{\boldsymbol{\omega}}_{(l),k})\Delta \approx \boldsymbol{\lambda}_i^{*(m)}(\hat{\boldsymbol{\omega}}_i)\Delta + \boldsymbol{\Lambda}_i^{*(m,m)}(\hat{\boldsymbol{\omega}}_i)\ \boldsymbol{X}_i \left(\hat{\boldsymbol{\omega}}_{(l),k}^{(m)} - \hat{\boldsymbol{\omega}}_i^{(m)}\right). \tag{18}$$

Substituting this approximation, the gradient evaluated at $\hat{\boldsymbol{\omega}}_{(l),k}$ becomes

$$\boldsymbol{g}_{(l),k}^{(m)} := \nabla_{\boldsymbol{\omega}^{(m)}} \ell_k^{\beta}(\hat{\boldsymbol{\omega}}_{(l),k}) \approx \sum_{i=1}^{k} \beta^{k-i}\left[\boldsymbol{X}_i{}'\boldsymbol{\varepsilon}_{(l),i}^{(m)} - \boldsymbol{X}_i{}'\boldsymbol{\Lambda}_i^{*(m,m)}\boldsymbol{X}_i\left(\hat{\boldsymbol{\omega}}_{(l),k}^{(m)} - \hat{\boldsymbol{\omega}}_i^{(m)}\right)\right], \tag{19}$$

where $\boldsymbol{\varepsilon}_{(l),i}^{(m)} := \boldsymbol{n}_i^{*(m)} - \boldsymbol{\lambda}_{(l),i}^{*\ (m)}\Delta$. Defining the quantities

$$\begin{aligned} \boldsymbol{b}_k^{(m)} &:= \beta \boldsymbol{b}_{k-1}^{(m)} + \boldsymbol{X}_k{}'\boldsymbol{\varepsilon}_{(l),k}^{(m)} + \boldsymbol{X}_k{}'\boldsymbol{\Lambda}_k^{*(m,m)}(\hat{\boldsymbol{\omega}}_{(l),k})\boldsymbol{X}_k\hat{\boldsymbol{\omega}}_{(l),k}^{(m)}, \\ \boldsymbol{B}_k^{(m)} &:= \beta \boldsymbol{B}_{k-1}^{(m)} + \boldsymbol{X}_k{}'\boldsymbol{\Lambda}_k^{*(m,m)}(\hat{\boldsymbol{\omega}}_{(l),k})\boldsymbol{X}_k, \end{aligned} \tag{20}$$

the parallelized, fully recursive gradient update rule for the m^{th} mark is thus

$$\begin{aligned} \boldsymbol{g}_{(l),k}^{(m)} =& \boldsymbol{b}_k^{(m)} - \boldsymbol{B}_k^{(m)}\hat{\boldsymbol{\omega}}_{(l),k}^{(m)} \\ =& \beta\left(\boldsymbol{b}_{k-1}^{(m)} - \boldsymbol{B}_{k-1}^{(m)}\hat{\boldsymbol{\omega}}_{(l),k}^{(m)}\right) + \boldsymbol{X}_k{}'\boldsymbol{\varepsilon}_{(l),k}^{(m)}. \end{aligned} \tag{21}$$

The AdOMP algorithm, and the utility of the recursive gradient update, is detailed in Algorithm 1. Let the initial support set of size r_0 be denoted by $S^{(0)}$, and the maximum support size by s^*. The maximum support size can be obtained by cross-validation. We assume that $S^{(0)}$ is non-empty here, where a reasonable initialization is made; it is then necessary to solve the preliminary problem in Step 3. Alternatively, we may initially set $S^{(0)} = \emptyset$ and $\hat{\boldsymbol{\omega}}_{(0),k} = \mathbf{0}$.

Additionally, the log-likelihoods of the full and reduced models can be computed in an online fashion. We use the Taylor approximation of $\ell_i(\hat{\boldsymbol{\omega}}_k)$ around $\hat{\boldsymbol{\omega}}_i$ to obtain the following recursion:

$$\ell_k^{\beta}(\hat{\boldsymbol{\omega}}_k) = a_k + \sum_{m=1}^{C^*}\left(\hat{\boldsymbol{\omega}}_k^{(m)\prime}\boldsymbol{b}_k^{(m)} - \frac{1}{2}\hat{\boldsymbol{\omega}}_k^{(m)\prime}\boldsymbol{B}_k^{(m)}\hat{\boldsymbol{\omega}}_k^{(m)}\right), \tag{22}$$

where the variable a_k is defined as

$$a_k := \beta a_{k-1} + \ell_k(\hat{\boldsymbol{\omega}}_k) - \sum_{m=1}^{C^*}\left(\hat{\boldsymbol{\omega}}_k^{(m)\prime}\boldsymbol{X}_k'\boldsymbol{\varepsilon}_k^{(m)} + \frac{1}{2}\hat{\boldsymbol{\omega}}_k^{(m)\prime}\boldsymbol{X}_k'\boldsymbol{\Lambda}^{*}{}_k^{(m)}\boldsymbol{X}_k\hat{\boldsymbol{\omega}}_k^{(m)}\right), \tag{23}$$

and $\boldsymbol{b}_k$, $\boldsymbol{B}_k$ have been previously defined. The bias term $\mathscr{B}_k = \left(\nabla\ell_k^{\beta}(\hat{\boldsymbol{\omega}}_k)\right)'\left(\nabla^2\ell_k^{\beta}(\hat{\boldsymbol{\omega}}_k)\right)^{-1}\left(\nabla\ell_k^{\beta}(\hat{\boldsymbol{\omega}}_k)\right)$ can also be computed recursively. The recursive computations of the log-likelihood and bias terms enable the adaptive de-biased deviance difference to be computed efficiently.

Algorithm 1 AdOMP : **Ad**aptive **O**rthogonal **M**atching **P**ursuit

Input: $\{\boldsymbol{n}_k^{*(m)}\}_{m=1}^{C^*}$, $\boldsymbol{X}_k$, $\{\boldsymbol{b}_{k-1}^{(m)}\}_{m=1}^{C^*}$, $\{\boldsymbol{B}_{k-1}^{(m)}\}_{m=1}^{C^*}$, $S^{(0)}$, s^*, β

Output: $\hat{\boldsymbol{\omega}}_k$, $\{\boldsymbol{b}_k^{(m)}\}_{m=1}^{C^*}$, $\{\boldsymbol{B}_k^{(m)}\}_{m=1}^{C^*}$

$r_0 = |S^{(0)}|$

$S_k^{(r_0)} = S^{(0)}$

$\hat{\boldsymbol{\omega}}_{(r_0),k} = \arg\max_{\text{supp}(\boldsymbol{\omega}_k) \subseteq S_k^{(r_0)}} \ell_k^{\beta}(\boldsymbol{\omega}_k)$

for $r = r_0 + 1$ to s^* **do**

  for $m = 1$ to C^* **do**

    $\boldsymbol{g}_{(r),k}^{(m)} = \beta \boldsymbol{g}_{(r-1),k-1}^{(m)} + \boldsymbol{X}_k{}' \boldsymbol{\varepsilon}_{(r-1),k}^{(m)}$

  end for

  $j = \arg\max_{i \notin S_k^{(r)}} \left| \{g_{(r),k}^{(i)}\} \right|$

  $S_k^{(r)} = S_k^{(r-1)} \cup \{j\}$

  $\hat{\boldsymbol{\omega}}_{(r),k} = \arg\max_{\text{supp}(\boldsymbol{\omega}_k) \subseteq S_k^{(r)}} \ell_k^{\beta}(\boldsymbol{\omega}_k)$

end for

for $m = 1$ to C^* **do**

  $\boldsymbol{\Lambda}_{(s^*),k}^{*}{}^{(m)} = diag\left(\boldsymbol{\lambda}_{(s^*),k}^{*}{}^{(m)} \Delta \odot (1 - \boldsymbol{\lambda}_{(s^*),k}^{*}{}^{(m)} \Delta)\right)$

  $\boldsymbol{b}_k^{(m)} = \beta \boldsymbol{b}_{k-1}^{(m)} + \boldsymbol{X}_k{}' \boldsymbol{\varepsilon}_{(s^*),k}^{(m)} + \boldsymbol{X}_k{}' \boldsymbol{\Lambda}_{(s^*),k}^{*}{}^{(m)} \boldsymbol{X}_k \hat{\boldsymbol{\omega}}_{(s^*),k}^{(m)}$

  $\boldsymbol{B}_k^{(m)} = \beta \boldsymbol{B}_{k-1}^{(m)} + \boldsymbol{X}_k{}' \boldsymbol{\Lambda}_{(s^*),k}^{*}{}^{(m)} \boldsymbol{X}_k$

end for

return $\hat{\boldsymbol{\omega}}_k = \hat{\boldsymbol{\omega}}_{(s^*),k}$, $\{\boldsymbol{b}_k^{(m)}\}_{m=1}^{C^*}$, $\{\boldsymbol{B}_k^{(m)}\}_{m=1}^{C^*}$

A.2 De-Sparsifying the AdOMP Estimate

We show that the AdOMP parameter estimates can be de-sparsified, as in [1], by inspecting the Karush-Kuhn-Tucker (KKT) conditions of the optimization problem. Note that $\hat{\boldsymbol{\omega}}_k = \hat{\boldsymbol{\omega}}_{(s^*),k}$, solves the maximization problem

$$\hat{\boldsymbol{\omega}}_{(s^*),k} = \underset{\text{supp}(\boldsymbol{\omega}_k) \subseteq S_k^{(s^*)}}{\arg\max} \ \ell_k^{\beta}(\boldsymbol{\omega}_k). \tag{24}$$

The support constraint is equivalent to requiring the parameter's i^{th} component $\omega_k^{(i)} = 0, i \notin S_k^{(s^*)}$. Collectively, these linear equality constraints can be expressed as $A_k^{(s^*)} \boldsymbol{\omega}_k = \mathbf{0}$, where the diagonal matrix $\left[A_k^{(s^*)}\right]_{i,i} = 0$ for $i \in S_k^{(s^*)}$ and 1 otherwise. The KKT conditions on the primal and dual parameters ($\boldsymbol{\omega}_k$ and $\boldsymbol{\nu}_k^{(s^*)}$, respectively) are straightforward to derive. Of particular relevance is the stationarity condition:

$$\nabla \ell_k^{\beta}(\boldsymbol{\omega}_k) - A_k^{(s^*)} \boldsymbol{\nu}_k^{(s^*)} = \mathbf{0}. \tag{25}$$

Substituting $\ell_k^{\beta}(\boldsymbol{\omega}_k)$ for its quadratic approximation around $\hat{\boldsymbol{\omega}}_{(s^*),k}$ into the stationarity condition and rearranging terms, we thus define the de-sparsified parameters as

$$\hat{\boldsymbol{w}}_k := \hat{\boldsymbol{\omega}}_{(s^*),k} - \left(\nabla^2 \ell_k^{\beta}(\hat{\boldsymbol{\omega}}_{(s^*),k})\right)^{-1} \left(\nabla \ell_k^{\beta}(\hat{\boldsymbol{\omega}}_{(s^*),k})\right), \tag{26}$$

in the same fashion as van de Geer et al. in [1]. The asymptotic normality of the de-sparsified AdOMP parameters is established in Theorem 1.

A.3 Forward Filtering and Backward Smoothing of $\hat{\gamma}_k^{(m)}$

To quantitatively characterize the null hypothesis that coordinated r^{th}-order spiking occurs at the same rate as between r independent neurons, recall that the reduced model is estimated by fixing $\mu_k^{(m)}$ at $\hat{\mu}_k^{(m)} - \hat{\gamma}_k^{(m)}$ for $m \in \mathcal{K}_r$. However, the estimated exogenous factors $\hat{\gamma}_k^{(m)}$ can vary abruptly between adjacent windows, noisily reflecting the null hypothesis. Hence, we apply Kalman forward/backward smoothing [2] for stability; the procedure is summarized in Algorithm 2 for the estimated exogenous factors of the m^{th} mark, $\{\hat{\gamma}_k^{(m)}\}_{k=1}^K$.

The estimated exogenous factors are smoothed independently, assuming a linear Gaussian forward model for each, as described in Eq. (27). The mark index is dropped here for clarity.

$$\begin{aligned} \hat{\gamma}_k &= x_k + v_k && v_k \sim \mathcal{N}(0, \sigma_v^2) \\ x_k &= x_{k-1} + w_k && w_k \sim \mathcal{N}(0, \sigma_w^2) \end{aligned} \tag{27}$$

Algorithm 2 Smoothing $\{\hat{\gamma}_k\}_{k=1}^K$ via Kalman Forward/Backward Algorithm

Input: $\{\hat{\gamma}_k\}_{k=1}^K$, $\sigma^2_{v,(0)}$, $\sigma^2_{w,(0)}$, L
Output: $\{\breve{\gamma}_k\}_{k=1}^K$, $\{\tilde{\gamma}_k\}_{k=1}^K$, $\sigma^2_{v,(L)}$, $\sigma^2_{w,(L)}$
1: **for** $l = 1$ to L **do**
2: $\quad\sigma^2_{0|0} = 1, x_{0|0} = 0$
3: $\quad$**for** $k = 1$ to K **do**
4: $\quad\quad x_{k|k-1} = x_{k-1|k-1}$
5: $\quad\quad \sigma^2_{k|k-1} = \sigma^2_{k-1|k-1} + \sigma^2_{w,(l-1)}$
6: $\quad\quad x_{k|k} = x_{k|k-1} + \frac{\sigma^2_{k|k-1}}{\sigma^2_{k|k-1}+\sigma^2_{v,(l-1)}}\left(\hat{\gamma}_k - x_{k|k-1}\right)$
7: $\quad\quad \sigma^2_{k|k} = \sigma^2_{k|k-1}\frac{\sigma^2_{v,(l-1)}}{\sigma^2_{k|k-1}+\sigma^2_{v,(l-1)}}$
8: $\quad$**end for**
9: $\quad$**for** $k = K$ to 1 **do**
10: $\quad\quad x_{k-1|K} = x_{k-1|k-1} + \frac{\sigma^2_{k-1|k-1}}{\sigma^2_{k|k-1}}\left(x_{k|K} - x_{k|k-1}\right)$
11: $\quad\quad \sigma^2_{k-1|K} = \sigma^2_{k-1|k-1} + \left(\frac{\sigma^2_{k-1|k-1}}{\sigma^2_{k|k-1}}\right)^2\left(\sigma^2_{k|K} - \sigma^2_{k|k-1}\right)$
12: $\quad$**end for**
13: $\quad\sigma^2_{w,(l)} = \frac{1}{K}\sum_{k=1}^K\left[\left(\hat{\gamma}_k - x_{k|K}\right)^2 + \sigma^2_{k|K}\right]$
14: $\quad\sigma^2_{w,(l)} = \frac{1}{K}\sum_{k=1}^K\left[\left(x_{k|K} - x_{k-1|K}\right)^2 + \left(1 - 2\frac{\sigma^2_{k-1|k-1}}{\sigma^2_{k|k-1}}\right)\sigma^2_{k|K} + \sigma^2_{k-1|K}\right]$
15: **end for**
16: $\{\breve{\gamma}_k\}_{k=1}^K = \{x_{k|k}\}_{k=1}^K$ and $\{\tilde{\gamma}_k\}_{k=1}^K = \{x_{k|K}\}_{k=1}^K$
17: **return** $\{\breve{\gamma}_k\}_{k=1}^K$, $\{\tilde{\gamma}_k\}_{k=1}^K$, $\sigma^2_{v,(L)}$, $\sigma^2_{w,(L)}$

Forward filtering is described in Steps 3–8, and backward smoothing in Steps 9–12. The Expectation-Maximization algorithm is used to update the initial values of the noise variances σ_w^2 and σ_v^2 over L iterations The backward-smoothed estimated exogenous factors, $\{\tilde{\gamma}_k\}_{k=1}^K$, were utilized to test for significant higher-order coordination.

B Supplementary Results

This appendix contains supporting results that demonstrate: the effects of varying the window length and forgetting factor (Section B.1); a statistical measure of model goodness-of-fit (Section B.2); and the utility of the algorithm in analyzing ensembles with more complex latent dynamics (Section B.3).

B.1 Varying Hyperparameters W and β

The hyperparameters W and β together control the effective integration window of the proposed adaptive model, defined as $\frac{W}{1-\beta}$. The role of the effective integration window is demonstrated in relation to the simulated example in Section 5.1. The simulated data is analyzed with varying values of β, as shown in Figure 1.

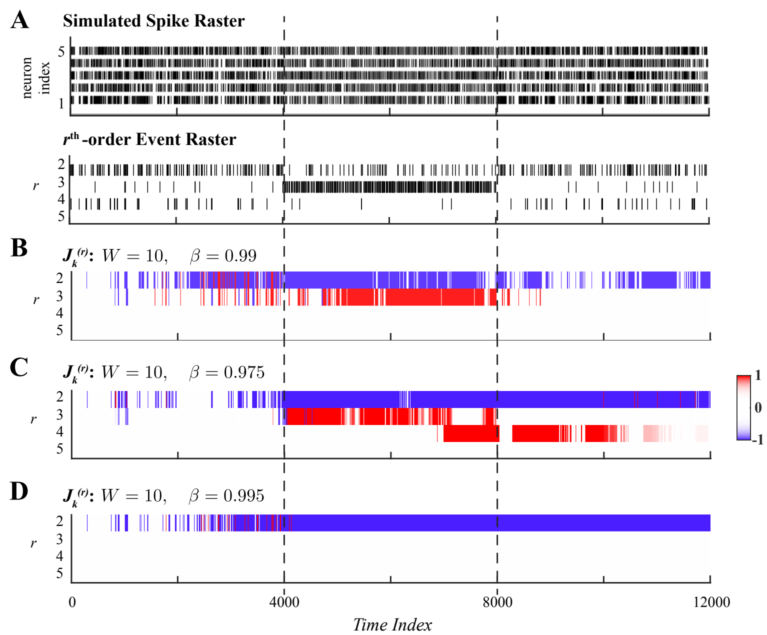

Figure 1: Varying W and β in history-dependent analysis of simulated data. **A.** Simulated ensemble spiking of five neurons (top) and the sum of the r^{th}-order simultaneous spiking events for $r = 2, 3, 4, 5$ (bottom). **B.** Significant r^{th}-order coordination with $W = 10$ and $\beta = 0.99$ (see also Fig. 2 in main text). **C.** Significant r^{th}-order coordination with $W = 10$ and $\beta = 0.975$. **D.** Significant r^{th}-order coordination with $W = 10$ and $\beta = 0.995$. Testing in **B**–**D** performed at level $\alpha = 0.001$.

The simulated spiking and rasters of r^{th}-order events for the example are reproduced in Figure 1–A, as are the main results of history-dependent analysis of higher-order coordination (Fig. 1–B) for which hyperparameters $W = 10$ and $\beta = 0.99$ were used. We first compare the main simulated results to an analysis with $\beta = 0.975$ (Fig. 1–C). The choice of smaller β results in faster dynamics: note that 3^{rd}-order coordination is detected at the start of the second epoch. However, as a consequence, the analysis is less stable (noting 3^{rd}-order coordination later in the second epoch) and more prone to false detection over short intervals (4^{th}-order coordination). Next we compare the main simulated results to an analysis with $\beta = 0.995$ (Fig. 1–D). The choice of larger β results in more stable detection of coordination over time (i.e. less jitter in 2^{nd}-order coordination), but has slower dynamics that make the analysis less sensitive to fast state transitions; consequently, no 3^{rd}-order coordination is detected.

B.2 Model Goodness-of-Fit

The goodness-of-fit of the estimated multinomial GLM for ensemble activity can be assessed by invoking the multivariate generalization of the time-rescaling theorem. For univariate point processes, the theorem states that interspike intervals (ISIs) rescaled with respect to the CIF (see [3] for details) are independent an exponentially distributed with unit rate. Hence, the uniformity of the ISIs rescaled by the estimated CIFs is assessed by comparison to uniform quantiles via the Kolmogorov-

Smirnov (KS) test. Additionally, the uncorrelatedness of empirically rescaled ISIs is assessed by the autocorrelation function (ACF) test.

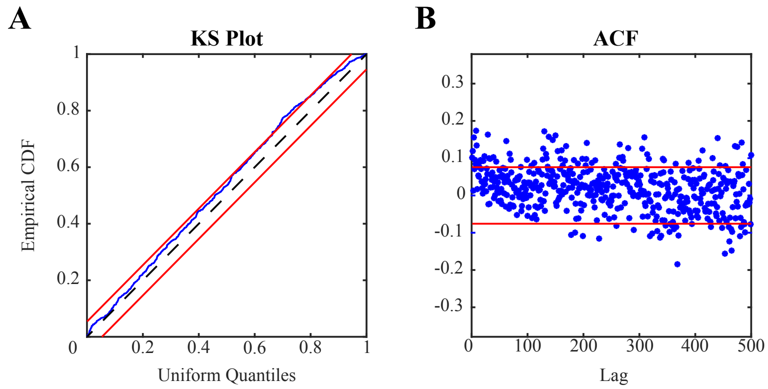

Figure 2: Assessing Goodness-of-Fit of History-Dependent Model in Simulation. **A.** Kolmogorov-Smirnov plot for mark index $m = 8$ with 95% confidence intervals (red). **B.** Autocorrelation Function for mark index $m = 8$ with 95% confidence intervals (red).

The time-rescaling theorem can be extended to multivariate point processes [4, 5], and in its generalization establishes that time-rescaling renders the marks mutually independent and rescaled marked ISIs (the interval between events of the same mark) are independent and exponentially distributed with unit rate. Consequently, the KS and ACF tests may be applied to each mark in order to assess goodness-of-fit. An example of the KS and ACF tests are shown in Figure 2 for the history-dependent analysis of simulated ensemble activity in Section 5.1 of the main text in the case of one mark.

B.3 Simulated Ensemble Spiking Data: Example 2

We present a second simulated example in which the ensemble spiking of 5 neurons was generated with more complex latent dynamics, shown in Figure 3. In the first simulated epoch (time bins $0 - 4000$), 4^{th}-order spiking events were excited by amplifying the default history-modulation parameters. In the second (bins $4000 - 6000$), no simultaneous spiking events were excited by history effects or exogenous influences. However, during the third epoch (bins $6000 - 8000$), 3^{rd}-order spiking events were induced by increasing the base rate parameter. The base rate parameter was similarly increased for 3^{rd}-order spiking events in the fourth epoch (bins $8000 - 12000$), while 4^{th}-order spiking events where concurrently excited by ensemble spiking history in the same manner as during the first epoch.

The simulated spiking activity, shown in Figure 3–A, does not observably reflect these latent dynamics precisely, though related changes in firing rate can be seen. The aggregate r^{th}-order marks (Fig. 3–B), however, do show increased rates of 3^{rd}- and 4^{th}-order spiking events.

Statistical analyses of r^{th}-order coordination for $r = 2, \ldots, 5$ using the history-independent model ($W = 10$; $\beta = 0.975$) indicates facilitated 3^{rd}-order coordination during the third and fourth epochs by large positive values of the J-statistics (Fig. 3–C). Facilitated 4^{th}-order coordination is also detected during the first and fourth epochs. By analyzing ensemble spiking using the history-dependent model ($W = 10$; $\beta = 0.99$) (Fig. 3–D), conditional facilitation of 3^{rd}-order coordination was correctly detected during the third and fourth epochs while 4^{th}-order coordination was correctly conditioned out. The history-dependent analysis also detected conditional suppression of 2^{nd}-order coordination.

The three control measures, however, are unable to capture the underlying dynamics. Significant pairwise correlations (Fig. 3–E) are indicated in each epoch except the second, but vary in magnitude with the average firing rate rather than based on the latent dynamics. Similarly, the spiking regularity (Fig. 3–F) indicates Poisson spiking statistics rather than coordinated activity. The 3^{rd}- and 4^{th}-order mark CIF differences (Fig. 3–G) once again only weakly reflect the changing rates of simultaneous events; closer inspection re-emphasizes the oscillatory nature of this measure that diminishes its reliability (Fig. 3–G, insets).

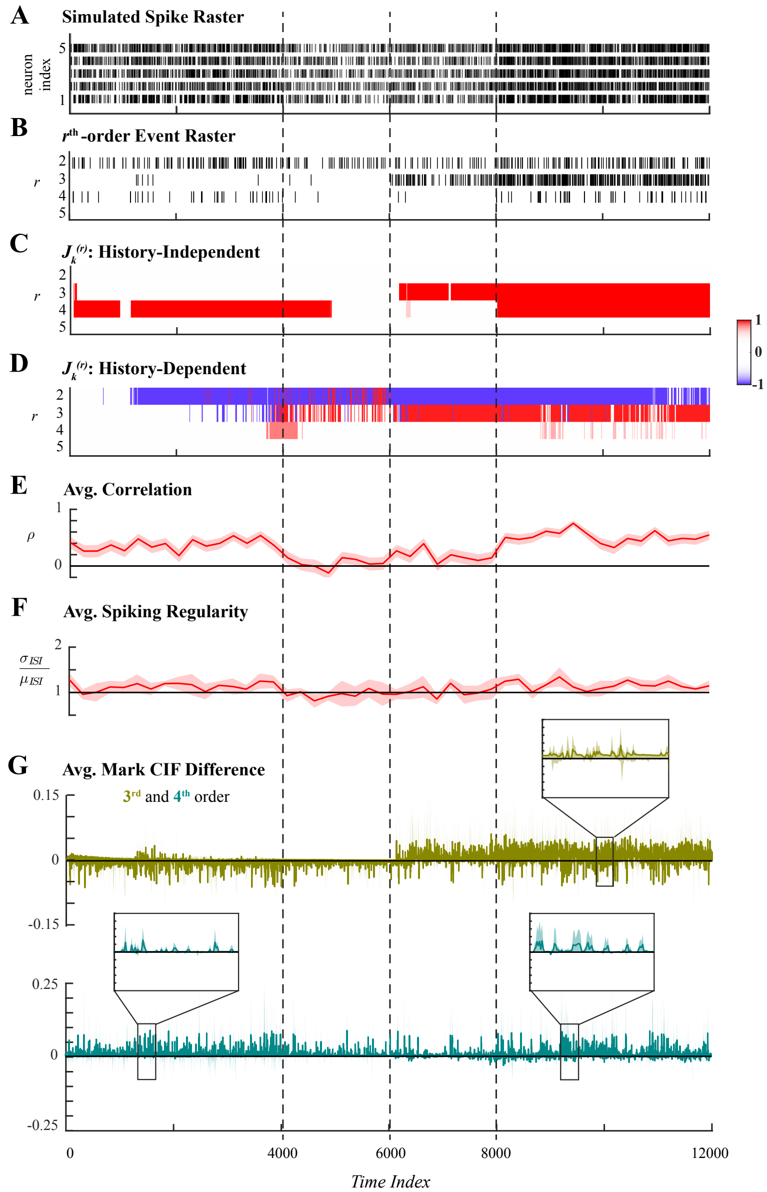

Figure 3: Analysis of ensemble spiking with concurrent 3$^{\text{rd}}$- and 4$^{\text{th}}$-order coordination. **A.** Simulated ensemble spiking of five neurons. **B.** Sum of the r^{th}-order simultaneous spiking events for $r = 2, 3, 4, 5$. Spiking coordination varies across 4 epochs, demarcated by vertical dashed lines. **C.** Significant r^{th}-order coordination neglecting ensemble history. **D.** Significant r^{th}-order coordination based on history-dependent ensemble spiking model. Statistical testing in **C**–**D** performed at level $\alpha = 0.001$. **E.** Average Pearson correlation with 95% confidence interval. **F.** Average spiking regularity: coefficient of variation ± 2 SEM. **G.** Average mark CIF differences of 3$^{\text{rd}}$- (green) and 4$^{\text{th}}$-order (teal) spiking interactions ± 2 SEM.

C Theoretical Results for the Statistical Inference of Higher-Order Spiking Coordination

The procedure for analyzing r^{th}-order coordination is summarized in Algorithm 3 below. Inferring coordinated spiking uses the limiting distributions of the adaptive de-biased deviance difference. This appendix contains the details of our theoretical analysis of limiting distributions. First, preliminary results are established. Namely, the limiting behavior of the Hessian matrix, $\nabla^2 \ell_k^{\beta}(\boldsymbol{\omega}_k^0)$, and gradient of the exponentially-weighted total data log-likelihood, $\nabla \ell_k^{\beta}(\boldsymbol{\omega}_k^0)$, are characterized in Section C.1. Theorem 1, which concerns asymptotic normality of de-sparsified estimates, is proven in Section C.2. Similarly, the limiting distributions of the adaptive de-biased deviance difference, i.e. Theorem 2, is proven in Section C.3. Finally, in Section C.4, testing for synchrony with contemporaneous-event models is addressed directly as a corollary to Theorem 2.

C.1 Preliminaries

The limiting behaviors of $\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0)$ and $\nabla \ell_k^\beta(\boldsymbol{\omega}_k^0)$ are characterized in Propositions 1 and 2, respectively. Limits are evaluated over a monotonically increasing sequence $\{\beta_l\}_{l=1}^\infty$ that converges to 1; the shorthand $\beta \to 1$ is used throughout for notational convenience. For simplicity, we analyze the case that the windows over which the model is piece-wise constant are of size $W = 1$ and indexed $i = k - N + 1, \ldots, k$.

Proposition 1. *As* $\beta \to 1$*,*

$$\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0) \to \mathbb{E}\left[\nabla^2 \ell_i(\boldsymbol{\omega}_k)\right] =: -\boldsymbol{\mathcal{I}}_k$$

Proof. First, note the following decomposition of the Hessian:

$$\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0) = (1-\beta) \sum_{i=k-N+1}^{k} \beta^{k-i} \nabla^2 \ell_i(\boldsymbol{\omega}_k^0).$$

Taking $\boldsymbol{V}_i = \beta^{k-i}\nabla^2 \ell_i(\boldsymbol{\omega}_k^0)$ and the sequence $c_\beta := \frac{1}{1-\beta}$ that tends to ∞ as $\beta \to 1$, a version of the LLN for ϕ-mixing random variables, [6], is invoked to find that

$$\begin{aligned}
\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0) - \mathbb{E}\left[\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0)\right] &= (1-\beta) \sum_{i=k-N+1}^{k} \left(\beta^{k-i}\nabla^2 \ell_i(\boldsymbol{\omega}_k^0) - \beta^{k-i}\mathbb{E}\left[\nabla^2 \ell_i(\boldsymbol{\omega}_k^0)\right]\right) \\
&= c_\beta^{-1} \sum_{i=k-N+1}^{k} \left(\boldsymbol{V}_i - \mathbb{E}\left[\boldsymbol{V}_i\right]\right) \xrightarrow{a.s.} 0.
\end{aligned}$$

Algorithm 3 Dynamic Analysis of r^{th}-Order Spiking Coordination

Input: $\{\boldsymbol{n}_k^*\}_{k=1}^K$, $\{\boldsymbol{X}_k\}_{k=1}^K$, r, β, α
Output: $\{J_k^{(r)}\}_{k=1}^K$, $\{\hat{\nu}_k^{(r)}\}_{k=1}^K$, $\{D_{k,\beta}^{(r)}\}_{k=1}^K$

```
𝒦_r = {m ∈ 𝒦 : Σ_{c=1}^{C} m_c = r} and M^(r) = |𝒦_r|
for k = 1 to K do
    h_k = 0
    Estimate ω̂_k^(F) using AdOMP; evaluate ℓ_k^β(ω̂_k^(F)) and ℬ_k^(F)
    for m ∈ 𝒦_r do
        Evaluate {u_t^(m)}_{t=(k-1)W+1}^{kW} and {u_{0,t}^(m)}_{t=(k-1)W+1}^{kW}
        Set γ̂_k^(m) = (1/W) Σ_{t=(k-1)W+1}^{kW} (u_t^(m) − u_{0,t}^(m)) and μ_{0,k}^(m) = μ̂_k^(m) − γ̂_k^(m)
    end for
    Estimate ω̂_k^(R) using AdOMP with constraint μ_k^(m) = μ_{0,k}^(m) for m ∈ 𝒦_r
    Evaluate ℓ_k^β(ω̂_k^(R)), ℬ_k^(R), and D_{k,β}^(r)(ω̂_k^(F), ω̂_k^(R)) as defined in Eq. (12) in the main text
    if F^{-1}_{χ²(M^(r))}(1 − α) < D_{k,β}^(r)(ω̂_k^(F), ω̂_k^(R)) then
        h_k = sgn(Σ_{m∈𝒦_r} γ̂_k^(m))
    end if
end for
Estimate {ν̂_k^(r)}_{k=1}^K via non-central χ² filtering/smoothing
J_k^(r) = h_k × (1 − α − F_{χ²(M^(r), ν̂_k)}(F^{-1}_{χ²(M^(r))}(1 − α)))
return {J_k^(r)}_{k=1}^K, {ν̂_k^(r)}_{k=1}^K, {D_{k,β}^(r)}_{k=1}^K
```

The limiting behavior of the Hessian is thus given by

$$\begin{aligned}\mathbb{E}\left[\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0)\right] &= \lim_{\beta\to 1}(1-\beta)\sum_{i=1}^{k}\beta^{k-i}\mathbb{E}\left[\nabla^2 \ell_i(\boldsymbol{\omega}_k^0)\right] \\ &= \mathbb{E}\left[\nabla^2 \ell_i(\boldsymbol{\omega}_k^0)\right].\end{aligned}$$

From the Fisher information equality, we have that $\mathbb{E}\left[\nabla^2 \ell_i(\boldsymbol{\omega}_k^0)\right] = -\boldsymbol{\mathcal{I}}_k$ □

Next, we establish the asymptotic normality of the score function.

Proposition 2. *As* $\beta \to 1$,

$$\sqrt{\frac{1+\beta}{1-\beta}}\nabla \ell_k^\beta(\boldsymbol{\omega}_k^0) \to \mathcal{N}\left(\mathbf{0}, \boldsymbol{\mathcal{I}}_k\right)$$

Proof. The result is established by invoking a version of the CLT for martingales [6, 7].

First, recall that the gradient with respect to the parameters of the m-th mark,

$$\begin{aligned}\frac{1}{1-\beta}\nabla_{\boldsymbol{\omega}^{(m)}}\ell_k^\beta(\boldsymbol{\omega}_k^0) &= \sum_{i=k-N+1}^{k}\beta^{k-i}\nabla_{\boldsymbol{\omega}^{(m)}}\ell_i(\boldsymbol{\omega}_k^0) \\ &= \sum_{i=k-N+1}^{k}\beta^{k-i}\boldsymbol{X}_i{}'\left[\boldsymbol{n}_i^{*(m)} - \boldsymbol{\lambda}_i^{*(m)}(\boldsymbol{\omega}_k^0)\Delta\right],\end{aligned} \tag{28}$$

and that the covariates $\boldsymbol{X}_t$ are the ensemble spiking history $\{n_j^{(c)}\}_{c=1,j=t-p}^{C,t-1}$. Let $\mathcal{F}_t$ be the σ-field generated by $\{n_t^{(c)}\}_{c=1,t=t-p}^{C,t-1}$. It is straightforward to verify that each element of $\nabla \ell_k^\beta(\boldsymbol{\omega}_k^0)$ is the sum of martingale differences with respect to the filtration $\{\mathcal{F}_t\}_{t=1}^k$.

Employing the Cramer-Wold device, it will suffice to characterize the limiting distribution of $Z_k^\beta = \boldsymbol{z}'\left(\frac{1}{1-\beta}\nabla \ell_k^\beta(\boldsymbol{\omega}_k^0)\right)$ for an arbitrary unit vector $\boldsymbol{z}$. Let

$$\boldsymbol{U}_t := \beta^{k-t}\nabla \ell_t(\boldsymbol{\omega}_k^0), \quad \boldsymbol{V}_t := \beta^{2(k-t)}\nabla^2 \ell_t(\boldsymbol{\omega}_k^0), \tag{29}$$

so that we may write

$$\frac{1}{1-\beta}\nabla \ell_k^\beta(\boldsymbol{\omega}_k^0) = \sum_{t=k-N+1}^{k}\boldsymbol{U}_t. \tag{30}$$

Defining $Y_t = Z_t^\beta - Z_{t-1}^\beta = \boldsymbol{z}'\boldsymbol{U}_t$, note that $\mathbb{E}\left[Y_t | \mathcal{F}_{t-1}\right] = \boldsymbol{z}'\mathbb{E}\left[\boldsymbol{U}_t | \mathcal{F}_{t-1}\right] = 0$; thus, Z_k^β is a martingale. Also, noting that

$$\mathbb{E}\left[Y_t^2|\mathcal{F}_{t-1}\right] = \boldsymbol{z}'\mathbb{E}\left[\boldsymbol{U}_t\boldsymbol{U}_t{}'|\mathcal{F}_{t-1}\right]\boldsymbol{z} = \boldsymbol{z}'\mathbb{E}\left[-\boldsymbol{V}_t|\mathcal{F}_{t-1}\right]\boldsymbol{z}, \tag{31}$$

we define

$$\begin{aligned}s_\beta^2 := \mathbb{E}\left[{Z_k^\beta}^2\right] &= \sum_t \mathbb{E}\left[{Y_t}^2\right] \\ &= \boldsymbol{z}'\mathbb{E}\left[-\sum_t \boldsymbol{V}_t\right]\boldsymbol{z} \\ &= \frac{1}{1-\beta^2}\boldsymbol{z}'\boldsymbol{\mathcal{I}}_k\boldsymbol{z}.\end{aligned} \tag{32}$$

Thus, we must establish that $Z_k^\beta/s_\beta \xrightarrow{d} \mathcal{N}(0,1)$ as $\beta \to 1$. The version of CLT for martingales in [7] stipulates this result if the following two conditions hold:

(i) $s_\beta^{-2}\sum_t \mathbb{E}\left[{Y_t}^2|\mathcal{F}_{t-1}\right] \xrightarrow{p} 1$, and

(ii) $s_\beta^{-2}\sum_t \mathbb{E}\left[{Y_t}^2\mathbb{1}\left\{|Y_t| \geq \varepsilon s_\beta\right\}\right] \to 0, \forall \varepsilon > 0$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function.

We first address condition (i). Substituting for s_β and ${Y_t}^2$, the condition is rewritten as

$$\frac{(1-\beta^2)\sum_t \boldsymbol{z}'\mathbb{E}\left[-\boldsymbol{V}_t|\mathcal{F}_{t-1}\right]\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{\mathcal{I}}_k\boldsymbol{z}} \to 1. \tag{33}$$

Letting $\rho = \beta^2$, and using the definition of $\boldsymbol{V}_t$, the condition may be equivalently expressed as

$$\frac{(1-\rho)\boldsymbol{z}'\sum_t \rho^{k-t}\left(\mathbb{E}\left[-\nabla^2\ell_t(\boldsymbol{\omega}_k^0)|\mathcal{F}_{t-1}\right] - \mathbb{E}\left[-\nabla^2\ell_k(\boldsymbol{\omega}_k^0)\right]\right)\boldsymbol{z}}{\boldsymbol{z}'\mathbb{E}\left[-\nabla^2\ell_k(\boldsymbol{\omega}_k^0)\right]\boldsymbol{z}} \xrightarrow{p} 0, \tag{34}$$

where we have used the Fisher information equality; this condition is implied by monotone convergence and the result of Proposition 1.

We next consider condition (ii), also known as the Lindeberg condition. As $\beta \to 1$, note that

$$s_\beta = \left(\frac{1}{1-\beta^2}\boldsymbol{z}'\boldsymbol{\mathcal{I}}_k\boldsymbol{z}\right)^{\frac{1}{2}} \to \infty,$$

while

$$|Y_t| = \left|\boldsymbol{z}'\beta^{k-t}\nabla\ell_t(\boldsymbol{\omega}_k^0)\right| \to \left|\boldsymbol{z}'\nabla\ell_t(\boldsymbol{\omega}_k^0)\right|.$$

Thus, $\mathbb{1}\{|Y_t| \geq \varepsilon s_\beta\} \to 0$ for all $\varepsilon > 0$, and the Lindeberg condition holds, thereby proving $Z_k^\beta/s_\beta \xrightarrow{d} \mathcal{N}(0,1)$. To relate the statement of the proposition to Z_k^β/s_β more explicitly, observe that

$$\begin{aligned}\frac{Z_k^\beta}{s_\beta} &= \frac{\boldsymbol{z}'\nabla\ell_k^\beta(\boldsymbol{\omega}_k^0)/(1-\beta)}{\left(\boldsymbol{z}'\boldsymbol{\mathcal{I}}_k\boldsymbol{z}/(1-\beta^2)\right)^{1/2}} \\ &= \sqrt{\frac{1+\beta}{1-\beta}}\frac{\boldsymbol{z}'\nabla\ell_k^\beta(\boldsymbol{\omega}_k^0)}{\sqrt{\boldsymbol{z}'\boldsymbol{\mathcal{I}}_k\boldsymbol{z}}}.\end{aligned} \tag{35}$$

□

C.2 Asymptotic Normality of the De-Sparsified AdOMP Estimates

Theorem 1 establishes the asymptotic normality of the de-sparsified greedy estimate based on the following set of technical conditions.

(C1) For $i = 1,\ldots,k$, define $\boldsymbol{z}_i := [\boldsymbol{z}_i^{(1)'},\ldots,\boldsymbol{z}_i^{(C^*)'}]'$ such that $\{\boldsymbol{z}_i^{(m)}\}_{m=1:C^*} := \{\boldsymbol{X}_i\boldsymbol{\omega}_k^{(m)}\}_{m=1:C^*}$. Writing $\ell_i(\boldsymbol{\omega}_k)$ equivalently as $\ell_i(\{\boldsymbol{z}_i^{(m)}\}_{m=1:C^*})$, the second derivatives $\nabla^2_{\boldsymbol{z}_i}\ell_i(\{\boldsymbol{z}_i^{(m)}\}_{m=1:C^*})$ exist and satisfy the following for all $\boldsymbol{z}_i$ and for some δ-neighborhood ($\delta > 0$):

$$\max_{\boldsymbol{z}_0\in\{\boldsymbol{z}_i\}_{i=1}^k}\left\{\sup_{|\hat{\boldsymbol{z}}-\boldsymbol{z}_0|\vee|\boldsymbol{z}-\boldsymbol{z}_0|<\delta}\frac{\|\nabla^2_{\boldsymbol{z}_i}\ell_i(\{\hat{\boldsymbol{z}}_i^{(m)}\}_{m=1}^{C^*}) - \nabla^2_{\boldsymbol{z}_i}\ell_i(\{\boldsymbol{z}_i^{(m)}\}_{m=1}^{C^*})\|}{\|\hat{\boldsymbol{z}}-\boldsymbol{z}\|}\right\} \leq 1$$

(C2) Assuming the true parameters, $\boldsymbol{\omega}_k \in \mathbb{R}^d$, to be (s,ξ)-compressible with $\xi < \frac{1}{2}$, as in [8], the error $\|\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0\|_1 = \mathcal{O}_{\mathbb{P}}(\zeta\sqrt{d})$, where $\zeta := \sqrt{(1-\beta)\log(s)\log(d)}$ and $\boldsymbol{\omega}_k^0$ is the maximum likelihood estimate.

(C3) Letting $\hat{\Sigma}_k := \nabla^2\ell_k^\beta(\hat{\boldsymbol{\omega}}_k)$, and $\Sigma_k := \mathbb{E}\left[\nabla^2\ell_k(\boldsymbol{\omega}_k)\right]$, one of the following two conditions holds. For every β_l in the monotonically increasing sequence $\{\beta_l\}_{l=1}^\infty$ that converges to 1, $\hat{\Theta}_k = \hat{\Sigma}_k^{-1}$ exists. Alternatively, $\hat{\Theta}_k$ is a consistent estimator of Σ_k^{-1}, and $\|\hat{\Theta}_k\hat{\Sigma}_k - \boldsymbol{I}\|_\infty = \mathcal{O}_{\mathbb{P}}(\sqrt{1-\beta})$.

(C4) The covariates satisfy $\|\boldsymbol{X}\|_\infty := \max\left|[\boldsymbol{X}]_{i,j}\right| = \mathcal{O}(c_1)$ for some constant c_1, where $\boldsymbol{X} := \left[{\boldsymbol{X}_1}',\ldots,{\boldsymbol{X}_k}'\right]'$ is the collection of history covariate matrices over all windows.

(C5) The covariates satisfy $\|\tilde{\boldsymbol{X}}\hat{\Theta}_k\|_\infty = \mathcal{O}_{\mathbb{P}}(c_1)$, where the $T \times d$ matrix $\tilde{\boldsymbol{X}}$ is obtained by tiling $\boldsymbol{X}$ horizontally C^* times.

(C6) The covariates satisfy $\|\tilde{\boldsymbol{X}}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\|^2 = \mathcal{O}_{\mathbb{P}}\left(c_2\zeta^2\right)$ for some constant c_2.

Theorem 1 is established by decomposing the difference $\left(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^0\right)$ into two components, one negligible and the non-negligible, using conditions (C1)-(C6). We also provide the rate of decay of negligible terms explicitly, but summarily treat them as $o_{\mathbb{P}}(1)$. Then, Propositions 1 and 2 are used show to the non-negligible component is asymptotically normal.

Proof. To begin, we analyze the de-sparsified estimates $\hat{\boldsymbol{w}}_k = \hat{\boldsymbol{\omega}}_k - \hat{\Theta}_k \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)$ element-wise, attending particularly to the latter term of the right hand side. For the j^{th} element,

$$\begin{aligned}(\hat{\Theta}_k)_j \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) =& (\hat{\Theta}_k)_j \nabla \ell_k^\beta(\boldsymbol{\omega}_k^0) \\ &+ (\hat{\Theta}_k)_j \nabla^2 \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)\left(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0\right) + \Delta_1^{(j)},\end{aligned} \tag{36}$$

where

$$\Delta_1^{(j)} = (\hat{\Theta}_k)_j \left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0) - \nabla^2 \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)\right)\left(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0\right). \tag{37}$$

Note that based on the expansion in Eq. (36), the difference between the desparsified and maximimum likelihood estimates (for the j^{th} component)

$$\begin{aligned}(\hat{\boldsymbol{w}}_k)_j - (\boldsymbol{\omega}_k^0)_j &= (\hat{\boldsymbol{\omega}}_k)_j - (\boldsymbol{\omega}_k^0)_j - (\hat{\Theta}_k)_j \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) \\ &= (\hat{\boldsymbol{\omega}}_k)_j - (\boldsymbol{\omega}_k^0)_j - (\hat{\Theta}_k)_j \nabla \ell_k^\beta(\boldsymbol{\omega}_k^0) \\ &\quad - (\hat{\Theta}_k)_j \nabla^2 \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0) - \Delta_1^{(j)} \\ &= -(\hat{\Theta}_k)_j \nabla \ell_k^\beta(\boldsymbol{\omega}_k^0) - \Delta_2^{(j)},\end{aligned} \tag{38}$$

where

$$\Delta_2^{(j)} = \Delta_1^{(j)} + \left((\hat{\Theta}_k)_j \nabla^2 \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) - \boldsymbol{e}_j'\right)(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0), \tag{39}$$

with $\boldsymbol{e}_j$ a standard basis vector. In the following, we derive uniform upper bounds on both $|\Delta_1^{(j)}|$ and $|\Delta_2^{(j)}|$ for all $j = 1, \ldots, d$.

We reintroduce two notations from conditions (C1) and (C5) to proceed. Recall that the collection of covariate matrices over all windows is denoted as $\boldsymbol{X} = \left[\boldsymbol{X}_1', \ldots, \boldsymbol{X}_k'\right]'$; letting $\tilde{\boldsymbol{X}}_i$ denote the matrix $\boldsymbol{X}_i$ tiled horizontally C^* times, we also define the $T \times d$ matrix $\tilde{\boldsymbol{X}} = \left[\tilde{\boldsymbol{X}}_1', \ldots, \tilde{\boldsymbol{X}}_k'\right]'$. Thus, for $i = 1, \ldots, k$, we defined $\boldsymbol{z}_i := \tilde{\boldsymbol{X}}_i \boldsymbol{\omega}_k$, and the log-likelihood function for the i^{th} window can be written as $\ell_i(\boldsymbol{\omega}_k) = \ell_i(\boldsymbol{z}_i)$.

For the i^{th} window, note that

$$\nabla_{\boldsymbol{z}} \ell_i(\hat{\boldsymbol{z}}_i) = \nabla_{\boldsymbol{z}} \ell_i(\boldsymbol{z}_i^0) + \nabla_{\boldsymbol{z}}^2 \ell_i(\tilde{\boldsymbol{z}}_i)\left(\hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^0\right), \tag{40}$$

for some $\tilde{\boldsymbol{z}}_i$ such that $\|\tilde{\boldsymbol{z}}_i - \hat{\boldsymbol{z}}_i\|_2 \leq \left\|\hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^0\right\|_2$. Using condition (B1), we have that

$$\begin{aligned}\left\|\left(\nabla_{\boldsymbol{z}}^2 \ell_i(\tilde{\boldsymbol{z}}_i) - \nabla_{\boldsymbol{z}}^2 \ell_i(\hat{\boldsymbol{z}}_i)\right)\left(\hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^0\right)\right\|_2 &\leq \|\tilde{\boldsymbol{z}}_i - \hat{\boldsymbol{z}}_i\|_2 \left\|\hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^0\right\|_2 \\ &\leq \left\|\hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^0\right\|_2^2.\end{aligned} \tag{41}$$

Additionally, note that, by the chain rule for derivatives,

$$\nabla_{\boldsymbol{\omega}}^2 \ell_i(\boldsymbol{\omega}_k) = \tilde{\boldsymbol{X}}_i' \nabla_{\boldsymbol{z}}^2 \ell_i(\boldsymbol{z}_i) \tilde{\boldsymbol{X}}_i. \tag{42}$$

Using these two relations, a uniform upper bound for $\{\Delta_1^{(j)}\}_{j=1}^d$ is derived as follows. The residual is expanded as

$$\Delta_1^{(j)} = (1-\beta)\sum_{i=1}^k \beta^{k-i} (\hat{\Theta}_k)_j \tilde{\boldsymbol{X}}_i' \left(\nabla_{\boldsymbol{z}}^2 \ell_i(\boldsymbol{z}_i^0) - \nabla_{\boldsymbol{z}}^2 \ell_i(\hat{\boldsymbol{z}}_i)\right) \tilde{\boldsymbol{X}}_i (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0). \tag{43}$$

First the i^{th} term in the summation, neglecting β^{k-i}, is addressed:

$$\begin{aligned}
&\left|(\hat{\Theta}_k)_j \tilde{\boldsymbol{X}}_i' \left(\nabla^2_{\boldsymbol{z}} \ell_i(\boldsymbol{z}_i^0) - \nabla^2_{\boldsymbol{z}} \ell_i(\hat{\boldsymbol{z}}_i)\right) \tilde{\boldsymbol{X}}_i (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right| \\
&\leq \left\|(\hat{\Theta}_k)_j \tilde{\boldsymbol{X}}_i'\right\|_\infty \left\|\left(\nabla^2_{\boldsymbol{z}} \ell_i(\boldsymbol{z}_i^0) - \nabla^2_{\boldsymbol{z}} \ell_i(\hat{\boldsymbol{z}}_i)\right) \tilde{\boldsymbol{X}}_i (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right\|_1 \\
&\leq \left\|(\hat{\Theta}_k)_j \tilde{\boldsymbol{X}}_i'\right\|_\infty \sqrt{d} \left\|\left(\nabla^2_{\boldsymbol{z}} \ell_i(\boldsymbol{z}_i^0) - \nabla^2_{\boldsymbol{z}} \ell_i(\hat{\boldsymbol{z}}_i)\right) \tilde{\boldsymbol{X}}_i (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right\|_2 \\
&\leq \left\|\hat{\Theta}_k \tilde{\boldsymbol{X}}'\right\|_\infty \sqrt{d} \left\|\tilde{\boldsymbol{X}}_i (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right\|_2^2,
\end{aligned}$$

where the first inequality is a consequence of Hölder's inequality, the second is due to the equivalence of norms, and the third follows by invoking condition (C1), as in Eq. (41). Using the fact that $\beta < 1$, it follows that

$$\begin{aligned}
\left|\Delta_1^{(j)}\right| &\leq (1-\beta) \sum_{i=1}^{k} \beta^{k-i} \left\|\hat{\Theta}_k \tilde{\boldsymbol{X}}'\right\|_\infty \sqrt{d} \left\|\tilde{\boldsymbol{X}}_i (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right\|_2^2 \\
&\leq \left\|\hat{\Theta}_k \tilde{\boldsymbol{X}}'\right\|_\infty \sqrt{d} \left\|\tilde{\boldsymbol{X}} (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right\|_2^2.
\end{aligned} \tag{44}$$

Consequently, the term $|\Delta_2^{(j)}|$ is upper-bounded as follows:

$$\begin{aligned}
\left|\Delta_2^{(j)}\right| \leq &\left\|\hat{\Theta}_k \tilde{\boldsymbol{X}}'\right\|_\infty \sqrt{d} \left\|\tilde{\boldsymbol{X}} (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right\|_2^2 \\
&+ \left\|(\hat{\Theta}_k)_j \nabla^2 \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) - \boldsymbol{e}_j'\right\|_\infty \left\|\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0\right\|_1.
\end{aligned} \tag{45}$$

Invoking conditions (C5) and (C6),

$$\begin{aligned}
\left\|\hat{\Theta}_k \tilde{\boldsymbol{X}}'\right\|_\infty \sqrt{d} \left\|\tilde{\boldsymbol{X}} (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0)\right\|_2^2 &= \mathcal{O}_{\mathbb{P}}\left(c_1 \sqrt{d}(1-\beta)\log(s)\log(d)\right) \\
&= o_{\mathbb{P}}(1),
\end{aligned} \tag{46}$$

and hence, $\left|\Delta_1^{(j)}\right|$ is negligible for each $j = 1, \ldots, d$. Similarly, condition (C2) implies that $\left\|\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^0\right\|_1 = \mathcal{O}_{\mathbb{P}}\left(\sqrt{1-\beta}\sqrt{\log(s)\log(d)}\sqrt{d}\right) = o_{\mathbb{P}}(1)$; invoking the consistency of the estimator $\hat{\Theta}_k$, the second part of condition (C3), it follows that $\left|\Delta_2^{(j)}\right| = o_{\mathbb{P}}(1)$. Of course, this conclusion follows immediately from Eq. (45) if $\nabla^2 \ell_k^\beta$ is invertible, as per the first part of condition (C3). Thus, we have shown that $\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^0 = -(\hat{\Theta}_k)\nabla \ell_k^\beta(\boldsymbol{\omega}_k^0) + o_{\mathbb{P}}(1) \cdot \mathbf{1}$.

Proposition 1 establishes that $\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^0) \to \mathbb{E}[\nabla^2 \ell_i(\boldsymbol{\omega}_k)]$, or equivalently that $\hat{\Sigma}_k \to \Sigma_k$; invoking condition (C3), it follows that $\hat{\Theta}_k \to \Sigma_k^{-1}$. Recall that this is a sequential limit in β, if $\hat{\Sigma}_k$ is invertible for each β, $\hat{\Theta}_k \to \Sigma_k^{-1}$ by the continuous mapping theorem. Alternatively, if $\hat{\Theta}_k$ is not invertible, the statement follows directly by second part of condition (C3). Proposition 2 establishes the asymptotic normality of $\nabla \ell_k^\beta(\boldsymbol{\omega}_k^0)$; in conjunction with the aforementioned limiting behavior of $\hat{\Theta}_k$ and Slutsky's theorem, we thus conclude that $\sqrt{\frac{1+\beta}{1-\beta}}\left(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^0\right)$ is asymptotically normal with zero-mean and covariance matrix $\Sigma_k^{-1}\boldsymbol{\mathcal{I}}_k\Sigma_k^{-1} = \boldsymbol{\mathcal{I}}_k^{-1}$. $\square$

C.3 Limiting Distributions of the Adaptive De-biased Deviance Difference

This proof of Theorem 2 characterizes the limiting behavior of the de-biased deviance difference statistic for testing r^{th}-order spiking coordination. However, the arguments also generalize to any nested model, being closely related to the treatment in [9] for the de-biased deviance difference in testing Granger causal links.

We first characterize the limiting behavior of the de-biased deviance difference under the null hypothesis; then, the treatment of Davidson and Lever [10] for a sequence of local alternatives is adapted to establish the result under the alternative hypothesis. In each case, we first characterize the limiting behavior of the de-biased deviance (between the estimated and true parameters) and subsequently establish the result for the deviance difference.

Proof. We begin with the adaptive de-biased deviance statistic for the estimated parameters $\hat{\boldsymbol{\omega}}_k$ and the true parameters $\boldsymbol{\omega}_k$:

$$D_{k,\beta}(\hat{\boldsymbol{\omega}}_k, \boldsymbol{\omega}_k) = \left(\frac{1+\beta}{1-\beta}\right) \left[2\left(\ell_k^\beta(\hat{\boldsymbol{\omega}}_k) - \ell_k^\beta(\boldsymbol{\omega}_k)\right) - \mathscr{B}_k\right], \tag{47}$$

where $\mathscr{B}_k := \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)' \left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k)\right)^{-1} \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)$. To write the deviance in a more convenient form, we first note the quadratic expansion of $\ell_k^\beta(\boldsymbol{\omega}_k)$:

$$\begin{aligned}\ell_k^\beta(\boldsymbol{\omega}_k) =& \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) + (\boldsymbol{\omega}_k - \hat{\boldsymbol{\omega}}_k)' \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) \\ &+ \frac{1}{2}(\boldsymbol{\omega}_k - \hat{\boldsymbol{\omega}}_k)' \left(\nabla^2 \ell_k^\beta(\tilde{\boldsymbol{\omega}}_k)\right)(\boldsymbol{\omega}_k - \hat{\boldsymbol{\omega}}_k).\end{aligned} \tag{48}$$

Substituting into Eq. (47) and expressing the bias correction term $\mathscr{B}_k$ explicitly, the deviance is equivalent to

$$\begin{aligned}\left(\frac{1-\beta}{1+\beta}\right) D_{k,\beta}(\hat{\boldsymbol{\omega}}_k, \boldsymbol{\omega}_k) =& 2(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k)' \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) \\ &- (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k)' \left(\nabla^2 \ell_k^\beta(\tilde{\boldsymbol{\omega}}_k)\right)(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k) \\ &- \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)' \left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k)\right)^{-1} \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k),\end{aligned} \tag{49}$$

where $\tilde{\boldsymbol{\omega}}_k$ intermediates $\hat{\boldsymbol{\omega}}_k$ and $\boldsymbol{\omega}_k$. By rearranging terms and recalling that the de-sparsified estimate $\hat{\boldsymbol{w}}_k = \hat{\boldsymbol{\omega}}_k - \left(\nabla^2 \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)\right)^{-1} \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)$, the deviance can be compactly expressed as:

$$\left(\frac{1-\beta}{1+\beta}\right) D_{k,\beta}(\hat{\boldsymbol{\omega}}_k, \boldsymbol{\omega}_k) = -(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k)' \nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k)(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k) + \Delta_1, \tag{50}$$

with $\Delta_1 = (\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k)' \left(\nabla^2 \ell_k^\beta(\tilde{\boldsymbol{\omega}}_k) - \nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k)\right)(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k)$. It can be shown based on the conditions of Theorem 1 that $|\Delta_1| = \mathcal{O}_{\mathbb{P}}\left(\|\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k\|^3\right) = o_{\mathbb{P}}\left((1-\beta)^{3/2}\right)$, and thus negligible. Once again, negligible terms are summarized as $o_{\mathbb{P}}(1)$ terms. Similar arguments were presented in the course of proving Theorem 1.

We now explicitly define the null and sequence of local alternative hypotheses in order to adapt Davidson and Lever's treatment [10]. Recall first that limits in β are understood to be sequential limits; i.e., they are evaluated for the sequence $\{\beta_l\}_{l=1}^\infty$ where $\beta_l \to 1$ as $l \to \infty$. Then, for window k, we test the null hypothesis $H_{0,k} : \boldsymbol{\omega}_k^0 = (\boldsymbol{\omega}_{0,k}, \boldsymbol{\omega}_{1,k})$ against the sequence of local alternative hypotheses $\left\{H_{1,k}^{\beta_l} : \boldsymbol{\omega}_k^{\beta_l} = \left(\boldsymbol{\omega}_{0,k}^*, \boldsymbol{\omega}_{1,k}^{\beta_l}\right)\right\}_{l=1}^\infty$, where $\boldsymbol{\omega}_{1,k}^{\beta_l} = \boldsymbol{\omega}_{1,k} + \sqrt{\frac{1-\beta_l}{1+\beta_l}}\boldsymbol{\delta}_k$; the limiting true parameter vector under the alternative is denoted by $\boldsymbol{\omega}_k^*$. The partition of the $d-$dimensional parameter vector corresponds to the free parameters under the null hypothesis ($\boldsymbol{\omega}_{0,k}$, for example) and the restricted sub-vector ($\boldsymbol{\omega}_{1,k}$). Thus, the statistical test seeks to detect local perturbations of order $\mathcal{O}\left(\sqrt{\frac{1-\beta}{1+\beta}}\right)$ from the null hypothesis. In the statistical inference of r^{th}-order coordination, $\boldsymbol{\omega}_{1,k}$ corresponds to the base rate parameters of r^{th}-order events, the number of which is denoted by $M^{(r)}$; however, since a similar partition can be made for any nested hypothesis test, the result shown here generalizes.

We now establish the limiting behavior of the de-biased deviance difference under the null hypothesis. By Proposition 1, $\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k) \to -\mathcal{I}_k$. Additionally, by Proposition 2, $\sqrt{\frac{1+\beta}{1-\beta}} \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) \to \mathcal{N}(\mathbf{0}, \mathcal{I}_k)$. Thus, recalling the definition of the the de-sparsified estimate $\hat{\boldsymbol{w}}_k$ and invoking Slutsky's Theorem, $\sqrt{\frac{1+\beta}{1-\beta}}(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k) \to \mathcal{N}(\mathbf{0}, \mathcal{I}_k^{-1})$ and consequently, the deviance $[D_{k,\beta}(\hat{\boldsymbol{\omega}}_k, \boldsymbol{\omega}_k)|H_{0,k}] \to \chi^2(d)$, asymptotically following a χ^2 distribution with d degrees of freedom. Finally, invoking classical results for likelihood ratio tests of nested models [11, 12], we conclude that under the null hypothesis, the adaptive de-biased deviance difference converges in distribution to a χ^2 random variable with $M^{(r)}$ degrees of freedom:

$$\left[D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\omega}}_k^{(F)}, \hat{\boldsymbol{\omega}}_k^{(R)}\right) \middle| H_{0,k}\right] \to \chi^2(M^{(r)}) \tag{51}$$

It remains to establish the limiting behavior of the adaptive de-biased deviance difference under the sequence of alternative hypotheses. Though Proposition 1 establishes that $\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*) \to -\boldsymbol{\mathcal{I}}_k^*$, the asymptotic normality of $\sqrt{\frac{1+\beta}{1-\beta}}(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^*)$ does not follow as readily. To establish asymptotic normality, we first derive an alternative expression for the difference $(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^*)$ using the following two Taylor expansions. The gradient of the log-likelihood evaluated at $\hat{\boldsymbol{\omega}}_k$ and $\boldsymbol{\omega}_k^\beta$ may equivalently be written as:

$$
\begin{aligned}
\nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) &= \nabla \ell_k^\beta(\boldsymbol{\omega}_k^*) + \nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\left(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^*\right) + \Delta_2 \\
\nabla \ell_k^\beta(\boldsymbol{\omega}_k^\beta) &= \nabla \ell_k^\beta(\boldsymbol{\omega}_k^*) + \nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\left(\boldsymbol{\omega}_k^\beta - \boldsymbol{\omega}_k^*\right) + \Delta_3,
\end{aligned} \tag{52}
$$

respectively. The remainder terms

$$
\Delta_2 := \left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k') - \nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\right)\left(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^*\right)
$$

and

$$
\Delta_3 := \left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k'') - \nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\right)\left(\boldsymbol{\omega}_k^\beta - \boldsymbol{\omega}_k^*\right),
$$

are negligible at rate $\|\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^*\|^2 = o_{\mathbb{P}}(1-\beta)$, which can also be shown based on the conditions of Theorem 1. Thus, we can rewrite

$$
\begin{aligned}
\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^* &= \hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}_k^* - \left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\right)^{-1} \nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k) \\
&= -\left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\right)^{-1} \nabla \ell_k^\beta(\boldsymbol{\omega}_k^*) + o_{\mathbb{P}}(1) \\
&= \boldsymbol{\omega}_k^\beta - \boldsymbol{\omega}_k^* - \left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\right)^{-1} \nabla \ell_k^\beta(\boldsymbol{\omega}_k^\beta) + o_{\mathbb{P}}(1).
\end{aligned} \tag{53}
$$

Denoting $\boldsymbol{\omega}_k^\beta - \boldsymbol{\omega}_k^*$ as $\tilde{\boldsymbol{\delta}}_k := \left[\mathbf{0}', \sqrt{\frac{1-\beta}{1+\beta}}\boldsymbol{\delta}_k'\right]'$, invoking Propositions 1 and 2, by Slutsky's theorem we find that $\sqrt{\frac{1+\beta}{1-\beta}}\left(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^*\right) \to \mathcal{N}\left(\tilde{\boldsymbol{\delta}}_k, (\boldsymbol{\mathcal{I}}_k^*)^{-1}\right)$. In contrast to the null case, the limiting normal distribution has non-zero mean; thus, we employ Davidson and Lever's approach [10] to establish the limiting behavior of the adaptive de-biased deviance difference.

To proceed, we first decompose the Fisher information matrix in accordance with the partition of the parameter vector introduced earlier. The parameters $\boldsymbol{\omega}_k^*$ were partitioned into $\boldsymbol{\omega}_{0,k}^*$, corresponding to the parameters free under the null hypothesis, and $\boldsymbol{\omega}_{1,k}^*$, the parameters restricted under the null hypothesis; $\boldsymbol{\mathcal{I}}_k^*$ is similarly decomposed as:

$$
\boldsymbol{\mathcal{I}}_k^* = \begin{bmatrix} \boldsymbol{\mathcal{I}}_{0,0,k}^* & \boldsymbol{\mathcal{I}}_{0,1,k}^* \\ \boldsymbol{\mathcal{I}}_{1,0,k}^* & \boldsymbol{\mathcal{I}}_{1,1,k}^* \end{bmatrix},
$$

where $\boldsymbol{\mathcal{I}}_{0,1,k}^* = \boldsymbol{\mathcal{I}}_{1,0,k}^*{}'$. Utilizing the quadratic form of the de-biased deviance in Eq. (50) and invoking Proposition 1, the adaptive de-biased deviance difference is expressed as:

$$
\begin{aligned}
D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\omega}}_k^{(F)}, \hat{\boldsymbol{\omega}}_k^{(R)}\right) = &\left(\frac{1+\beta}{1-\beta}\right)\left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right)' \boldsymbol{\mathcal{I}}_k^* \left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right) \\
&- \left(\frac{1+\beta}{1-\beta}\right)\left(\hat{\boldsymbol{w}}_{0,k}^{(R)} - \boldsymbol{\omega}_{0,k}^*\right)' \boldsymbol{\mathcal{I}}_{0,0,k}^* \left(\hat{\boldsymbol{w}}_{0,k}^{(R)} - \boldsymbol{\omega}_{0,k}^*\right) \\
&+ o_{\mathbb{P}}(1).
\end{aligned} \tag{54}
$$

In the following steps, an equivalent expression for the reduced model deviance is derived in terms of $\left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right)$. Recalling the Taylor expansion of $\nabla \ell_k^\beta(\hat{\boldsymbol{\omega}}_k)$ around $\boldsymbol{\omega}_k^*$ in Eq. (52), we see that

$$
\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^* = -\left(\nabla^2 \ell_k^\beta(\boldsymbol{\omega}_k^*)\right)^{-1} \nabla \ell_k^\beta(\boldsymbol{\omega}_k^*) + o_{\mathbb{P}}(1-\beta),
$$

and that

$$
\nabla \ell_k^\beta(\boldsymbol{\omega}_k^*) = \boldsymbol{\mathcal{I}}_k^*\left(\hat{\boldsymbol{w}}_k - \boldsymbol{\omega}_k^*\right) + o_{\mathbb{P}}(1-\beta),
$$

having invoked Proposition 1. It follows that the partial gradient $\nabla_{\boldsymbol{\omega}_{0,k}} \ell_k^{\beta}(\boldsymbol{\omega}_k^*) = \boldsymbol{\mathcal{I}}_{0,0,k}^* \left(\hat{\boldsymbol{w}}_{0,k} - \boldsymbol{\omega}_{0,k}^*\right) + o_{\mathbb{P}}(1-\beta)$, and so the second term of Eq. (54) is equivalent to

$$\left(\frac{1+\beta}{1-\beta}\right) \left(\nabla_{\boldsymbol{\omega}_{0,k}} \ell_k^{\beta}(\boldsymbol{\omega}_k^*)\right)' \left(\boldsymbol{\mathcal{I}}_{0,0,k}^*\right)^{-1} \left(\nabla_{\boldsymbol{\omega}_{0,k}} \ell_k^{\beta}(\boldsymbol{\omega}_k^*)\right) + o_{\mathbb{P}}(1). \tag{55}$$

Note that the partial gradient is a subvector of the gradient, i.e., $\nabla_{\boldsymbol{\omega}_{0,k}} \ell_k^{\beta}(\boldsymbol{\omega}_k^*) = \left(\nabla \ell_k^{\beta}(\boldsymbol{\omega}_k^*)\right)_0$. Hence,

$$\begin{aligned} \nabla_{\boldsymbol{\omega}_{0,k}} \ell_k^{\beta}(\boldsymbol{\omega}_k^*) &= \left(\boldsymbol{\mathcal{I}}_k^* \left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right)\right)_0 + o_{\mathbb{P}}(1-\beta) \\ &= \boldsymbol{\mathcal{I}}_{0,\cdot,k}^* \left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right) + o_{\mathbb{P}}(1-\beta), \end{aligned} \tag{56}$$

where $\boldsymbol{\mathcal{I}}_{0,\cdot,k}^* = \left[\boldsymbol{\mathcal{I}}_{0,0,k}^*, \boldsymbol{\mathcal{I}}_{0,1,k}^*\right]$. Equation (55) can then be expressed as

$$\left(\frac{1+\beta}{1-\beta}\right) \left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right)' \boldsymbol{A} \left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right) + o_{\mathbb{P}}(1), \tag{57}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\mathcal{I}}_{0,0,k}^* & \boldsymbol{\mathcal{I}}_{0,1,k}^* \\ \boldsymbol{\mathcal{I}}_{1,0,k}^* & \boldsymbol{\mathcal{I}}_{1,0,k}^* \left(\boldsymbol{\mathcal{I}}_{0,0,k}^*\right)^{-1} \boldsymbol{\mathcal{I}}_{0,1,k}^* \end{bmatrix}.$$

Thus, the adaptive de-biased deviance difference is equal to

$$D_{k,\beta}^{(r)} = \left(\frac{1+\beta}{1-\beta}\right) \left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right)' \left(\boldsymbol{\mathcal{I}}_k^* - \boldsymbol{A}\right) \left(\hat{\boldsymbol{w}}_k^{(F)} - \boldsymbol{\omega}_k^*\right) + o_{\mathbb{P}}(1). \tag{58}$$

Note that

$$\boldsymbol{\mathcal{I}}_k^* - \boldsymbol{A} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{I}}_{1,1,k}^* - \boldsymbol{\mathcal{I}}_{1,0,k}^* \left(\boldsymbol{\mathcal{I}}_{0,0,k}^*\right)^{-1} \boldsymbol{\mathcal{I}}_{0,1,k}^* \end{bmatrix}, \tag{59}$$

and so, defining $\tilde{\boldsymbol{\mathcal{I}}}_k^* := \boldsymbol{\mathcal{I}}_{1,1,k}^* - \boldsymbol{\mathcal{I}}_{1,0,k}^* \left(\boldsymbol{\mathcal{I}}_{0,0,k}^*\right)^{-1} \boldsymbol{\mathcal{I}}_{0,1,k}^* = \left(\boldsymbol{\mathcal{I}}_k^{*-1}\right)_{1,1}^{-1}$, the deviance difference simplifies to

$$D_{k,\beta}^{(r)} = \left(\frac{1+\beta}{1-\beta}\right) \left(\hat{\boldsymbol{w}}_{1,k}^{(F)} - \boldsymbol{\omega}_{1,k}^*\right)' \tilde{\boldsymbol{\mathcal{I}}}_k^* \left(\hat{\boldsymbol{w}}_{1,k}^{(F)} - \boldsymbol{\omega}_{1,k}^*\right) + o_{\mathbb{P}}(1). \tag{60}$$

That $\sqrt{\frac{1+\beta}{1-\beta}} \left(\hat{\boldsymbol{w}}_{1,k}^{(F)} - \boldsymbol{\omega}_{1,k}^*\right) \to \mathcal{N}\left(\boldsymbol{\delta}_k, \left(\boldsymbol{\mathcal{I}}_k^{*-1}\right)_{1,1}\right)$ follows from earlier arguments. Using this fact, we conclude that, under the sequence of local alternatives $H_{1,k}^{\beta}$, the adaptive de-biased deviance difference converges to a non-central χ^2 random variable with $M^{(r)}$ degrees of freedom and non-centrality parameter $\nu_k^{(r)} := \boldsymbol{\delta}_k' \left(\boldsymbol{\mathcal{I}}_k^{*-1}\right)_{1,1} \boldsymbol{\delta}_k$:

$$\left[D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\omega}}_k^{(F)}, \hat{\boldsymbol{\omega}}_k^{(R)}\right) \middle| H_{1,k}^{\beta}\right] \to \chi^2(M^{(r)}, \nu_k^{(r)}). \tag{61}$$

□

C.4 Limiting Distributions of the Adaptive Deviance Difference: A Special Case for History-Independent Models

As has been previously proposed [13], the statistical inference procedure summarized in Algorithm 3 can be adapted to history-independent analysis. The two hypotheses,

$$\begin{aligned} H_0 &: r^{\text{th}}\text{-order simultaneous spikes occur as frequently as they would between independent units} \\ H_1 &: r^{\text{th}}\text{-order simultaneous spikes occur at a significantly different rate than they would between independent units} \end{aligned} \tag{62}$$

are quantified as in Section 4.1, excepting the history covariates. The distinction between this test and that specified in (10) in the main text is that the former seeks only to determine if observed rates of simultaneous spiking events are facilitated or suppressed, while the latter seeks to determine if observed rates of simultaneous spiking events are attributable to unobserved or neglected processes.

Estimates of the history-independent model parameters are unbiased, as no sparsity constraints are imposed to solve this special case of the maximum likelihood problem, (7). Hence, the adaptive deviance difference, $D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\mu}}_k^{(F)}, \hat{\boldsymbol{\mu}}_k^{(R)}\right) := \left(\frac{1+\beta}{1-\beta}\right)\left[2\left(\ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k^{(F)}) - \ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k^{(R)})\right)\right]$, is used as the test statistic. It is shown here, as a corollary to Theorem 2, that the asymptotic distributions of the adaptive deviance difference under the null and alternative hypotheses are characterized similarly.

Corollary 0.1. *Let $\hat{\boldsymbol{\mu}}_k^{(F)}$ and $\hat{\boldsymbol{\mu}}_k^{(R)}$ respectively be the full and reduced maximum-likelihood estimates of the history-independent model parameters at window k, where $\hat{\boldsymbol{\mu}}_k^{(R)}$ assumes conditionally independent r^{th}-order simultaneous spiking. Then, as $\beta \to 1$,*

i) if r^{th}-order synchrony matches independent r^{th}-order interactions, the adaptive deviance difference $D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\mu}}_k^{(F)}, \hat{\boldsymbol{\mu}}_k^{(R)}\right) \xrightarrow{d} \chi^2(M^{(r)})$, and

ii) if r^{th}-order synchrony diverges from independent r^{th}-order interactions, and assuming the base rate parameters of r^{th}-order interactions scale at least as $\mathcal{O}(\sqrt{\frac{1-\beta}{1+\beta}})$, the adaptive deviance difference $D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\mu}}_k^{(F)}, \hat{\boldsymbol{\mu}}_k^{(R)}\right) \xrightarrow{d} \chi^2(M^{(r)}, \nu_k^{(r)})$,

where $\nu_k^{(r)}$ is the non-centrality parameter at time k that depends only on the true parameters, and $M^{(r)} = |\mathcal{K}_r|$ is the difference in the cardinalities of the full and reduced support sets.

As in Theorem 2, we address the case when the window over which parameters are constant are of length $W = 1$. In the following, we prove the corollary result, with emphasis on the points of departure from Theorem 2.

Proof. Maximum-likelihood estimation of the parameters eliminates the need to de-bias the deviance, since the gradient evaluated at the maximum-likelihood estimate (and consequentially, the bias terms) is exactly zero. Hence, the test statistic reduces to the adaptive deviance difference:

$$D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\mu}}_k^{(F)}, \hat{\boldsymbol{\mu}}_k^{(R)}\right) = \left(\frac{1+\beta}{1-\beta}\right)\left[2\left(\ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k^{(F)}) - \ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k^{(R)})\right)\right] \tag{63}$$

As in the course of proving Theorem 2, we begin with the deviance between the estimated and true parameters, $D_{k,\beta}\left(\hat{\boldsymbol{\mu}}_k, \boldsymbol{\mu}_k\right) = \left(\frac{1+\beta}{1-\beta}\right)\left[2\left(\ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k^{(F)}) - \ell_k^{\beta}(\boldsymbol{\mu}_k)\right)\right]$. Noting that

$$\ell_k^{\beta}(\boldsymbol{\mu}_k) = \ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k) + \frac{1}{2}(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k)'\nabla^2\ell_k^{\beta}(\tilde{\boldsymbol{\mu}}_k)(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k), \tag{64}$$

the deviance can be expressed as

$$D_{k,\beta}\left(\hat{\boldsymbol{\mu}}_k, \boldsymbol{\mu}_k\right) = -\left(\frac{1+\beta}{1-\beta}\right)(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k)'\nabla^2\ell_k^{\beta}(\boldsymbol{\mu}_k)(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k) + o_{\mathbb{P}}(1). \tag{65}$$

To proceed, we first argue that asymptotic results analogous to Propositions 1 and 2 hold. To determine the limiting behavior of the Hessian, note that it can be written as $\nabla^2\ell_k^{\beta}(\boldsymbol{\mu}_k) = (1-\beta)\sum_{t=1}^{k}\beta^{k-t}\nabla^2\ell_t(\boldsymbol{\mu}_k)$; evaluating both sides, it can be shown that $\nabla^2\ell_k^{\beta}(\boldsymbol{\mu}_k) = \nabla^2\ell_k(\boldsymbol{\mu}_k)$ and is not a function of the simultaneous spiking observations. Hence, $\nabla^2\ell_k^{\beta}(\boldsymbol{\mu}_k) = \mathbb{E}\left[\nabla^2\ell_k^{\beta}(\boldsymbol{\mu}_k)\right] = -\boldsymbol{\mathcal{I}}_k$. It also follows that $\nabla^2\ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k) = -\boldsymbol{\mathcal{I}}_k$ by the asymptotic consistency of $\hat{\boldsymbol{\mu}}_k$.

The normality of the gradient, i.e., that $\sqrt{\frac{1+\beta}{1-\beta}}\nabla\ell_k^{\beta}(\hat{\boldsymbol{\mu}}_k) \to \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{I}}_k)$, is a consequence of Proposition 2. Note that, trivially, $\mathbb{E}\left[\nabla\ell_t(\hat{\boldsymbol{\mu}}_k)|\mathcal{F}_{t-1}\right] = 0$, where the filtration $\{\mathcal{F}_t\}_{t=0}^{k}$, as previously defined, corresponds to past spiking observations. Hence, the martingale CLT of Proposition 2 is applicable and the required conditions can be shown to hold. Alternatively, because $\nabla \ell_k^\beta(\hat{\boldsymbol{\mu}}_k)$ is the sum of independent but not identically distributed random variables, the Lindeberg-Feller CLT may be invoked to the same effect.

The limiting distribution of the adaptive deviance difference can now be established. First, we address the null hypothesis $H_{0,k} : \boldsymbol{\mu}_k^0 = (\boldsymbol{\mu}_{0,k}, \boldsymbol{\mu}_{1,k})$, where the parameters are partitioned into the free ($\boldsymbol{\mu}_{0,k}$) and restricted ($\boldsymbol{\mu}_{1,k}$) subsets as before. Through a series of arguments similar to those presented in proving Theorem 2, the limiting behaviors of $\nabla^2 \ell_k^\beta(\boldsymbol{\mu}_k)$ and $\sqrt{\frac{1+\beta}{1-\beta}} \nabla \ell_k^\beta(\hat{\boldsymbol{\mu}}_k)$ imply that $D_{k,\beta}(\hat{\boldsymbol{\mu}}_k, \boldsymbol{\mu}_k) \to \chi^2(d)$, where d is the dimensionality of the parameter vector; it likewise follows that the adaptive deviance difference is asymptotically χ^2-distributed with $M^{(r)}$ degrees of freedom

$$\left[D_{k,\beta}^{(r)} \left(\hat{\boldsymbol{\mu}}_k^{(F)}, \hat{\boldsymbol{\mu}}_k^{(R)} \right) \middle| H_{0,k} \right] \to \chi^2(M^{(r)}), \tag{66}$$

where, as before, $M^{(r)}$ is the dimensionality of the subvector $\boldsymbol{\mu}_{1,k}$.

Next, we address the limiting distribution of the adaptive deviance difference under the sequence (in β) of local alternatives $\left\{ H_{1,k}^\beta : \boldsymbol{\mu}_k^\beta = \left(\boldsymbol{\mu}_{0,k}^*, \boldsymbol{\mu}_{1,k} + \sqrt{\frac{1-\beta}{1+\beta}} \boldsymbol{\delta} \right) \right\}$, where β converges to unity and $\boldsymbol{\mu}_k^*$ is the limiting true parameter vector. The deviances considered in this case are between the estimated and true limiting parameter, $D_{k,\beta}(\hat{\boldsymbol{\mu}}_k, \boldsymbol{\mu}_k^*)$.

The characterization of the limiting behavior of the Hessian is true under the sequence of alternatives, but the asymptotic normality of $\sqrt{\frac{1+\beta}{1-\beta}} \left(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k^* \right)$ must be established. To this end, we derive an equivalent expression based on the following Taylor expansions of the gradient at $\hat{\boldsymbol{\mu}}_k$ and $\boldsymbol{\mu}_k^\beta$ around $\boldsymbol{\mu}_k^*$:

$$\begin{aligned} \nabla \ell_k^\beta(\hat{\boldsymbol{\mu}}_k) &= \nabla \ell_k^\beta(\boldsymbol{\mu}_k^*) + \nabla^2 \ell_k^\beta(\boldsymbol{\mu}_k^*) \left(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k^* \right) + o_{\mathbb{P}}(1-\beta) \\ \nabla \ell_k^\beta(\boldsymbol{\mu}_k^\beta) &= \nabla \ell_k^\beta(\boldsymbol{\mu}_k^*) + \nabla^2 \ell_k^\beta(\boldsymbol{\mu}_k^*) \left(\boldsymbol{\mu}_k^\beta - \boldsymbol{\mu}_k^* \right) + o_{\mathbb{P}}(1-\beta). \end{aligned} \tag{67}$$

By rearranging the terms of these two equations, and by noting that $\nabla \ell_k^\beta(\hat{\boldsymbol{\mu}}_k) = 0$, the following equivalence is derived:

$$\begin{aligned} \hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k^* &= -\left(\nabla^2 \ell_k^\beta(\boldsymbol{\mu}_k^*) \right)^{-1} \nabla \ell_k^\beta(\boldsymbol{\mu}_k^*) + o_{\mathbb{P}}(1-\beta) \\ &= \boldsymbol{\mu}_k^\beta - \boldsymbol{\mu}_k^* - \left(\nabla^2 \ell_k^\beta(\boldsymbol{\mu}_k^*) \right)^{-1} \nabla \ell_k^\beta(\boldsymbol{\mu}_k^\beta) + o_{\mathbb{P}}(1-\beta). \end{aligned} \tag{68}$$

Denoting $\boldsymbol{\mu}_k^\beta - \boldsymbol{\mu}_k^*$ as $\tilde{\boldsymbol{\delta}}_k := \left[\mathbf{0}', \sqrt{\frac{1-\beta}{1+\beta}} \boldsymbol{\delta}_k' \right]'$, it follows from the limiting behavior of $\nabla^2 \ell_k^\beta(\boldsymbol{\mu}_k^*)$, the asymptotic normality of $\sqrt{\frac{1+\beta}{1-\beta}} \nabla \ell_k^\beta(\boldsymbol{\mu}_k^\beta)$ (by Proposition 2, or alternatively by the Lindeberg-Feller CLT), and by invoking Slutksy's theorem that $\sqrt{\frac{1+\beta}{1-\beta}} \left(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k^* \right) \to \mathcal{N}\left(\tilde{\boldsymbol{\delta}}_k, (\boldsymbol{\mathcal{I}}_k^*)^{-1} \right)$.

Recalling the partition of the parameter vector into subsets of free and restricted parameters under the null hypothesis and the corresponding decomposition of the Fisher information matrix $\boldsymbol{\mathcal{I}}_k^*$, the deviance difference between the full and reduced models can be expressed as

$$\begin{aligned} D_{k,\beta}^{(r)} \left(\hat{\boldsymbol{\mu}}_k^{(F)}, \hat{\boldsymbol{\mu}}_k^{(R)} \right) = &\left(\frac{1+\beta}{1-\beta} \right) \left(\hat{\boldsymbol{\mu}}_k^{(F)} - \boldsymbol{\mu}_k^* \right)' \boldsymbol{\mathcal{I}}_k^* \left(\hat{\boldsymbol{\mu}}_k^{(F)} - \boldsymbol{\mu}_k^* \right) \\ &- \left(\frac{1+\beta}{1-\beta} \right) \left(\hat{\boldsymbol{\mu}}_{0,k}^{(R)} - \boldsymbol{\mu}_{0,k}^* \right)' \boldsymbol{\mathcal{I}}_{0,0,k}^* \left(\hat{\boldsymbol{\mu}}_{0,k}^{(R)} - \boldsymbol{\mu}_{0,k}^* \right) \\ &+ o_{\mathbb{P}}(1). \end{aligned} \tag{69}$$

As before, we focus on the latter term. The first of the Taylor expansions in Eq. (67) implies that

$$\nabla \ell_k^\beta(\boldsymbol{\mu}_k^*) = \boldsymbol{\mathcal{I}}_k^* \left(\boldsymbol{\mu}_k - \hat{\boldsymbol{\mu}}_k^* \right) + o_{\mathbb{P}}(1),$$

and so the reduce model deviance can be written as

$$\left(\frac{1+\beta}{1-\beta} \right) \nabla_{\boldsymbol{\mu}_{0,k}} \ell_k^\beta(\boldsymbol{\mu}_k^*)' \boldsymbol{\mathcal{I}}_{0,0,k}^{-1} \nabla_{\boldsymbol{\mu}_{0,k}} \ell_k^\beta(\boldsymbol{\mu}_k^*) + o_{\mathbb{P}}(1). \tag{70}$$

Following a series of arguments analogous to those presented in the proof of Theorem 2, the deviance difference can be shown to be equivalent to

$$D_{k,\beta}^{(r)} = \left(\frac{1+\beta}{1-\beta}\right)\left(\hat{\boldsymbol{\mu}}_{1,k}^{(F)} - \boldsymbol{\mu}_{1,k}^{*}\right)' \tilde{\boldsymbol{\mathcal{I}}}_{k}^{*}\left(\hat{\boldsymbol{\mu}}_{1,k}^{(F)} - \boldsymbol{\mu}_{1,k}^{*}\right) + o_{\mathbb{P}}(1), \tag{71}$$

where $\tilde{\boldsymbol{\mathcal{I}}}_{k}^{*} := \boldsymbol{\mathcal{I}}_{1,1,k}^{*} - \boldsymbol{\mathcal{I}}_{1,0,k}^{*}\left(\boldsymbol{\mathcal{I}}_{0,0,k}^{*}\right)^{-1}\boldsymbol{\mathcal{I}}_{0,1,k}^{*} = \left(\boldsymbol{\mathcal{I}}_{k}^{*-1}\right)_{1,1}^{-1}$. That $\sqrt{\frac{1+\beta}{1-\beta}}\left(\hat{\boldsymbol{\mu}}_{1,k}^{(F)} - \boldsymbol{\mu}_{1,k}^{*}\right) \to \mathcal{N}\left(\boldsymbol{\delta}_{k}, \left(\boldsymbol{\mathcal{I}}_{k}^{*-1}\right)_{1,1}\right)$ follows from previous arguments. We thus conclude that, in the special case of the contemporaneous-event model, the adaptive deviance difference converges to a non-central χ^2-distributed random variable with $M^{(r)}$ degrees of freedom and non-centrality parameter $\nu_k^{(r)} := \boldsymbol{\delta}_k' \left(\boldsymbol{\mathcal{I}}_{k}^{*-1}\right)_{1,1}\boldsymbol{\delta}_k$ under a sequence of local alternative hypotheses:

$$\left[D_{k,\beta}^{(r)}\left(\hat{\boldsymbol{\mu}}_{k}^{(F)}, \hat{\boldsymbol{\mu}}_{k}^{(R)}\right) \Big| H_{1,k}^{\beta}\right] \to \chi^2(M^{(r)}, \nu_k^{(r)}). \tag{72}$$

□

References

[1] S. van de Geer, P. Buhlmann, Y. Ritov, and R. Dezeure, “On asymptotically optimal confidence regions and tests for high-dimensional models,” *Ann. Statist.*, vol. 42, no. 3, pp. 1166–1202, 2014.

[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Courier Corporation, 2012.

[3] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank, “The time-rescaling theorem and its application to neural spike train data analysis,” *Neural Computation*, vol. 14, pp. 325–346, 02 2002.

[4] D. Ba, S. Temereanca, and E. N. Brown, “Algorithms for the analysis of ensemble neural spiking activity using simultaneous-event multivariate point-process models,” *Frontiers in Computational Neuroscience*, vol. 8, no. 6, 2014.

[5] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, vol. 1. New York, NY: Springer, 2nd ed., 2003.

[6] M. J. Crowder, “Maximum likelihood estimation for dependent observations,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 38, no. 1, pp. 45–53, 1976.

[7] B. M. Brown, “Martingale central limit theorems,” *Annals of Mathematical Statistics*, vol. 42, pp. 59–66, February 1971.

[8] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[9] 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*, vol. 115, no. 17, pp. E3869 – E3878, 2018.

[10] R. R. Davidson and W. E. Lever, “The limiting distribution of the likelihood ratio statistic under a class of local alternatives,” *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 32, no. 2, pp. 209–224, 1970.

[11] S. S. Wilks, “The large-sample distribution of the likelihood ratio for testing composite hypotheses,” *Ann. Math. Statist.*, vol. 9, no. 1, pp. 60–62, 1938.

[12] 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.

[13] S. Mukherjee and B. Babadi, “A statistical approach to dynamic synchrony analysis of neuronal ensemble spiking,” in *Proceedings of the 2019 Asilomar Conference on Signals, Systems, and Computers, Nov. 3-6, Pacific Grove, CA*, 2019.