A Bayesian approach for Simultaneous Spike Extraction and Sorting

Steven Le Cam[1], Harry Tran[1], Radu Ranta[1], and Valérie Louis-Dorr[1]

***Abstract*— The information processing in the brain is governed by large neural ensembles organized in networks. The firing of the neuronal cells in each neural population bring insight on the role of this structure for a given task or on its involvement in a pathology. Such investigations require the detection and sorting of neuronal spikes from the recordings. This is known as a challenging task due to the high level of surrounding local field potential (LFP) and to the strong similarities between the patterns to classify. In this article we present a new method for spike classification leaning on the separation between the spike waveform and the background LFP. The approach is rooted in a Bayesian framework integrating prior on the LFP spectral density and considering mixture of Gaussian for the classification. The algorithm provides both the posterior mean of the spike waveforms for each cluster as well as the posterior mean of each singular spike waveforms.**

***Index Terms*— Spike/LFP Separation, Neural Spike Classification, Bayesian Approach, Variational Approach**

I. Introduction

Extra-cellular micro-recordings give the opportunity to study the brain activity at the level of the neuronal cells. Identifying the cells at the origin of each spike gives insight on their involvement in the processing of a specific task or stimulus. Due to different characteristics of the recorded neurons (types of neurons, morphologies, relative positions with respect to the micro-electrode), the spiking waveforms of two different neurons will be seen in the signal with different shapes and amplitudes. It is then possible to distinguish them. The task is nonetheless challenging because the local field potential (LFP) generally masks most of the spiking activities, making it harder to distinguish between similar spike waveforms originating from different cells.

Classical approaches first consider a pass-band filter (generally in the band 300-6000 Hz): the activity below $300Hz$ is assumed to stand mainly for the LFP (*i.e.*, the activity of the neural tissue surrounding the electrode), while the higher frequency activities are considered to hold the main part of the spiking activities. The time instants of the spikes are identified by applying a threshold to the high frequency part of the signal. After rejection of false detection (artefacts), discriminant features are then extracted from the detected spike shapes, considering about 2 to $3ms$ time window around the maximum peak of the spike shape. Most efficient approaches have been reported to use wavelet decomposition and to select a given number of wavelet coefficients with distributions most differing from a normal distribution (*e.g.*, based on a Kolmogorov-Smirnov test) [1]. A clustering method is then applied on these features to identify the different neurons (cluster) from which the spikes originate. Clustering approach such as K-means or Superparamagnetic clustering (SPC) have been considered. The SPC approach has the advantage to consider non-parametric cluster shapes (in particular the gaussianity of the spike features is not assumed) and satisfactory performance are reported with this approach [2]. However it requires user parameter tuning (such as the parameter of temperature or stop criterion), and manual expert intervention are still necessary at the end of the clustering process for achieving satisfactory classification performance [2], [1]. Few parametric and non-parametric Bayesian inference methods have also been developed for spike sorting purposes [3], [4]. Unlike common approaches which produce hard label for each identified spikes, Bayesian approaches provide (posterior) probabilities for the labelling of each individual spike, which can be considered in subsequent analyses. Also, such probabilistic framework naturally take into account the inherent uncertainties in the recording. Most of the methods previously cited are considering only the high frequency part of the signal, however some significant features of the spikes may lie in the low frequency range. For this reason it is preferable to separate properly the LFP and the spiking activity while proceeding to the classification of the spikes. In this paper we propose an iterative Bayesian approach to properly separate the LFP from the spiking activities as well as classifying the spikes simultaneously. Similarly as in [5], we use a prior on the power spectral density of the LFP to proceed to its separation from the spikes. We rely a a classical spike detection step (300-6000 Hz band thresholding) to identify the spike support of the spikes, and we then proceed to the separation and to the classification iteratively from the wide band data. The spike waveforms are considered as a mixture of gaussian, and a Maximum A Posteriori (MAP) decision strategy is applied for the classification. We demonstrate that separating the spikes from the LFP indeed enhance the classification of the spikes. The advantage of separating the spiking waveforms from the LFP is evaluated on a realistic set of simulated data available online [6].

II. The generative model

In this section we present the general Bayesian formalization of the problem. We consider that each recorded channel y of length T can be decomposed as the addition of the LFP w and the spiking signal S, up to ϵ modelling the additive noise:

$$y = w + S + \epsilon \tag{1}$$

[1]Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France
steven.le-cam at univ-lorraine.fr

The spiking signal S can be written as a sum of N spike waveforms with a time support of few milliseconds, or alternatively as a sum of $N = \sum_{k=1}^{K} N_k$ waveforms originating from K different units (neurons):

$$S = \sum_{n=1}^{N} s_n * \delta_n = \sum_{k=1}^{K} \sum_{n=1}^{N_k} s_n^k * \delta_n^k \quad (2)$$

δ_n (or δ_n^k) being a dirac centered at the time location of the spike with waveform s_n (or s_n^k), and $*$ is the convolution product. $\delta^k = \sum_{n=1}^{N_k} \delta_n^k$ is the k^{th} neuron spike train. These waveforms can overlap in time, except for two waveforms originating from the same neuron, due to the assumption of refractory intervals. The waveforms s_n^k are supposed to follow a gaussian distribution over a mean waveform μ_k with uniform variance σ_k^2 (accounting for the uncertainties in the data, small variations of the spike waveform, drifts of the electrodes...). The spike waveforms S extracted from the signal y are then supposed to be a mixture of Gaussian. Let introduce the hidden variable Z, taking its values in the set $\{\lambda_1, ..., \lambda_K\}$, λ_k being the class of spike waveforms originating from the kth neuron:

$$p(S|Z) = \prod_{n=1}^{N} p(s_n|z_n) \quad (3)$$

$$p(s_n|z_n = \lambda_k) = \mathcal{N}(\mu_k, \sigma_k^2 \mathbf{I}_{T_s}) \quad (4)$$

with T_s the length of the spike waveform, $\mathbf{I}_{T_s}$ the $T_s \times T_s$ identity matrix. The posterior probabilities $p(z_n|s_n)$ are estimated providing a soft labelling of each spike. If a hard decision is needed, a Maximum A Posteriori (MAP) strategy can be used for the classification decision. Such as in [5], w follows a gaussian distribution $\mathcal{N}(0, \gamma\Sigma_w)$, where Σ_w is computed from the PSD of the signal y and γ is a scaling parameter. Finally, the noise ϵ follows a i.i.d gaussian distribution with mean μ_ϵ and variance σ_ϵ^2. This PSD being roughly learned from the data, we will let the method adjust between γ (accounting for the signal space) and σ_ϵ^2 (accounting for the noise space). The likelihood of the observation given the parameters of the model then writes:

$$p(y|w, S) \propto \exp \frac{1}{2\sigma_\epsilon^2} ||y - w - S||_2^2 \quad (5)$$

and the full probability law of the model is given by:

$$p(y, w, S, Z) = p(y|w, S)p(w)p(S|Z)p(Z) \quad (6)$$

$p(Z)$ being the prior probability over the coefficient of the gaussian mixture. The optimization of the model is carried out through a variational Bayesian procedure detailed in the next section.

III. Bayesian separation/classification

A. Variational optimization

A variational Bayesian (VB) approach is used to maximize the log-likelihood of the model, providing an iterative two step optimization scheme very similar to Expectation-maximization. Note that an EM procedure is not eligible here because it is not possible to find a closed-form for the joint posterior density of the latent variables (w, S, Z). The VB approach consists in introducing the probability density $q(w, S, Z) = q(w) \prod_{n=1}^{N} q(s_n)q(z_n)$, assuming the posterior independence of the latent variables. This density is optimized so that it approximates the true posterior density $p(w, S, Z|y)$. The update of these densities (E-step) are obtained as follows:

$$log(q(w)) = \langle log(p(y, w, S, Z) \rangle_{S,Z} \quad (7)$$

$$log(q(s_n)) = \langle log(p(y, w, S, Z) \rangle_{w,Z,S_{\setminus s_n}} \quad (8)$$

$$log(q(z_n)) = \langle log(p(y, w, S, Z) \rangle_{w,S,Z_{\setminus z_n}} \quad (9)$$

We do not to deal with overlapping spikes, thus we can consider the independence of the densities over the N spike S and label Z. The update of $q(z_n)$ and $q(s_n)$ can be simplified to:

$$log(q(s_n)) = \langle log(p(y, w, S, Z) \rangle_{w,z_n} \quad (10)$$

$$log(q(z_n)) = \langle log(p(y, w, S, Z) \rangle_{w,s_n} \quad (11)$$

The densities $q(w)$ and $q(s_n)$ can be identified as multivariate Gaussian distributions with mean $\hat{w}$ and $\hat{s}_n$ and covariance $\hat{\Sigma}_w$ and $\hat{\sigma}_n$ respectively:

$$\hat{w} = \gamma\Sigma_w(\gamma\Sigma_w + \sigma_\epsilon^2 \mathbf{I}_T)^{-1}(y - \sum_{n=1}^{N} \hat{s}_n * \delta_n - \mu_\epsilon) \quad (12)$$

$$\hat{\Sigma}_w = \sigma_\epsilon^2 \gamma\Sigma_w(\gamma\Sigma_w^{-1} + \sigma_\epsilon^2 \mathbf{I}_T)^{-1} \quad (13)$$

$$\begin{aligned} \hat{s}_n &= \hat{\sigma}_n^2((\sigma_\epsilon^{-2}(y - \hat{w} - \mu_\epsilon) * \delta_{-n})\Pi_{T_s} \\ &\quad + \textstyle\sum_{k=1}^{K} q(z_n = \lambda_k)\sigma_k^{-2}\mu_k) \end{aligned} \quad (14)$$

$$\hat{\sigma}_n^{-2} = \sigma_\epsilon^{-2} + \sum_{k=1}^{K} q(z_n = \lambda_k)\sigma_k^{-2} \quad (15)$$

with Π_{T_s} is the indicator function $\mathbb{1}_{[-T_s/2,\ T_s/2]}$, $(x * \delta_{-n})\Pi_{T_s}$ is the segment of the signal x on the time support of the n^{th} spike. We suppose that the matrix Σ_w is circulant [5], then the computation of the posterior covariance matrix of w (requiring inversion of large matrix) can be efficiently computed with Fast Fourier Transform (see section III-B.2). The approximate discrete posterior probabilities $q(z_n)$ are given by:

$$q(z_n = \lambda_k) \propto \pi_k \sigma_k^{-T_s} \exp^{\frac{-0.5}{\sigma_k^2}(||\hat{s}_n - \mu_k||_2^2 + T_s\hat{\sigma}_n^2)} \quad (16)$$

up to a normalization constant summing these probabilities to 1. The M-step provides an update of the parameters of the model $\theta = \{\gamma, \mu_\epsilon, \sigma_\epsilon^2, \mu_{1..K}, \Sigma_{1..K}\}$, by deriving the log likelihood of the full model under mean field approximation (*i.e.,* taking its expectation over the latent variables: $\langle log(p(y, w, S, Z; \theta^{old})) \rangle_{w,S,Z}$) with respect to each param-

eter and equating to 0:

$$\mu_\epsilon = \frac{1}{T}(y - \hat{w} - \sum_{n=1}^{N} \hat{s}_n * \delta_n) \quad (17)$$

$$\sigma_\epsilon^2 = \frac{1}{T}||y - \hat{w} - \sum_{n=1}^{N} \hat{s}_n * \delta_n - \mu_\epsilon||_2^2 + \text{trace}(\hat{\Sigma}_w) + T_s \sum_{n=1}^{N} \hat{\sigma}_n^2 \quad (18)$$

$$\gamma = \frac{1}{T}\text{trace}((\hat{w}^t\hat{w} + \hat{\Sigma}_w)\Sigma_w^{-1}) \quad (19)$$

$$\mu_k = \frac{\sum_{n=1}^{N} q(z_n = \lambda_k)\ \hat{s}_n}{\sum_{n=1}^{N} q(z_n = \lambda_k)} \quad (20)$$

$$\sigma_k^2 = \frac{1}{T_s}\frac{\sum_{n=1}^{N} q(z_n = \lambda_k)\ ||\hat{s}_n - \mu_k||_2^2}{\sum_{n=1}^{N} q(z_n = \lambda_k)} \quad (21)$$

with $\hat{w}^t\hat{w}$ the outer product of the line vector $\hat{w}$.

B. Initialization

1) Filtering and Spike detection: Following [7], [1], [2], the detection of the spikes is carried out by thresholding the high frequency part of the signal in the range $300-6000Hz$. The filter is a two pole butterworth band-pass filter with cut-off frequency set to $300Hz$ and $6000Hz$. The threshold is computed as a multiple of the backround noise standard-deviation σ_n within this high frequency band. The estimation of this parameter must be as robust as possible to artifact and to the spiking activity, and the Donoho-Johnstone universal thresholding method is used [8], [2]:

$$\sigma_n = median(\frac{|x|}{0.6745}) \quad (22)$$

and the threshold is set as $4 \times \sigma_n$ [2].

2) LFP covariance matrix and circulant approximation: The prior over the LFP signal is a gaussian distribution with zero mean and covariance $\gamma\Sigma_w$. The matrix Σ_w is fixed while the scaling factor γ is estimated by the algorithm and balance the respective contributions of the LFP and the background noise to the observed signal y [5]. Given the huge size of Σ_w ($T \times T$, with T the number of samples of the signal of the order of several thousands), we shall avoid its manipulation (matrix product, inversion, trace calculus), and we follow [5] by constraining Σ_w to be circulant, benefiting of interesting properties of Toeplitz matrices [9]. Under circulant approximation, inversion and product involving Toeplitz matrices can be carried out efficiently using the FFT of the sequence that defines the circulant. This FFT is known to be the expected Power Spectral Density (PSD) of the LFP, denoted g, and has to be estimated from the observed data y. There are different possibilities for the estimation of g. In this paper, an autoregressive power spectral density estimate (Yule-Walker method) is used, with a relatively low order (8 in this paper). This way, we preserve the main characteristics of the LFP spectrum (the LFP being the main contributor to y) while cutting out the contribution of the higher frequencies due to the spiking activities. Given an estimate of the PSD g, equations (12), (13), (18) and (19) can be efficiently computed using the Fast Fourier Transform.

3) Estimation of the number of clusters (single units) and cluster parameters initialization: The number K of spiking neurons (SU) has to be pre-determined. Using a Gaussian Mixture Model (GMM) and Expectation Minimization (EM) with the high-passed filtered spike waveforms as an input, we determine the number of SU based on the MDL criterion [10]. The true number of units are rather well estimated for lower values (up to about ten SU), and tend to provide a sub-estimation when higher number of neurons are to be identified. This approach do not benefit from the LFP/spike separation provided by our algorithm. Starting from the (sub-)estimate given by the GMM-EM algorithm, we iteratively increase the number of clusters K and we compute the MDL criterion after convergence of our algorithm by considering the marginal joint log-likelihood of $(\hat{s}_n, Z)$:

$$MDL_{VB}(K) = -\sum_{n=1}^{N} log(\sum_{k=1}^{K} p(\hat{s}_n|z_n = \lambda_k)\pi_k) + 0.5 * (K * L - 1) * log(N * T_s) \quad (23)$$

with $L = K(1 + T_s + T_s(T_s - 1)/2) - 1$ the number of continuously valued real numbers required to specify the cluster parameters $\{\pi_k, \mu_k, \sigma_k^2\}$. We stop the iteration when $MDL_{VB}(K+1) > MDL_{VB}(K)$.

The spike shapes are initialized based on a naive approach: the peak values of each spike waveform in s_n^r are ordered and divided in K slices with even number of spikes N_s/K. Within each of these K bins about 10 spike shapes are randomly selected, and their average are computed and affected to the initial mean of the clusters. The initial variance of the cluster are taken equal as the tenth of the standard deviation of the original spike waveforms s_n^r.

IV. Results

We evaluate our algorithm on the data set from Camunas et al. [6], which provide realistic simulation of micro recordings containing units (SU) as well as multi-units (MU) originating from the surrounding neuronal population. In these simulations, the presence of 2 up to 20 units are simulated, and for each number of unit 5 signals of 10 minutes length (24Khz sampling rate) are available (https://www135.lamp.le.ac.uk/hgr3/). This provides a common ground-truth of 95 signals for comparing various spike sorting approaches [7], [11]. After the detection procedure, the number of false positives (artifacts) ranges from 1.2% to 3.4% of the total number of detected spikes (respectively from 5 to 123 false detections). On average over the 95 signals, only 75.7% of all spikes (MU+SU) are detected (std: 7.4%), while 96.5% of the SU spikes are detected (std: 1.9%).

The performance of our algorithm is compared with these of the GMM-EM algorithm applied on band-passed time samples (used to initialize the proposed approach, see subsection III-B.3). Both for GMM and for the proposed ap-

proach, we consider a time window of $1ms$ around the maximum peak of the detected spikes ($0.5ms$ before and after). Our method can be seen as a combination of a GMM clustering algorithm with a signal separation between the LFP and the spike waveforms, thus comparing it with GMM alone evaluate the contribution of separating the spikes from the LFP and of considering low frequency features for classification purpose. We also provide comparison with Combinato [7], which proves to be more accurate than the approach presented in [11], [2].

The method is evaluated using the criteria from [7], [11]. A cluster is counted as a hit if it contains more than 50% of a given simulated SU, and if at least 50% of the spikes originating from this unit are classified in this cluster. If a cluster does not match any unit and is not formed of at least 50% of MU spikes, it is counted as a false positive. We provide p-value of a Wilcoxon signed-rank test to evaluate the significance of our results.

Our method significantly outperforms both the GMM approach and Combinato, providing 78.1% (std: 10.7%) hits, while GMM provides 70.6% (std: 8.2%) hits (p=0.0009) and Combinato achieves 71.5% (std: 13.8%, p=0.001). Note that Combinato performs slighly better than GMM, however not significantly (p=0.97). Considering only the data with high number of units (above 8), our approach achieves 70.5% (std: 9.2%) hits, GMM reaches 66.2% (std: 5.5%) hits (p=0.0034), and 64.5% (std: 8.2%) for Combinato (p=0.0007). The difference between GMM and Combinato is again not significant (p=0.53). Figure 1, the average number of hits for each number of simulated units in the signal is given. Our approach proves to be consistently better than Combinato or GMM, the difference in performance being increased with the number of simulated spikes. We also evaluate the method in term of false positives (dashed lines, not provided for Combinato). The average number of false positives vary in the same ranges (from 0 to 3) and, as expected, tends to increase with the number of simulated units. No significant difference are observed between GMM and the proposed approach (p=0.87). Note that the presence of MU generates 1.67 (std: 0.76) additional clusters on average (*i.e.*, clusters with at least 50% of MU spikes). For these MU clusters as well as for the false positives, additional criteria such as those described in [12] should be used to discard them.

V. Conclusion

We propose a method for both extracting and classifying spiking waveforms originating from several units. We demonstrate that the proper extraction of low frequency features brings more accuracy to the classification task. The gain in performance is about 8% on average over all number of simulated units. The method also proves to be competitive with well-established methods of the literature. The version of the approach presented in this paper considers temporal features for the classification step, while it remains true that more discriminant features could be extracted, *e.g.* using principal component analysis or wavelet transform. Another perspective not considered in this paper is that a despiked version of the signal is provided by the method through the variable w. It might then provide a refined version of the STA as proposed by Zanos *et al.* [5] for further spike/LFP relationship analysis [13].

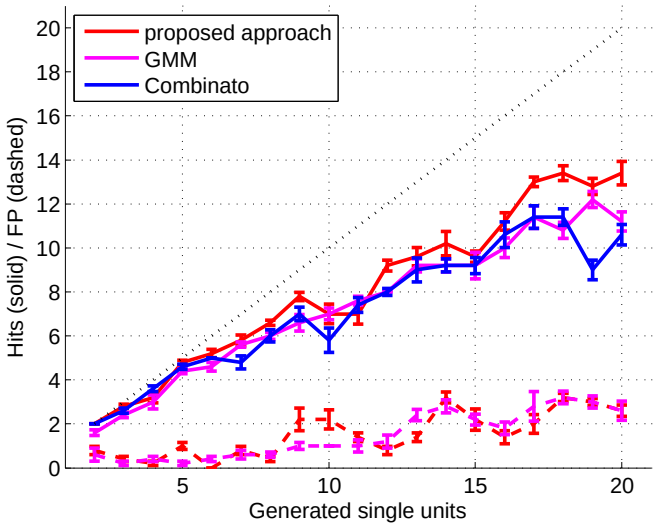

Fig. 1. Average of hits (solid lines) and FP (dashed lines) vs number of simulated units.

References

[1] H. G. Rey, C. Pedreira, and R. Q. Quiroga, "Past, present and future of spike sorting techniques," *Brain research bulletin*, vol. 119, pp. 106–117, 2015.

[2] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural computation*, vol. 16, no. 8, pp. 1661–1687, 2004.

[3] F. Wood and M. J. Black, "A nonparametric bayesian alternative to spike sorting," *Journal of neuroscience methods*, vol. 173, no. 1, pp. 1–12, 2008.

[4] T. Takekawa, Y. Isomura, and T. Fukai, "Accurate spike sorting for multi-unit recordings," *European Journal of Neuroscience*, vol. 31, no. 2, pp. 263–272, 2010.

[5] T. P. Zanos, P. J. Mineault, and C. C. Pack, "Removal of spurious correlations between spikes and local field potentials," *Journal of neurophysiology*, vol. 105, no. 1, pp. 474–486, 2010.

[6] L. A. Camuñas-Mesa and R. Q. Quiroga, "A detailed and fast model of extracellular recordings," *Neural computation*, vol. 25, no. 5, pp. 1191–1212, 2013.

[7] J. Niediek, J. Boström, C. E. Elger, and F. Mormann, "Reliable analysis of single-unit recordings from the human brain under noisy conditions: tracking neurons over hours," *PloS one*, vol. 11, no. 12, p. e0166598, 2016.

[8] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[9] R. M. Gray *et al.*, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[10] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *The Annals of statistics*, pp. 416–431, 1983.

[11] C. Pedreira, J. Martinez, M. J. Ison, and R. Q. Quiroga, "How many neurons can we see with current spike sorting algorithms?" *Journal of neuroscience methods*, vol. 211, no. 1, pp. 58–65, 2012.

[12] D. N. Hill, S. B. Mehta, and D. Kleinfeld, "Quality metrics to accompany spike sorting of extracellular signals," *Journal of Neuroscience*, vol. 31, no. 24, pp. 8699–8705, 2011.

[13] G. T. Einevoll, C. Kayser, N. K. Logothetis, and S. Panzeri, "Modelling and analysis of local field potentials for studying the function of cortical circuits," *Nature Reviews Neuroscience*, vol. 14, no. 11, p. 770, 2013.