COMPRESSED AND DISTRIBUTED SENSING OF MULTIVARIATE NEURAL POINT PROCESSES

Karim G. Oweiss

Electrical and Computer Engineering Department
Michigan State University, USA

ABSTRACT

This paper proposes a new approach to simultaneously estimate time-varying intensity functions of multiple point processes from their continuous-time signal representation. We use models of neural response properties in the cortex to illustrate the theory of the proposed approach. Based on sparse representation of the continuous-time signals in the context of compression, it is shown that intensity functions can be approximated reasonably well without the need to decompress and classify the source signals. The approach is best suited for the case when multiple point processes are characterized by non-binary spike waveforms observed with an array of sensors. When spike waveforms from different sources are correlated, the estimated intensities can be inaccurate due to spike classification errors. We therefore build on our previous work for separating correlated spike waveforms to enable enhanced separation of those intensity functions. We finally show that this framework leads to substantial savings in computational complexity for real time operation in resource constrained signal processing systems.

Index Terms- point process, rate estimation, sparse representation, neural recordings, brain machine interface, compressed sensing.

1. INTRODUCTION

Advanced sensing methodologies have increased the shear amount of data that needs to be acquired from distributed signal sources to infer useful information about phenomena of interest. In attempting to circumvent the information retrieval problem from large volumes of measured signals, one is interested in inferring the parameters of the random process underlying the generation the observed data. In many applications, point processes have been typically used to describe discrete event data. Discharge rate of nerve cells is a classical example where a point process is used to describe the neuronal response to some intrinsic and extrinsic conditions over time.

For the case of neuronal point processes, a neuron "fires" an action potential (AP) in the form of a *spike* waveform indicating that the membrane potential has exceeded the voltage threshold due to the influence of ionic current from gated channels in the membrane. The information in the spike train lies foremost in the time of occurrences of each spike. It is thus of paramount importance to estimate the intensity function underlying the point process to understand how the information is encoded in a larger neuronal population. It is believed that this function is the fundamental source of information coding in the nervous system.

Recently, Brain Machine Interface (BMI) applications demonstrated the utility of implantable high-density microelectrode arrays in recording ensemble neural activity in the cortex that can be subsequently decoded to control external devices [1]. A challenging problem in this application is the ability to transmit large throughput neural data from the implanted device to the outside world for further analysis. Brain implants are thus a novel example of resource-constrained signal processing systems in which severe restrictions in size, power dissipation, and energy consumption are imposed. It is therefore desired to minimize the computational complexity of the implantable system, yet infer the useful information, in this case the intensity functions of the recorded population, early in the data stream.

Previous attempts to estimate rate functions of neuronal point processes from single trials have been reported [2]. Thereof, an empirical approach using arbitrary chosen kernel functions was proposed. It was reported that the main parameters influencing the quality of the estimator were the shape of the basis and its width, which determines the temporal resolution of the rate estimator. In a previous study [3], we proposed a distributed compression technique relying on sparse representation of the spatiotemporal patterns of neural activity. The compressed signals, typically resulting from multiple neurons, have to be decompressed once received outside the cortex, then separated to multiple single-neuron spike trains.

The objective of this paper is to propose a new technique based on the compression scheme proposed in [3] for directly estimating intensity functions of neural point processes from continuous-time neural data that have been regularly sampled and compressed with a sparse representation operator. Herein, the shape of the kernel will depend on the particular choice of the basis that *sparsifies* the data as much as possible.

2. THEORY

2.1 Single point process model

For a point process with deterministic continuous-time intensity function $r(t)$, the integral over a finite interval $[T_a, T_b]$ represents the expected number of events N encountered during that interval

$$E[N] = \int_{T_a}^{T_b} r(t)dt \tag{1}$$

In many applications, the observations consist of the time of occurrence of the N events. From this knowledge, the goal is to estimate the intensity function that resulted in the observed discrete event process. In practice, this is done by binning the data in equal intervals and counting the number of events within that interval. Generally speaking, the large variability observed in the point process data mandates averaging across multiple trials to reduce the statistical variations in the estimated intensity function. An example is illustrated in Figure 1 for 30 trials based estimator.

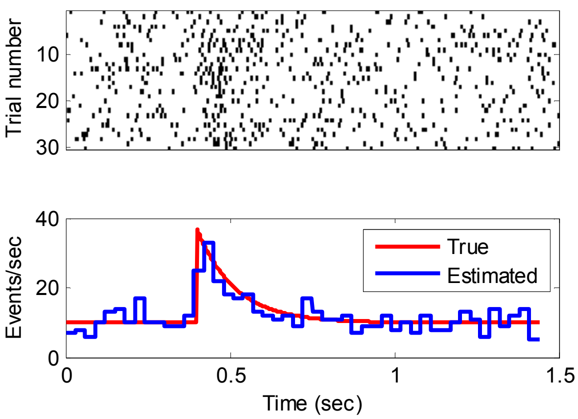

Figure 1: Intensity function and associated point process for 30 trials. Estimate using a 30 ms bin width averaged across trials.

Outside the range where the intensity function is time varying, the variability observed in Figure 1 can be regarded as "process noise". We'll use the additive noise model below to describe the observed point process as

$$\boldsymbol{y} = \sum_{i=1}^{N} \delta(n - n_i) = \boldsymbol{e} + \boldsymbol{z} \qquad (2)$$

where both $\boldsymbol{e}$ and $\boldsymbol{z}$ are binary. The occurrence of 1's in $\boldsymbol{e}$ is very "regular" in the sense that $r(t)$ can be reliably estimated from observing $\boldsymbol{e}$ without trial averaging. This is feasible if $r(t)$ has some discontinuities that is translated into a sudden change in the event count. On the other hand, the occurrence of 1's in $\boldsymbol{z}$ can be regarded as *spurious* events that do not positively contribute to estimating strong variations in $r(t)$ in a given trial and therefore their effect is reduced through trial averaging. This is most pronounced when $r(t)$ is very smooth and approaches a constant.

2.2 Sparse representation

In certain applications, the occurrence of an event in the point process is typically represented by a short rapid transient signal waveform - a spike- with duration T_e. Measurement noise can sometimes obscure the spike waveforms if the Signal-to-Noise Ratio (SNR) is low. To estimate the event occurrence times and hence the intensity, a threshold is calculated based on the noise level. An event is declared whenever the spike surpasses that threshold. The specific time of event occurrence is arbitrarily taken from within the spike interval T_e (e.g. the zero crossing of the spike). However, the measurement noise may lead to erroneous event times or excess event counts thereby contributing to a poor estimator of the rate.

Assuming that the continuous-time signals are adequately sampled by f_s, each event is thus represented by the $T_e f_s \times 1$ dimensional vector $\boldsymbol{s}$. If $\boldsymbol{s}$ is known to be highly compressible in a known basis (such as wavelets), a sparse representation operator is desired to map $\boldsymbol{s}$ to the smallest number of transform coefficients, ideally *one* most significant coefficient centered at the event time. If a fixed universal basis is used, the sparse representation results in possibly more than one transform coefficient depending on an acceptable level of reconstruction error [4]. When the events are "regularly" spaced and the basis is matched to $\boldsymbol{s}$, the sparse representation can be an efficient way to approximate any sudden jump in the rate. This can be seen by observing that: first, the most significant coefficients in this best basis are much higher in magnitude than any other basis; and second, they occur following the same regularity pattern as the events in $\boldsymbol{e}$. Since sudden jumps in the rate are associated with a large increase/decrease in the number of events, this results in *ridges* in the sparse representation that indicate the presence of a sudden jump in the rate**.**

Let's explore this case further. With the sparse representation, (2) can be expressed as

$$\boldsymbol{y}_j = \boldsymbol{e}_j + \boldsymbol{z}_j \qquad j = 0,\dots,J-1 \qquad (3)$$

where J is the total number of basis. The rate in a given basis span (subband j) is expressed as

$$r_j(n) = \sum_{i=1}^{N_t} w_j(n - n_i) + \sum_{k=1}^{N_s} w_j(n - n_k) \qquad (4)$$

where w_j is the wavelet basis spanning subband j, N_t is the number of events in $\boldsymbol{e}$, and N_s is the number of events in $\boldsymbol{z}$ so that $N = N_t + N_s$. Equation (4) is a direct result of convolving the basis kernel with the irregularly spaced impulse train $\boldsymbol{y}$ in (2). When $\boldsymbol{y}$ is non-binary, (2) and (4) can be used to obtain

$$r_j(n) = \sum_{i=1}^{N_t} \boldsymbol{s}_j(n - n_i) + \sum_{k=1}^{N_s} \boldsymbol{v}_j(n - n_k) \qquad (5)$$

where $\boldsymbol{s}_j$ and $\boldsymbol{v}_j$ denote the projection of the spike waveforms with time of occurrences in $\boldsymbol{e}$ and $\boldsymbol{z}$, respectively onto the span of the basis w_j. The effect of the process noise term $\boldsymbol{z}$ is averaged out if the kernel support is large enough and will depend on the "resting time" of the rate function to some constant value.

2.3 Single sensor of multiple point processes

Let's now consider the case where we observe $P > 1$ point processes characterized by their observed spike trains. Each point process is characterized by a distinct spike waveform $\boldsymbol{s}_p$. The model in (2) can be expanded to yield

$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{V} \qquad (6)$$

where the rows of $\boldsymbol{S}$ and $\boldsymbol{V}$ are the spike trains from the P sources. The sparse representation of $\boldsymbol{Y}$ will therefore be

$$\boldsymbol{Y}_j = \boldsymbol{S}_j + \boldsymbol{V}_j \qquad j = 0,\dots,J-1 \qquad (7)$$

From $\boldsymbol{Y}_j$, the goal now is to estimate the *P*x*N* matrix of intensity functions $\boldsymbol{R}_j$ where each entry is given by

$$r_j(p,n) = \sum_{i=1}^{N_t} \boldsymbol{S}_j[p, n-n_i] + \sum_{k=1}^{N_s} \boldsymbol{V}_j[p, n-n_k] \qquad (8)$$

When a single sensor is used, the problem becomes more challenging since only one estimator can be computed as

$$r_j(n) = \sum_{p=1}^{P} \sum_{i=1}^{N_t} \boldsymbol{S}_j[p, n-n_i] + \sum_{p=1}^{P} \sum_{k=1}^{N_s} \boldsymbol{V}_j[p, n-n_k] \qquad (9)$$

This estimator is a mixture of rate estimators of every point process. For a particular point process *q*, the goodness of its individual rate estimator in (9) will heavily depend on minimizing the contribution of the estimators from other processes, i.e.

$$\sum_{\substack{p=1\\p\neq q}}^{P} \sum_{i=1}^{N_t} \boldsymbol{S}_j[p, n-n_i] + \sum_{\substack{p=1\\p\neq q}}^{P} \sum_{k=1}^{N_s} \boldsymbol{V}_j[p, n-n_k] \rightarrow 0 \qquad (10)$$

If knowledge of the identity of the spike waveforms (i.e. to which point process they belong) and their locations in time is available, one may be able to sort out the specific time indices of event occurrences. However, the objective is to be able to estimate the rate of each individual process without actually detecting and sorting the individual spike waveforms. For a particular point process, forcing a minimum contribution from other sources has to rely on a pronounced difference in the spike waveform projections, if any, on the basis $\boldsymbol{w}_j$. Specifically, the degree to which the rate of the *q*th point process can be faithfully estimated from the mixed rate in (9) depends largely on the ratio of $< s_q, w_j >$ to $< s_p, w_j >, q \neq p$, where $< \cdot >$ denotes a dot product.

2.4 Multiple sensors of multiple point processes

The problem of sorting out distinct intensity functions from a mixed rate estimator obtained from a single sensor becomes very challenging if the spike waveform shapes associated with different sources are correlated. In this case, the ratio of $< s_q, w_j >$ to $< s_p, w_j >, q \neq p$, will be a function of the correlation coefficient between the spike waveforms. Separating multiple rates from the estimated mixture can be eased if an array of sensors is utilized to record the population. The model in (6) can be generalized in this case to incorporate *M* sensor array observation matrix expressed as

$$r_j[p,n] = \sum_{p=1}^{P} \sum_{i=1}^{N_t} \boldsymbol{A}\boldsymbol{S}_j[p, n-n_i] + \sum_{k=1}^{N_s} \boldsymbol{A}\boldsymbol{V}_j[p, n-n_k] \qquad (11)$$

where $\boldsymbol{A}$ denotes a full rank *M*x*P* mixing matrix. When the mixing matrix is not equal to the identity, each of the *P* spike trains can be represented on every sensor with energy inversely proportional to the distance from the signal source. For the scope of this paper, it will be assumed that the mixing is stationary, i.e., $\boldsymbol{A}$ is time invariant during the period where the intensity functions are being estimated. Therefore, $\boldsymbol{A}$ can be factored out to yield

$$r_j[p,n] = \sum_{p=1}^{P} \boldsymbol{A}\Big(\sum_{i=1}^{N_t} \boldsymbol{S}_j[p, n-n_i] + \sum_{k=1}^{N_s} \boldsymbol{V}_j[p, n-n_k]\Big) \qquad (12)$$

Now following (10), the contribution of the estimators from other processes can be reduced by observing that for those subbands where $< s_q, w_j >$ dominate over $< s_p, w_j >, q \neq p$, the signal subspace spans that of the *q*th column of $\boldsymbol{A}$, denoted $\boldsymbol{a}_q$, or equivalently the *q*th row of $\boldsymbol{S}$. A best basis search is performed to find those bases where the *q*th signal subspace remains invariant. A best basis tree is obtained in which the nodes with the best rate estimator for the *q*th process can be found. The process is repeated for other columns of $\boldsymbol{A}$. For the lack of space, the interested reader is referred to the description of the MASSIT technique reported in [6].

2.5 Process Noise

As stated before, estimating $\boldsymbol{R}$ reliably depends on minimizing the effect of measurement and process noises. The first effect is automatically minimized during the best basis search outlined above in which denoising [7] is further performed. The second effect is minimized by selecting a particular node among the best basis tree where the effect of irregularities caused by process noise is minimized. This would be proportional to the duration in which the rate has sudden jumps so that the regularity term dominates. If the time resolution of the basis is wide enough to capture the fastest time constant of the underlying intensity function, it is anticipated that within the best basis subset, the particular basis with support simultaneously matching the time constant of the regularity term and having the highest compression capability of the spike waveform will be the best in suppressing the process noise effect. The goal of the sparse operator can be therefore summarized as:

1- Reduce the contribution of the noise term on the rate estimator.
2- Determine the optimal resolution level in which the rate estimation error is minimized.

As a metric of performance, we'll use the mean square estimation error

$$MSE_{T_{ab}} = \tfrac{1}{Q}\sum_{i=1}^{Q}\sum_{T_a}^{T_b}(r[i]-\hat{r}[i])^2$$

where Q is the number of samples used to discretize $r(t)$.

3. RESULTS

We present here results from simulated data with actual non-binary spike waveforms. The point processes followed the model in [2] in which the event occurrences consisted of a background neural response and a *phasic* response as illustrated in Figure 1. The Poisson rate model $r[n] = \beta + \alpha\Psi(n - n_0)$ was based on three parameters β , α and n_0 denoting the background rate, the strength, and the onset of the evoked response, respectively. Figure 2 illustrates an example of the measured signals for noiseless single point process data. Figure 3 illustrates the MSE relative to the time domain rate estimator for two neurons. The spike waveform of each neuron is also shown. We used a fixed *symmlet*4 basis for the sparse representation operator. Minimum error rate estimators are illustrated for the best basis projections. As can be seen, basis expressing the highest compression for each signal coupled with a support that lasts over a period comparable to the phasic response duration are preferred for capturing the regularity in event occurrences. We assessed in Figure 4 the performance with respect to the background rate. It was observed that the MSE increases for increasing background rate β . This is expected since the variance of the Poisson process is proportional to its average rate.

4. CONCLUSION

A new approach for simultaneously estimating rate functions of multiple discrete point processes was presented. Particular emphasis was given to the case when the point processes are characterized by non-binary spike waveforms in which sparse representation was shown to yield a better alternative to classical averaging across trials. A model of the point process was proposed that involves a regular term useful to estimate abrupt changes in the process, and an irregularity term modeling statistical variations across trials and is useful to estimate the constant portion of the time varying rate. Aside from compression advantages in a bandwidth-limited system, it was shown that the sparse representation of the non-binary spike trains yields additional advantages in simultaneously enabling both the regular and irregular components of the rate function to be reliably estimated. The irregular component can be estimated from relatively short basis support by averaging across a small number of trials, while the regular component is estimated from relatively longer basis support that depends on the time constant of the time varying portion of the rate. Besides the sparsity introduced in time, it was also argued that multiple sensors improve the ability to discern the desired rates based on additional sparsity in space. This would be obtained through a low rank approximation of the signal subspace components. The technique proposed is applicable to a wide variety of applications where multivariate spatiotemporal point processes with time varying intensity functions are observed in a distributed environment.

ACKNOWLEDGEMENT

This work was supported by NIH-NINDS grant # NS047516-01.

REFERENCES

[1] Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, et al., "Learning to Control a Brain–Machine Interface for Reaching and Grasping by Primates" (2003) *PLoS Biology* Vol. 1, No. 2, e42 doi:10.1371/journal.pbio.0000042.

[2] Nawrot M., et al., "Single trial estimation of neuronal firing rates: From single-neuron spike trains to population activity, J. of Neuro. Meth., 94, pp. 81-92, 1999

[3] K. G. Oweiss, "A Systems Approach for Data Compression and Latency Reduction in Cortically Controlled Brain Machine Interfaces," *IEEE Transactions on Biomedical Engineering,* vol. 53, no.7, pp. 1364 – 1377, July 2006

[4] D.L. Donoho, " Compressed Sensing" IEEE Transactions on Information Theory, Volume 52, Issue 4, April 2006, Pages: 1289 - 1306

[5] M. Vetterli, and C. Herley, "Wavelets and filter banks: Theory and Design," *IEEE Trans. On SP*, vol. 40, no. 9, pp. 2207-2232, 1992

[6] K G. Oweiss and D J. Anderson, "Tracking Signal Subspace Invariance for Blind Separation and Classification of Nonorthogonal Sources in Correlated Noise" EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 37485, 20 pages, 2007.

[7] D.L. Donoho, "De-Noising by Soft Thresholding," *IEEE Trans. on IT*, vol. 41, pp.613-627, 1995.

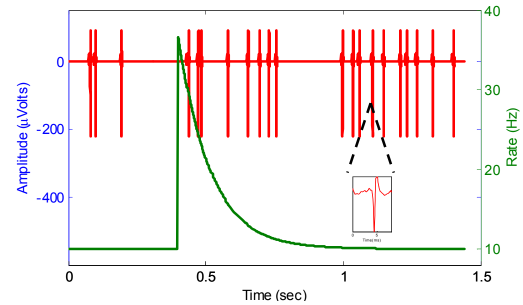

Figure 2: Sample rate function with non binary events.

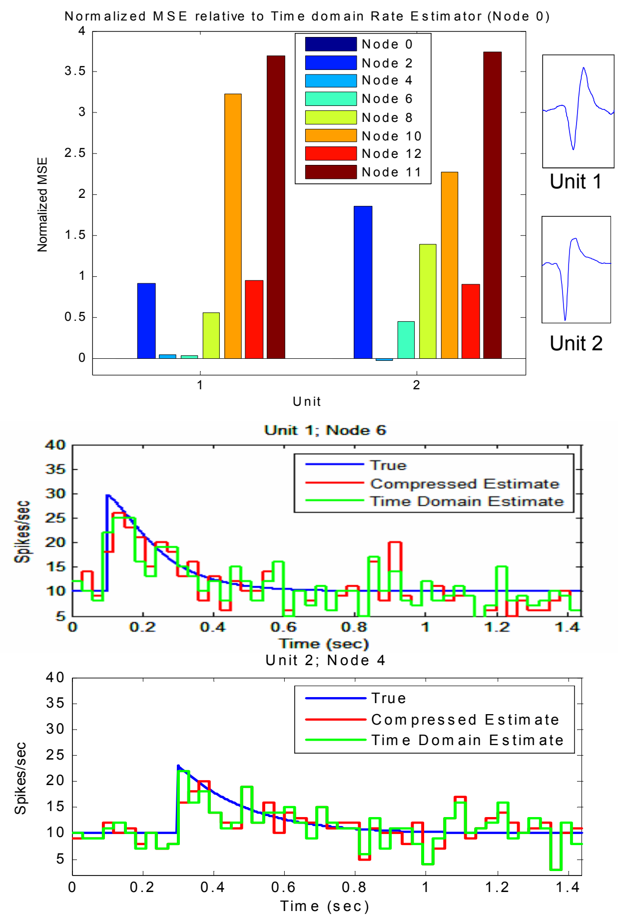

Figure 3: Performance for two non-binary spike trains.

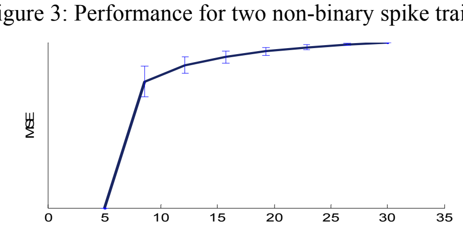

Figure 4: Performance versus background rate