

Disambiguating Different Covariation Types

Carlos D. Brody*

Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA 91125, U.S.A.

Covariations in neuronal latency or excitability can lead to peaks in spike train covariograms that may be very similar to those caused by spike timing synchronization (see companion article). Two quantitative methods are described here. The first is a method to estimate the excitability component of a covariogram, based on trial-by-trial estimates of excitability. Once estimated, this component may be subtracted from the covariogram, leaving only other types of contributions. The other is a method to determine whether the covariogram could potentially have been caused by latency covariations.

1 Introduction ---

A companion article, "Correlations Without Synchrony," contained elsewhere in this issue, has described how covariations in neuronal latency or excitability can lead to peaks in covariograms¹ that are very similar to peaks caused by spike synchronization. Since such peaks should be interpreted very differently to spike synchronization peaks, it is important to tell them apart. This note describes two methods that attempt to do this. The central idea is to use trial-by-trial information (e.g., number of spikes fired by each cell in each trial) as well as trial-averaged information (e.g., the joint peristimulus time histogram JPSTH) in trying to distinguish the various cases from each other. Friston (1995; see also Vaadia, Aertsen, & Nelken, 1995) has previously proposed a method to identify excitability covariations. It is in its use of trial-by-trial data that the excitability covariations method proposed here is most importantly different from that proposed by Friston.

The two methods described in this article differ in both the conclusions that can be drawn from them and their computational complexity. The excitability covariations method is computationally very simple, and when it indicates the presence of excitability covariations, it does so unequivocally. The latency covariations method is much more computationally demanding, and although it can determine whether latency covariations could po-

* Present address: Instituto de Fisiología Celular, UNAM, México D. F. 04510, México.

¹ Both here and in the companion article *covariogram* is used as an abbreviation for *shuffle-corrected cross-correlogram*, and is represented with the letter *V*.

tentially have generated the covariogram being analyzed, it cannot prove that they did so.

The notational conventions used here are as follows: $S_i^r(t)$ is the binned spiking response of cell i during trial r , the symbol $\langle \rangle$ represents averaging over trials r , the symbol \odot represents cross-correlation, and $\text{cov}(a, b)$ represents the covariance of two scalars a and b . The covariogram of two spike train sets is defined as $V = \langle S_1^r \odot S_2^r \rangle - \langle S_1^r \rangle \odot \langle S_2^r \rangle$, and the unnormalized JPSTH matrix is $J(t_1, t_2) = \langle S_1^r(t_1) S_2^r(t_2) \rangle - \langle S_1^r(t_1) \rangle \langle S_2^r(t_2) \rangle$.

2 Excitability Covariations

Let us model the responses of a cell as the sum of a stimulus-induced component plus a background firing rate (see the companion article in this issue)

$$\underbrace{F^r(t)}_{\text{Firing rate during trial } r} = \underbrace{\zeta^r Z(t)}_{\text{Stimulus induced}} + \underbrace{\beta^r B}_{\text{Background}}. \quad (2.1)$$

$F^r(t)$ is the model's expected response when its parameters are fixed at values appropriate for trial r , $Z(t)$ is the typical stimulus-induced firing rate, B is a constant function over the time of a trial, representing the typical background firing rate, and two gain factors, ζ^r and β^r , which may be different for different trials r , represent possible changes over trials in the state of the cell. Again following the companion article, when two such model cells (indexed by the subscripts 1 and 2) interact only through their gain factors, their covariogram is described as being due to excitability covariations and is:

$$\begin{aligned} V = & \text{cov}(\zeta_1, \zeta_2) Z_1 \odot Z_2 + \text{cov}(\zeta_1, \beta_2) Z_1 \odot B_2 \\ & + \text{cov}(\beta_1, \zeta_2) B_1 \odot Z_2 + \text{cov}(\beta_1, \beta_2) B_1 \odot B_2. \end{aligned} \quad (2.2)$$

Now let us take the experimental data, and in order to estimate the excitability component of a covariogram, let us characterize each of the two recorded cells using models of the form of equation 2.1. We must fit the parameters ζ^r , β^r , $Z(t)$, and B to each cell. This can be done separately for each cell.

It will be assumed that spikes during a short time preceding each trial have been recorded; this time will be written as $t < t_0$. (For sensory neurons, t_0 may be set to be the stimulus start time, but when recording from more central or motor neurons during complex behavioral tasks, it is necessary to set t_0 to be the very beginning of the entire trial—possibly far removed from the time period of interest, making the appropriateness of the estimate to be made questionable.) The mean background B can then be estimated from the average number of spikes per bin during $t < t_0$. In turn, the mean stimulus-induced component $Z(t)$ can be estimated from the PSTH (peristimulus time

histogram), labelled $P(t)$), as

$$Z(t) = P(t) - B. \quad (2.3)$$

Let $S^r(t)$ be the experimentally observed spike train in trial r . To be consistent with the number of pretrial spikes observed in that trial r , set β^r so that

$$\sum_{t < t_0} \beta^r B = \sum_{t < t_0} S^r(t), \quad (2.4)$$

and to be consistent with the total number of spikes observed in trial r , set ζ^r so that

$$\sum_t [\beta^r B + \zeta^r Z(t)] = \sum_t S^r(t). \quad (2.5)$$

Doing this for all trials sets all the necessary parameters that characterize the cell's response. The characterization is in terms of firing rates ($Z(t)$ and B) and across-trial changes in the firing rates (ζ^r and β^r). Once these parameters are set for both cells, the modeled excitability covariogram can be calculated from equation 2.2.

The excitability covariogram can be compared to the experimental covariogram, and subtracting it from the experimental covariogram can be thought of as removing the excitability components. Figure 1 illustrates the application of this straightforward procedure to two artificial cases: one with pure excitability covariations and one with both excitability and spike timing covariations (see the companion article).

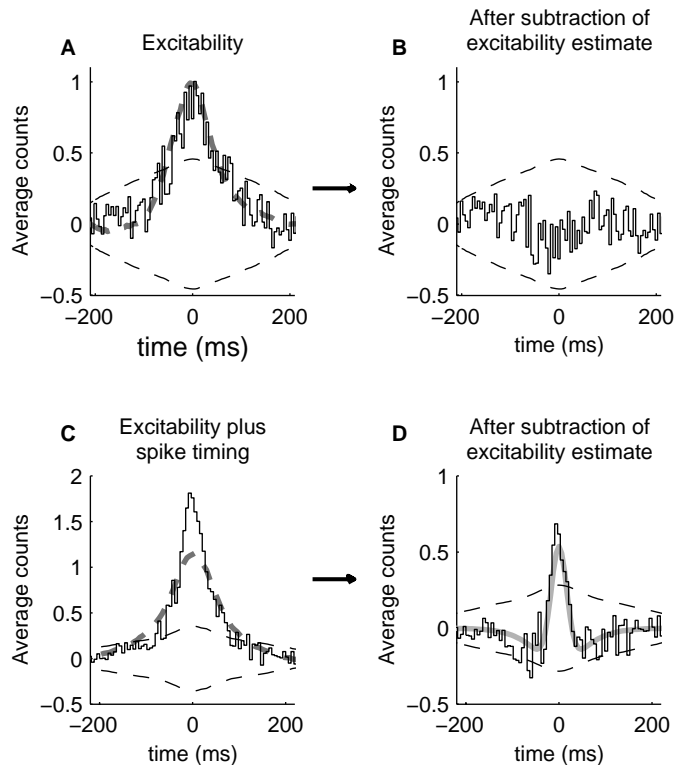
3 Previous Work on Excitability Corrections

Excitability covariations lead to (unnormalized) JPSTH matrices which are linear sums of separable components. That is, if t_1 and t_2 are the running times for cell 1 and 2, each of the JPSTH components can be factored into a function of t_1 times a function of t_2 (see equation 3.4 in the companion article and Friston, 1995). Given that a particular JPSTH matrix, and ensuing covariogram, are suspected of having been caused by excitability covariations, the question is, How can the JPSTH be split into a sum of separable components? An infinity of possible solutions exists.

Friston (1995; see also Vaadia et al., 1995) has described one solution choice, based on singular value decomposition (SVD) of the JPSTH matrix. The SVD, a well-known process, decomposes any matrix into a sum of separable, mutually orthogonal components by finding the sequence of such components that capture the most amount of squared power in the matrix. For example, the first component will be the separable matrix with the smallest possible sum of squared differences between its elements and those of the original matrix; the next component operates on

the same principle after having subtracted the first component from the original matrix; and so on. Using the SVD has two major advantages: (1) the first component is guaranteed to be the best single separable description of the original JPSTH matrix, in the squared-error sense just described; and (2) as many components as are necessary to describe the JPSTH matrix will be produced.² However, using the SVD has at least one major disadvantage (Friston, 1995): the components it produces will be orthogonal to each other. There is no reason to suppose that physiological components would be orthogonal in this sense. Furthermore, it must be remembered that while excitability covariations imply JPSTH separability, the converse is not necessarily true. The JPSTH is obtained through averaging over trials, and the average of a set of matrices being well described by a few separable components does not imply that each of the matrices that were

² As Vaadia et al. (1995) point out in their reply to Friston (1995), if too many components are needed to describe the JPSTH matrix, a spike timing synchronization interpretation may be far simpler and more parsimonious than the SVD-derived one.



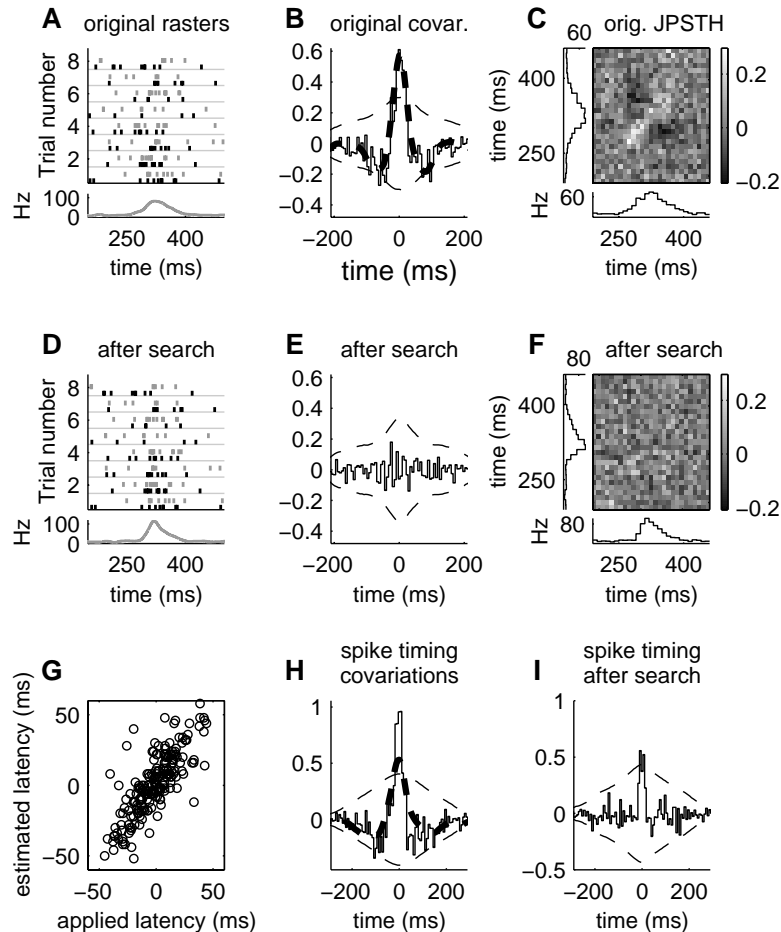
averaged (the individual trials) was also well described by the same components.

An alternative choice of separable components was made here. We required that only four components (which need not be orthogonal to each other) be used and that they be based on two time-independent and two time-dependent functions: B_1 , B_2 , $Z_1(t)$, and $Z_2(t)$. The form of these functions was estimated by assuming that the B 's represent background firing rates and the Z 's stimulus-induced responses. Most importantly, these physiological-interpretation-based assumptions allowed estimating the magnitude of each component from trial-by-trial information available in the data (equations 2.4 and 2.5). In contrast, the SVD method ignores trial-by-trial information. In most cases, one stimulus-induced component will be dominant, and the approximation of describing the data using only one B and one Z per cell will be a good one. However, if there is more than one important stimulus-induced excitability component, the method proposed

Figure 1: *Facing page*. (A) Covariogram of artificial spike trains, generated using excitability covariations only. On each trial, the time-varying firing rate of two independent Poisson cells was first multiplied by the same scalar gain factor ζ , drawn anew from each trial from a gaussian with unit mean and standard deviation. ζ was set to zero if negative. For details of spike train generation, see Figure 3 of the companion article. Overlaid on the covariogram as a thick dashed line is the excitability covariogram estimate from equation 2.2, using the procedure described in the text. Thin dashed lines are significance limits. (B) Same covariogram as (A) after subtraction of the estimated excitability component. (C) Covariogram of artificial spike trains constructed with both spike timing *and* excitability covariations (see the companion article). Although the shape does not obviously indicate two separate components, we can use more than just the shape to separate the two. The thick dashed line is the excitability covariogram estimate. Spike train details: On each of two hundred trials, a scalar ζ was drawn from a gaussian with unit mean and unit standard deviation (ζ was set to zero if negative). A spike train was then drawn from a Poisson source with time-varying firing rate $\zeta \cdot (70 \text{ Hz}) \cdot ((t - 120)/30) \cdot \exp((150 - t)/30)$ if $t > 120$, zero otherwise, with t in milliseconds. Spike times were then jittered twice, by a gaussian with zero-mean and 12 ms standard deviation; the result of the first jittering was assigned to cell 1, the result of the second to cell 2. Finally, 10 Hz background uncorrelated firing was added to both cells. (D) Same covariogram as in (C) after subtraction of the excitability component. A clear peak, indicative of a covariation other than an excitability covariation, can be seen. Since the spike trains were artificial, we know that this is a spike timing covariation and can predict the expected covariogram shape based on knowledge of the spike timing covariation parameters used to construct the rasters. The predicted shape is shown as a thick gray line. It matches the residual covariogram well. Subtracting the excitability covariogram estimate has accurately revealed the spike timing component of the covariogram.

here will not describe the data well; in such cases, the SVD method may be the more robust one.³

³ The referees informed me that although the fact remains unpublished, the JPSTH software used and distributed by Aertsen and colleagues contains an excitability correction term equal to $\text{cov}(n_1, n_2)P_1(t_1)P_2(t_2)$, where $\text{cov}(n_1, n_2)$ is the covariance in the total spike counts of the two cells and P_1 and P_2 are the two cells' PSTHs. In the absence of background firing, the JPSTH equivalent of equation 2.2 reduces to Aertsen et al.'s term. Thus, taking proper account of background firing when it is present is the principal extension provided here. The large number of covariograms in the literature to which a correction such as Aertsen and colleagues' or the one described here have not been applied (see the companion article) attests to the unfortunate fact that most investigators remain unaware of the need for them.



4 Latency Covariations

When estimating excitability covariations, the number of spikes fired provides a convenient measure of excitability for each individual trial and for each cell. For latency covariations, there may not be such a straightforward measure of latency available. Let us assume for the moment that there is one and that the estimated latency of cell i during trial r has been labeled t_i^r . Then, removing the effect of latency variations from the covariogram is simply a matter of back-shifting the spike trains: take the original spike trains $S_1^r(t)$ and $S_2^r(t)$ and shift them in time so as to build the covariogram of the set of spike trains $S_1^r(t - t_1^r)$ and $S_2^r(t - t_2^r)$.

Even when estimates of t_i^r are not directly available, we may wish to ask whether the observed covariogram could have been caused by latency covariations. For this to be the case, there must exist a set of time shifts t_i^r such that:

- The covariogram of $S_1^r(t - t_1^r)$ and $S_2^r(t - t_2^r)$ is zero within sampling noise.
- The covariogram predicted by the averages $\langle S_1^r(t - t_1^r) \rangle$ and $\langle S_2^r(t - t_2^r) \rangle$ and the shifts t_i^r must be similar to the original covariogram. Recalling that $\langle \rangle$ represents averaging over trials r and defining $\hat{P}_i(t) = \langle S_i^r(t - t_i^r) \rangle$, the predicted covariogram is

$$V = \hat{P}_1(t) \odot \hat{P}_2(t) - \langle \hat{P}_1(t + t_1^r) \rangle \odot \langle \hat{P}_2(t + t_2^r) \rangle.$$

The first condition ensures that the particular spike trains obtained from the experiment are consistent with their covariogram's being due to latency covariations; the second condition ensures that the covariogram is well predicted

Figure 2: *Facing page*. Latency search results. (A) Original rasters of artificial spike trains, constructed with latency covariations. Two independent Poisson cells were simulated; the raster pair for each trial was then shifted in time by a random amount, drawn anew for each trial, from a gaussian distribution with mean 0 ms and standard deviation 15 ms. For details of spike train generation, see Figure 2 of the companion article. (B) Covariogram of original rasters. Overlaid on it as a thick dashed line is the prediction derived from the latency search results (see condition 2.2 in the text). Thin dashed lines are significance limits. (C) Original JPSTH. Gray scale is correlation coefficient. (D) Rasters from (A), back-shifted by the latencies estimated from the search. (E) Covariogram of back-shifted spike trains, as in (D); no significant peaks are left. (F) JPSTH of back-shifted spike trains as in (D). (G) Scatter plot of estimated latencies versus applied latencies (the latter known since these spike trains were artificial). In both construction and estimation, the latencies of both cells were the same. (H) Covariogram of spike trains constructed with spike timing covariations (see the companion article). Overlaid on it as a thick dashed line is the prediction derived from the latency search procedure, run on these rasters even though they were known to contain spike timing covariations only. (I) Covariogram of spike trains used in (H) after applying the latency search procedure. Although the peak is reduced, it is still clearly there.

by global (latency) interactions only, and not through individual spike timing coordination between the two cells. Note that the existence of a set of time shifts satisfying conditions 2.1 and 2.2 merely shows that latency covariations could have generated the covariogram. It does not prove that they did so.

In the example illustrated in Figure 2, time shifts satisfying condition 2.1 were found by searching for the minimum of the cost function $G = \int V^2(\tau) d\tau$, where $V(\tau)$ is the covariogram of the set of spike trains $S_1^r(t - t_1^r)$ and $S_2^r(t - t_2^r)$. The minimization was an iterated line search along coordinate dimensions, with the search space reduced by using the restriction $t_1^r = t_2^r$, based on the assumption that the largest latency covariation peaks, and hence the best match to the covariogram, would be achieved when the latency shifts in both neurons were perfectly correlated.⁴ For each trial r , the cost function was evaluated at values of t^r ranging from -100 to $+100$ ms, in steps of 10 ms; t^r was then set to the shift that generated the smallest value of G ; and the search then proceeded to test shifts for the next trial. G typically asymptoted at a minimum after four or five loops through the whole set of trials. Using spike train sets that had 200 trials for each cell, search time using C on a 266 MHz Pentium running under Linux was approximately 10 minutes. Panel E in Figure 2 shows how the covariogram satisfies condition 2.1 after the minimization. The result of this search also satisfied condition 2.2, as shown in panel B.

Panels H and I in Figure 2 are based on spike trains constructed without latency covariations, using instead spike timing covariations only (see the companion article). The latency search process was run on these spike trains; as can be seen on both panels H and I, it cannot fully account for the covariogram. However, these panels also show that the latency search generated a covariogram that began to approximate the original one; for weaker yet still significant spike timing covariations, the latency search process could have been successful. This underscores that a positive result from the latency search can be taken as suggestive but never conclusive. It is only the negative result that can be taken as conclusive, since it demonstrates that the covariogram was not due to latency interactions alone.

Acknowledgments

I am grateful to John Hopfield and members of his group, Sanjoy Mahajan, Sam Roweis, and Erik Winfree, for discussion and critical readings of the manuscript for this article. I also thank John Hopfield for support. I thank George Gerstein, Kyle Kirkland, and Adam Sillito for discussion, and the anonymous reviewers for helpful comments. This work was supported by a Fulbright/CONACYT graduate fellowship and by NSF Cooperative Agreement EEC-9402726.

⁴ In fact the artificial rasters used here were constructed with $t_1^r = t_2^r$, so the assumption was known to be correct. Peter König (personal communication) has suggested initializing the search by aligning the rasters so as to minimize the width of individual PSTHs. This would generate a sharp shuffle corrector K and would thus be consistent with a tall latency covariations peak (see section 3.1 in the companion article).

All simulations and analyses were done in Matlab 5 (Mathworks, Inc., Natick, MA), except for the latency search, which also used some subroutines hand-compiled into C. The code for all of these, including the code to reproduce each of the figures, can be found at <http://www.cns.caltech.edu/~carlos/correlations.html>.

References

- Friston, K. J. (1995). Neuronal transients. *Proceedings of the Royal Soc. of London Series B Biological Sciences*, 261, 401–405.
- Vaadia, E., Aertsen, A., & Nelken, I. (1995). "Dynamics of neuronal interactions" cannot be explained by "neuronal transients." *Proceedings of the Royal Soc. of London Series B Biological Sciences*, 261, 407–410.

Received October 20, 1997; accepted November 25, 1998.