

ESTIMATING THE NUMBER OF NEURONS IN MULTI-NEURONAL SPIKE TRAINS¹

BY MENGXIN LI AND WEI-LIEM LOH

National University of Singapore and National University of Singapore

Dedicated to Charles M. Stein on his ninetieth birthday

A common way of studying the relationship between neural activity and behavior is through the analysis of neuronal spike trains that are recorded using one or more electrodes implanted in the brain. Each spike train typically contains spikes generated by multiple neurons. A natural question that arises is “what is the number of neurons ν generating the spike train?” This article proposes a method-of-moments technique for estimating ν . This technique estimates the noise nonparametrically using data from the silent region of the spike train and it applies to isolated spikes with a possibly small, but nonnegligible, presence of overlapping spikes. Conditions are established in which the resulting estimator for ν is shown to be strongly consistent. To gauge its finite sample performance, the technique is applied to simulated spike trains as well as to actual neuronal spike train data.

1. Introduction. In the field of neuroscience, it is generally acknowledged that neurons are the basic units of information processing in the brain. They play this role by generating highly peaked electric action potentials or, more simply, spikes [cf. Brillinger (1988), Dayan and Abbott (2001)]. A sequence of such spikes over time is called a spike train. A typical method of recording spike trains is by inserting electrodes into the brain. In the analysis of a spike train, Brown, Kass and Mitra (2004) note three goals: (i) identify each spike as “signal” (versus pure noise), (ii) determine the number of neurons being recorded, and (iii) assign each spike to the neuron(s) that produced it. (i), (ii) and (iii) are collectively termed *spike sorting* in the neuroscience literature, and, as remarked by Brown, Kass and Mitra (2004), are mandatory for all multi-neuronal spike train data analyses. The accuracy of spike sorting critically affects the accuracy of all subsequent analyses. For spike sorting to be well defined, it is generally assumed that each neuron generates a spike with a characteristic voltage waveshape (apart from noise) and that distinct neurons have distinct spike waveshapes. For example, Figure 3 in Section 4 presents the different spike waveshapes of 5 neurons that were estimated from real data.

Received October 2009; revised June 2010.

¹Supported in part by National University of Singapore Grant R-155-000-094-112.

Key words and phrases. Consistency, eigenvalue, isolated spike, method-of-moments, mixture distribution, neuronal spike train, overlapping spike, spike sorting, trigonometric moment matrix.

This article assumes that (i) has been achieved, that is, each spike in the spike train has been identified. The objective here is to estimate the number of neurons that produced the spike train using the data from (i). This problem has the following feature that makes it a nonstandard clustering problem. Most often, a spike is produced by one (and only one) neuron, as no other neuron spikes around that time. Such a spike is called an *isolated spike*. On the other hand, there may be instances where two or more neurons spike in synchrony, that is, at almost the same time [cf. Lewicki (1998), page R68]. Then the resulting spike is an additive superposition of the spikes generated by this group of neurons. This spike is called an *overlapping spike*. Consequently, if considered as a clustering problem, the number of clusters will not in general be equal to the number of neurons producing the spike train.

For definiteness we shall assume that there are ν neurons, labeled 1 to ν , generating the spike train and that n spikes, say, S_1, \dots, S_n , are detected and recorded in the spike train. Here the S_i 's are aligned according to their peaks and each $S_i \in \mathbb{R}^d$ for some $d \in \mathbb{Z}^+$. The d components of S_i are the measurements of the voltage values of the spike on a regular grid of time-points around the peak (or maximum) of the spike. Writing $S_i = (S_{i,1}, \dots, S_{i,d})' \in \mathbb{R}^d$, we assume that

$$(1) \quad S_{i,j} = \Theta_{i,j} + \eta_{i,j} \quad \forall i = 1, \dots, n, j = 1, \dots, d,$$

where $\eta_{i,j}, i = 1, \dots, n, j = 1, \dots, d$, are i.i.d. noise random variables with mean 0, variance σ^2 and $\Theta_i = (\Theta_{i,1}, \dots, \Theta_{i,d})' \in \mathbb{R}^d, i = 1, \dots, n$, are i.i.d. random vectors. The Θ_i 's and $\eta_{i,j}$'s are assumed to be all independent. Θ_i denotes the denoised spike shape of S_i and is random because the denoised spike shape is a function of the particular neuron(s) generating it and time lag in the spiking times if there are ≥ 2 neurons. Let $\alpha \in \mathbb{R}^d$ be a constant vector such that $\alpha' \alpha = 1$. In Sections 4 and 5, α is taken to be the first principal component of S_1, \dots, S_n and $0.01n$ vector 0's. However, other choices of α are possible too. For each spike S_i , define

$$(2) \quad X_i = \alpha' S_i \quad \forall i = 1, \dots, n,$$

which is the projection of S_i onto α . It follows from (1) that X_1, \dots, X_n are i.i.d. random variables. As observed by Ventura (2009), either implicitly or explicitly, almost all spike sorting methods assume that X_i has a mixture distribution with probability density function of the form

$$(3) \quad \sum_{q=1}^{\nu} \sum_{1 \leq j_1 < \dots < j_q \leq \nu} \pi_{j_1, \dots, j_q} h_{j_1, \dots, j_q}(x) \quad \forall x \in \mathbb{R},$$

where π_{j_1, \dots, j_q} is the probability that S_i is generated by (and only by) neurons j_1, \dots, j_q . The h_{j_1, \dots, j_q} 's are usually assumed to be Gaussian densities [cf. Lewicki (1994, 1998)], even though t densities have also been proposed [cf. Shoham, Fellows and Normann (2003)].

In this article we shall not assume that $h_j(\cdot)$, $j = 1, \dots, \nu$, are Gaussian or t densities but only that $h_j(\cdot) = f_{\mu_j, \sigma^2}(\cdot)$ belongs to a location family of probability densities [cf. Lehmann (1983), page 154]. It is noted that this location family contains both Gaussian and t densities. Here μ_j and σ^2 denote the mean and variance induced by the density f_{μ_j, σ^2} .

Due to the complex nature of spike overlap, we do not think it is accurate to model each h_{j_1, \dots, j_q} , $2 \leq q \leq \nu$, as a Gaussian or a t density (or, indeed, any other parametric density having only a small number of unknown parameters). For example, $h_{1,2}$ depends on noise as well as on the time lag between the spiking times of neurons 1 and 2. As the time lag is also random and that the phenomenon of neurons spiking in close proximity to each other is still not well understood, we feel it is more appropriate to model $h_{1,2}$ using a nonparametric density (rather than a parametric one). In particular, in this article, we shall assume that (3) is of the form

$$(4) \quad f(x) = \sum_{j=1}^{\nu} \pi_j f_{\mu_j, \sigma^2}(x) + \sum_{q=2}^{\nu} \sum_{1 \leq j_1 < \dots < j_q \leq \nu} \pi_{j_1, \dots, j_q} \int_{\mathbb{R}} f_{\mu, \sigma^2}(x) dG_{j_1, \dots, j_q}(\mu),$$

where $\pi_1 \geq \pi_2 \geq \dots \geq \pi_{\nu} > 0$ and G_{j_1, \dots, j_q} 's are unknown absolutely continuous probability distributions. Consequently, (4) leads to a nonparametric location mixture.

The objective is to estimate ν in (4) using the sample X_1, \dots, X_n and an independent auxiliary sample of i.i.d. observations Y_1, \dots, Y_m obtained from the silent region of the spike train. The silent region is defined to be the sections of the spike train where there are no spikes. Hence, the Y_l 's are defined as

$$(5) \quad Y_l = (\eta_{l,1}^*, \dots, \eta_{l,d}^*)\alpha \quad \forall l = 1, \dots, m,$$

where $\eta_{l,1}^*, \dots, \eta_{l,d}^*$ are noise voltage measurements on a regular grid of d consecutive time-points of the silent region of the spike train. We assume that $\eta_{l,j}^*, \eta_{i,k}$, $1 \leq l \leq m, 1 \leq i \leq n, 1 \leq j, k \leq d$, are i.i.d. random variables with mean 0 and variance σ^2 . In this article a method-of-moments estimator $\hat{\nu}$ for ν is proposed. This estimator has a number of attractive properties. First, this article establishes a reasonably transparent theory justifying/supporting $\hat{\nu}$. In particular, $\hat{\nu}$ is a strongly consistent estimator for ν under mild conditions. Second, the estimator can be computed without first (or concurrently) computing the other unknown quantities in (4). Consequently, it is computationally very fast relative to, say, EM or Markov chain Monte Carlo (MCMC) algorithms.

We would like to add that the above problem of estimating ν can be regarded as robust and (yet) consistent estimation of the number of components of a finite mixture where the latter is subjected to a small but nonnegligible contamination

by a nuisance distribution. While quite a number of papers have been written on estimating the number of components of a finite mixture [cf. Dacunha-Castelle and Gassiat (1997) and references cited therein], we are not aware of any work in the statistics literature that deals with this problem when the mixture is contaminated by another distribution.

On the neuroscience literature side, numerous algorithms for spike sorting have been proposed. A review of spike sorting algorithms and a discussion of their strengths and limitations can be found in Lewicki (1998). In particular, a considerable number of spike sorting algorithms assume that the proportion of overlapping spikes is negligible relative to the proportion of isolated spikes and, hence, the possible presence of overlapping spikes is ignored. With this assumption, many spike sorting algorithms further assume that the number of neurons ν is known and the problem reduces to a standard classification problem. If ν is unknown, other spike sorting algorithms use various EM or MCMC methods for determining ν as well as for assigning spikes to the neurons [cf. Pouzat, Mazor and Laurent (2002), Nguyen, Frank and Brown (2003), Wood and Black (2008)]. However, Brown, Kass and Mitra [(2004), page 456] noted that MCMC techniques have yet to be widely tested in spike sorting.

Spike sorting algorithms that take overlapping spikes into account usually involve significant user input [cf. Mokri and Yen (2008)]. Section 5 of Lewicki (1998) discusses the use of templates, independent component analysis and neural networks to handle overlapping spikes and the limitations of these methods. If ν is known, there are at least two model-based approaches for handling overlapping spikes. The first approach considers a $(\nu + 1)$ -component mixture distribution with the first ν components modeling the spike waveforms of the ν neurons. The $(\nu + 1)$ th component is a uniform density over a suitably large region of the sample space [cf. Sahani (1999), page 95] that serves as an approximate model for the overlapping spikes. The second approach is a trimming method where a number of the largest observations (outliers) are omitted [cf. Gallegos and Ritter (2005), García-Escudero et al. (2008)]. The rationale is that most of the outliers correspond to overlapping spikes and, hence, the remaining observations should be comprised essentially of isolated spikes.

The rest of the article is organized as follows. Section 2 introduces a number of trigonometric moment matrices. Theorem 1 derives some explicit error bounds for their eigenvalues. Motivated by these error bounds, Section 3 proposes a method-of-moments estimator $\hat{\nu}$ for ν . Theorem 2 shows that $\hat{\nu}$ is a strongly consistent estimator for ν under mild conditions. A point of note is that Theorem 2 does not require the proportion of overlapping spikes to be asymptotically negligible as sample size tends to infinity.

Section 4 presents a detailed simulated spike train study that investigates the finite sample accuracy of $\hat{\nu}$. Each spike train is generated by ν neurons where ν varies from 1 to 5. The spike shapes of these neurons are estimated from real data. There are 8 experiments in the study which present a variety of different

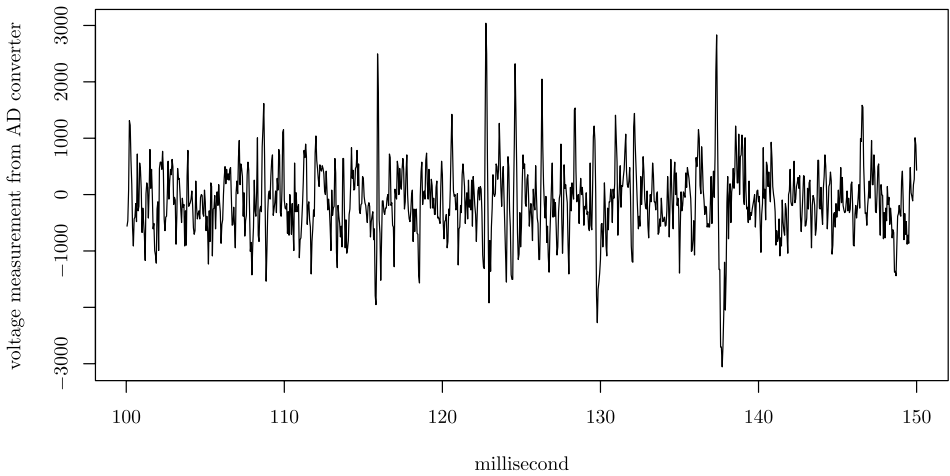


FIG. 1. A portion of Lewicki's actual spike train recording.

spike train situations depending on the proportion of overlapping spikes, the spike detection technique and sample sizes n, m . Table 1 summarizes the 8 experiments as a factorial design. This study finds that $\hat{\nu}$ has very good accuracy with regard to these 8 experiments for moderately large sample sizes such as $n = 1000$ and $m = 2000$. As a comparison, we have also applied the SpikeOMatic software [cf. Pouzat, Mazor and Laurent (2002); Pouzat et al. (2004)] to obtain an alternative estimate $\hat{\nu}_1$ for ν in 2 of these experiments.

Section 5 considers two spike train data sets taken from Lewicki (1994). The first is an actual 40-second spike train recording and the second is a synthesized recording using 6 spike shapes estimated from the first data set. Figure 1 presents a portion of the actual spike train recording. Lewicki (1994) inferred that ν is 6 for the actual recording. Three estimators are used to estimate ν from each spike train, namely, $\hat{\nu}_2$ by Lewicki's spike sorting algorithm [cf. Lewicki (1994)], $\hat{\nu}_1$ by SpikeOMatic software [cf. Pouzat, Mazor and Laurent (2002)] and $\hat{\nu}$. For the synthesized recording, we obtain $\hat{\nu}_1 = \hat{\nu}_2 = 5$ while $\hat{\nu} = 4$ or 5 depending on whether the threshold of $\hat{\nu}$ is set to 1.0 or 0.8, respectively. On the other hand, for the actual spike train recording, we obtain $\hat{\nu}_1 = 12$, $\hat{\nu}_2 = 9$ while $\hat{\nu} = 4$ or 5 depending on whether the threshold of $\hat{\nu}$ is set to 1.0 or 0.8, respectively. Thus, relative to $\hat{\nu}_1$ and $\hat{\nu}_2$, $\hat{\nu}$ is a more stable estimate with respect to these 2 data sets.

The article ends with some concluding remarks in Section 6. The symbols \mathbb{R}, \mathbb{C} denote the set of real numbers, complex numbers respectively, and all proofs are deferred to the Appendix.

2. Some trigonometric moment matrices. This section constructs a number of trigonometric moment matrices and derives explicit error bounds for their eigenvalues. Following the notation in (4), let θ denote a random variable with

cumulative distribution function

$$(6) \quad F_\theta(\mu) = \sum_{j=1}^v \pi_j \mathcal{I}\{\mu \geq \mu_j\} + \sum_{q=2}^v \sum_{1 \leq j_1 < \dots < j_q \leq v} \pi_{j_1, \dots, j_q} G_{j_1, \dots, j_q}(\mu) \quad \forall \mu \in \mathbb{R},$$

where $\pi_1 \geq \pi_2 \geq \dots \geq \pi_v > 0$ and $\mathcal{I}\{\cdot\}$ denotes the indicator function. For simplicity, let π_{cont} be the proportion of overlapping spikes and $\pi_{\text{cont}} F_{\text{cont}}$ be the continuous component of F_θ . Then

$$\begin{aligned} \pi_{\text{cont}} &= \sum_{q=2}^v \sum_{1 \leq j_1 < \dots < j_q \leq v} \pi_{j_1, \dots, j_q} = 1 - \sum_{j=1}^v \pi_j, \\ \pi_{\text{cont}} F_{\text{cont}}(\mu) &= \sum_{q=2}^v \sum_{1 \leq j_1 < \dots < j_q \leq v} \pi_{j_1, \dots, j_q} G_{j_1, \dots, j_q}(\mu) \quad \forall \mu \in \mathbb{R}. \end{aligned}$$

Let $p \geq v$ be an integer. Motivated by the ideas of Lindsay (1989a), (1989b) on moment matrices and finite mixtures, we define a function $T_p : \mathbb{R} \rightarrow \mathbb{C}^{(p+1) \times (p+1)}$ by

$$\begin{aligned} T_p(x) &= \begin{pmatrix} 1 & e^{ix} & e^{i2x} & \dots & e^{ipx} \\ e^{-ix} & 1 & e^{ix} & \dots & e^{i(p-1)x} \\ e^{-i2x} & e^{-ix} & 1 & \dots & e^{i(p-2)x} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-ipx} & e^{-i(p-1)x} & e^{-i(p-2)x} & \dots & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ e^{-ix} \\ e^{-i2x} \\ \vdots \\ e^{-ipx} \end{pmatrix} (1, e^{ix}, e^{i2x}, \dots, e^{ipx}), \end{aligned}$$

where $\mathbf{i} = \sqrt{-1}$. We further define matrices corresponding to the discrete component and the continuous component of F_θ by

$$(7) \quad \begin{aligned} M_{p, \text{disc}} &= \sum_{i=1}^v \pi_i T_p(\mu_i) = \begin{pmatrix} 1 & \dots & 1 \\ e^{-i\mu_1} & \dots & e^{-i\mu_v} \\ e^{-i2\mu_1} & \dots & e^{-i2\mu_v} \\ \vdots & \ddots & \vdots \\ e^{-ip\mu_1} & \dots & e^{-ip\mu_v} \end{pmatrix} \begin{pmatrix} \pi_1 & 0 & \dots & 0 \\ 0 & \pi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_v \end{pmatrix} \\ &\times \begin{pmatrix} 1 & e^{i\mu_1} & e^{i2\mu_1} & \dots & e^{ip\mu_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{i\mu_v} & e^{i2\mu_v} & \dots & e^{ip\mu_v} \end{pmatrix}, \end{aligned}$$

$$M_{p,\text{cont}} = \int_{\mathbb{R}} T_p(\mu) f_{\text{cont}}(\mu) d(\mu) = \int_0^{2\pi} T_p(\mu) \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j) d(\mu),$$

where f_{cont} is the probability density function of the distribution F_{cont} . Finally, we define

$$M_p = E[T_p(\theta)] = M_{p,\text{disc}} + \pi_{\text{cont}} M_{p,\text{cont}}.$$

Let $\lambda_i(A)$ denote the i th largest eigenvalue of A where A is an arbitrary $(p + 1) \times (p + 1)$ Hermitian matrix. Hence, $\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_{p+1}(A)$.

THEOREM 1. *With the above notation, suppose that μ_1, \dots, μ_ν are all distinct, $0 \leq \mu_1, \dots, \mu_\nu < 2\pi$. Then for $i = 1, \dots, \nu$, we have*

$$\begin{aligned} & (p + 1)\pi_i + 2\pi \pi_{\text{cont}} \left\{ \min_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j) \right\} \\ & - \sqrt{2 \sum_{1 \leq j < k \leq \nu} \pi_j \pi_k \left| \frac{1 - e^{i(p+1)(\mu_j - \mu_k)}}{1 - e^{i(\mu_j - \mu_k)}} \right|^2} \\ (8) \quad & \leq \lambda_i(M_p) \\ & \leq (p + 1)\pi_i + 2\pi \pi_{\text{cont}} \left\{ \max_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j) \right\} \\ & + \sqrt{2 \sum_{1 \leq j < k \leq \nu} \pi_j \pi_k \left| \frac{1 - e^{i(p+1)(\mu_j - \mu_k)}}{1 - e^{i(\mu_j - \mu_k)}} \right|^2}. \end{aligned}$$

Also, for $i = \nu + 1, \dots, p + 1$, we have

$$\begin{aligned} & 2\pi \pi_{\text{cont}} \min_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j) \\ (9) \quad & \leq \lambda_i(M_p) \leq 2\pi \pi_{\text{cont}} \max_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j). \end{aligned}$$

The following is an immediate corollary of Theorem 1.

COROLLARY 1. *Suppose the conditions of Theorem 1 are satisfied. Then for any constant $\gamma > 0$, there exists a positive integer p_γ such that*

$$\lambda_\nu(M_p) > \gamma \sqrt{p + 1} > \lambda_{\nu+1}(M_p) \quad \forall p \geq p_\gamma.$$

Also, $\lambda_{\nu+1}(M_p)$ is bounded uniformly in p . Finally, $\lambda_i(M_p) \sim (p + 1)\pi_i, \forall 1 \leq i \leq \nu$, as $p \rightarrow \infty$.

Corollary 1 gives, at least in principle, a way for estimating ν by estimating the eigenvalues of M_p .

3. A method-of-moments estimator for ν . Let $\eta_{i,j}$, $i = 1, \dots, n$, $j = 1, \dots, d$, and Θ_i , $i = 1, \dots, n$, be as in (1) and $X_i = \alpha' S_i$, $i = 1, \dots, n$, be as in (2). Then X_1, \dots, X_n is an i.i.d. sequence of observations from the mixture distribution given by (4). We observe that

$$X_i = \theta_i + \tilde{Y}_i \quad \forall i = 1, \dots, n,$$

where $\theta_i = \alpha' \Theta_i$ and $\tilde{Y}_i = (\eta_{i,1}, \dots, \eta_{i,d})' \alpha$ are independent random variables having cumulative distribution function F_θ , given by (6), and probability density function f_{0,σ^2} , respectively. θ_i can be regarded as the signal and \tilde{Y}_i the zero-mean noise. Here we assume that $E(e^{-ik\tilde{Y}_1}) \neq 0$ for all $k \in \{1, \dots, p\}$. This is a very weak assumption and is satisfied by, for example, mean-centered normal, t and Gamma distributions. Because θ_1 and \tilde{Y}_1 are independent, we have

$$E(e^{-ik\theta_1}) = E(e^{-ikX_1})[E(e^{-ik\tilde{Y}_1})]^{-1}.$$

Let Y_1, \dots, Y_m be as in (5). We observe that the Y_l 's are obtained from the voltage measurements at different time-points of the silent region of the spike train. Then Y_1, \dots, Y_m are i.i.d. with density f_{0,σ^2} and are also independent of X_1, \dots, X_n . As Y_1 has the same distribution as \tilde{Y}_1 , we have

$$(10) \quad E(e^{-ikY_1}) \neq 0 \quad \forall k \in \{1, \dots, p\}.$$

Since $M_p = E[T_p(\theta)]$, we shall estimate M_p using its sample analog, that is, the $(p + 1) \times (p + 1)$ matrix \hat{M}_p whose (j, k) th element is given by

$$(11) \quad (\hat{M}_p)_{j,k} = \frac{n^{-1} \sum_{i=1}^n e^{-i(j-k)X_i}}{m^{-1} \sum_{l=1}^m e^{-i(j-k)Y_l}}.$$

\hat{M}_p is a Hermitian matrix and, hence, its eigenvalues are real numbers. The method-of-moments estimator $\hat{\nu}$ for ν is as follows. Let γ , p_γ and $p \geq p_\gamma$ be as in Corollary 1. Define

$$(12) \quad \hat{\nu} = \#\{i : 1 \leq i \leq p + 1, \lambda_i(\hat{M}_p) > \gamma_{\text{threshold}}\},$$

where $\gamma_{\text{threshold}} = \gamma\sqrt{p + 1}$ and $\#\{\cdot\}$ denotes set cardinality. We call $\gamma_{\text{threshold}}$ the threshold parameter of $\hat{\nu}$.

THEOREM 2. *Let $\hat{\nu}$ be as in (12) with $p \geq p_\gamma$. Suppose (10) and the conditions of Theorem 1 hold. Then $\hat{\nu} \rightarrow \nu$ almost surely as $\min(m, n) \rightarrow \infty$.*

For $\hat{\nu}$ to perform well, it is necessary to obtain a good choice of the threshold parameter $\gamma_{\text{threshold}}$. The usual methods, such as cross-validation, for determining the value of the threshold parameter do not seem to work here. Instead we shall compute explicit error bounds for $\lambda_i(\hat{M}_p)$, $i = 1, \dots, p + 1$, below. These error bounds shall serve as guidelines for setting the value of $\gamma_{\text{threshold}}$.

LEMMA 1. *Let $\varepsilon > 0$ be a constant,*

$$\Omega_{j,\varepsilon} = \left\{ \left| \frac{m^{-1} \sum_{l=1}^m e^{-ijY_l}}{E(e^{-ijY_1})} - 1 \right| > \varepsilon \right\} \quad \forall j = 1, \dots, p,$$

and $\Omega_\varepsilon = \bigcup_{j=1}^p \Omega_{j,\varepsilon}$. Then

$$P(\Omega_\varepsilon) \leq \sum_{j=1}^p \min \left\{ \frac{6}{m^2[\varepsilon|E(e^{-ijY_1})|]^4} \left(1 + O\left(\frac{1}{m}\right) \right), \right. \\ \left. \frac{42}{m^3[\varepsilon|E(e^{-ijY_1})|]^6} \left(1 + O\left(\frac{1}{m}\right) \right) \right\}.$$

It is interesting to note that for the parameter values and sample sizes that we are concerned with in this article, the inequality of Lemma 1 gives a smaller upper bound than those obtained via the Hoeffding or Bernstein exponential-type inequalities.

THEOREM 3. *Let Ω_ε be as in Lemma 1, Ω_ε^c be its complement and $E^{\Omega_\varepsilon^c}$ denote the conditional expectation given Ω_ε^c . Then with the assumptions of Theorem 1,*

$$(13) \quad \sqrt{\frac{E^{\Omega_\varepsilon^c} \sum_{i=1}^{p+1} [\lambda_i(\hat{M}_p) - \lambda_i(M_p)]^2}{p+1}} \\ \leq \sqrt{\frac{2}{n(1-\varepsilon)^2} \sum_{j=1}^p \frac{p-j+1}{(p+1)|\psi_Z(\sigma j)|^2} + \frac{p\varepsilon^2}{(1-\varepsilon)^2}},$$

where $Z = Y_1/\sigma$, $\psi_Z(t) = Ee^{-itZ}$ for all $t \in \mathbb{R}$ and, hence, $Ee^{-ijY_1} = \psi_Z(\sigma j)$.

We remark that the upper bounds of the inequalities of Lemma 1 and Theorem 3, though relatively simple, are conservative in that the quantities on the left-hand side are substantially smaller than those on the right-hand side. Nonetheless, we shall end this section with an example which computes the upper bounds in Lemma 1 and Theorem 3 explicitly. Suppose $\varepsilon = 0.05$, $p = 20$, $\sigma = 0.1$, $Y_1 \sim N(0, \sigma^2)$, $n = 1000$ and $m = n^2$. Then $P(\Omega_\varepsilon) \leq 0.01$ and

$$E^{\Omega_\varepsilon^c} \frac{\sum_{i=1}^{p+1} [\lambda_i(\hat{M}_p) - \lambda_i(M_p)]^2}{p+1} \leq 0.12.$$

4. Simulated spike train study. In this section we shall study the finite sample performance of the method-of-moments estimator \hat{v} given by (12) via simulated spike trains.

4.1. *Spike train generation.* For realism, we use spike shapes estimated from real data to generate the spike train. The spike shapes are obtained by applying Pouzat's software SpikeOMatic which is available at www.biomedicale.univ-paris5.fr/SpikeOMatic.html to a tetrode data set recorded from the locust (*Schistocerca americana*) antennal lobe. This tetrode data is distributed with the SpikeOMatic software at www.biomedicale.univ-paris5.fr/SpikeOMatic/Data.html. It is a 20-second recording sampled at 15 kHz from a tetrode filtered between 300 Hz and 5 kHz. Pouzat, Mazor and Laurent (2002) and Pouzat et al. (2004) are two papers behind the software SpikeOMatic.

SpikeOMatic is applied to the locust data and we selected the 6 spike shapes with the largest numbers of spikes. These (four-channel) spike shapes are shown in Figure 2. As the two spike shapes with the smallest first channel positive peak heights have significant difference only in the fourth channel, we will delete one of them in order to generate single channel data in which distinct spike shapes are different enough to be distinguished. The resulting 5 spike shapes are shown in Figure 3 with the noise standard deviation $\sigma = 1$.

Once the spike shapes are obtained, a Poisson process is simulated to select the time of spike events. Given that a spike occurred, a random variable with categorical distribution is generated indicating which neuron has spiked. The categorical distribution is assumed to have equal probabilities for each of ν possible spike shapes where ν is the number of neurons (i.e., $\pi_1 = \dots = \pi_\nu \leq 1/\nu$). Here ν ranges from 1 to 5. The π_i 's can be set to different values but they should not be too close to zero. This is necessary because if one of the neurons has infinitesimally small

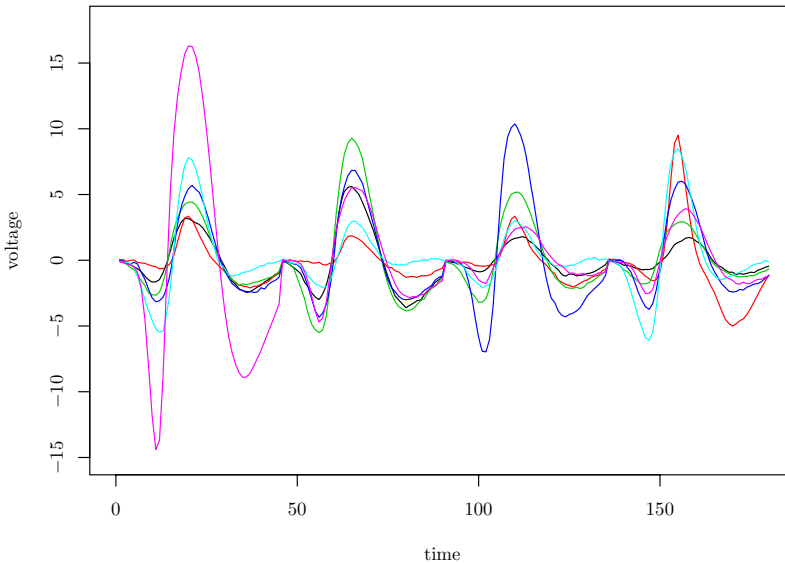


FIG. 2. Six four-channel spike shapes from the locust data.

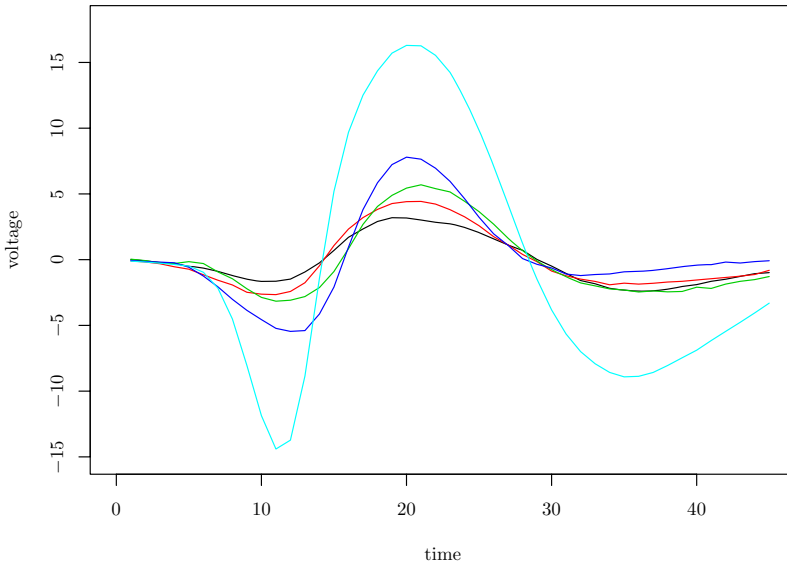


FIG. 3. Five one-channel spike shapes selected with $\sigma = 1$.

probability of spiking, then it would appear that no algorithm can estimate ν well with a finite sample.

As noted previously, Pouzat’s locust data is a 20-second recording sampled at 15 kHz. Thus, the recording has 300,000 time-points on a regular grid. There are about 1000 spikes detected. Hence, the collective firing rate is about $1000/300,000 = 1/300$. In our simulations, we choose a similar collective firing rate of $1/400$, as this gives about 10% overlapping spikes.

SpikeOMatic rescales the data such that the noise standard deviation $\sigma = 1$. In the generation of a spike train, independent standard Gaussian noise, or t_5 distributed noise (with degree of freedom 5) multiplied by $\sqrt{3/5}$, is added to the “signal.” Multiplication by $\sqrt{3/5}$ is needed since t_5 distribution does not have standard deviation 1. Here “signal” refers to either the spike shape or 0 (in the case of pure noise).

4.2. Spike detection. As mentioned in the Introduction, one of the tasks of spike sorting is to identify each spike in the spike train. This is known as spike detection. The function `find.spikes.with.template` provided by SpikeOMatic is used to detect spikes from spike train recordings. In our study, the detection threshold is set to 2.25σ (which is one of the recommended values of SpikeOMatic) where σ is the noise standard deviation. This is because the spike shape with the smallest positive peak height (in Figure 3) is close to 3σ and we do not want to consistently miss detecting spikes with this spike shape. The function `find.spikes.with.template` uses template to detect spikes and, hence, there should not be too many false positives

with detection threshold $= 2.25\sigma$ if the noise is Gaussian. However, this threshold would present problems if the noise has a heavy-tailed distribution such as the t_5 distribution (cf. Experiment 6 of Section 4.4). For more details of spike detection via SpikeOMatic, we refer the reader to the documentation for the software.

4.3. *Setting the tuning parameters of \hat{v} .* We observe from the definition of \hat{v} in (12) that \hat{v} is not scale invariant and that \hat{v} has three tuning parameters, namely, σ , p and $\gamma_{\text{threshold}}$, to be determined in order to use \hat{v} . First we observe that it is usually the case that the silent region forms a sizeable part of a spike train. Since data from the silent region can be used to estimate σ , we shall, without loss of generality, assume in this subsection that σ is known.

The function `make.sweeps` of SpikeOMatic extracts the spikes and the pure noise observations and automatically rescales them such that the noise standard deviation is 1. After extracting the spikes, we shall further rescale them such that the noise standard deviation $\sigma = 0.1$. The rationale for this will be explained later. Next we set $\gamma_{\text{threshold}} = 1$ with the implicit assumption that $(p + 1)\pi_\nu > \gamma_{\text{threshold}}$. We observe from (8) that the latter is a rough proxy for $\lambda_\nu(\hat{M}_p) > \gamma_{\text{threshold}}$. This implicit assumption is indeed satisfied for all the 8 experiments. (If this does not hold, we would have to set a smaller value for $\gamma_{\text{threshold}}$ with a corresponding increase in false positives.)

Let $S_1, \dots, S_n \in \mathbb{R}^d$ denote the extracted spikes. The (normalized) first principal component $\alpha \in \mathbb{R}^d$ of $S_1, \dots, S_n, 0, \dots, 0$ is computed. Here $0 \in \mathbb{R}^d$ and the number of 0's used to compute α is taken to be $0.01n$ (to the nearest integer). The 0's are needed in the computation of α because if not, α is not well defined in the case of $\nu = 1$ and i.i.d. Gaussian noise. Next define $X_j = \alpha' S_j$ as in the [Introduction](#). The rationale for projecting the spikes onto the first principal component is the hope that this direction will best separate the spike shapes from one another as well as from pure noise (which is represented by 0's).

Now we argue that σ should be neither too small nor too large after rescaling. If σ is large, then $\psi_Z(\sigma j)$ will be small. [For example, if the noise is Gaussian, we have $\psi_Z(\sigma j) = e^{-j^2\sigma^2/2}$.] Thus, the bound as on the right-hand side of (13) is large. This indicates that the estimation error of the eigenvalues will be large, resulting in poor performance of \hat{v} .

On the other hand, if σ is scaled too small, the distance between some pair of μ_j 's would likely be small, that is, for some $j \neq k$, $|1 - e^{i(\mu_j - \mu_k)}|$ will be small. This implies that the lower bound in (8) can be less than the threshold parameter $\gamma_{\text{threshold}} = 1$. This again results in poor performance of \hat{v} . Consequently, we would like the following condition to be satisfied:

$$\text{CONDITION (I): } \sqrt{\frac{2}{0.95^2 n} \sum_{j=1}^p \frac{p-j+1}{(p+1)|m^{-1} \sum_{l=1}^m e^{-ipY_l}|^2} + \frac{0.05^2 p}{0.95^2}} \leq \frac{1}{3}.$$

Since m is large, we observe that $|E(e^{-ipY_1})| \approx |m^{-1} \sum_{l=1}^m e^{-ipY_l}|$. Consequently, the left-hand side of the inequality in Condition (I) is an approximation for the right-hand side of (13) with $\varepsilon = 0.05$ and can be used as an approximate upper bound for the standard error of the eigenvalue estimates in (13). Thus, if Condition (I) holds, $\gamma_{\text{threshold}} = 1$ is very likely to exceed 3 times this standard error and that, for sufficiently small π_{cont} , $\lambda_{\nu+1}(\hat{M}_p) \leq 1$. Finally we choose $p = p_{\text{max}}$ where p_{max} is the largest value of p satisfying Condition (I). Numerical experiments on the accuracy of $\hat{\nu}$ with these values of $\sigma = 0.1$, $p = p_{\text{max}}$, and $\gamma_{\text{threshold}} = 1$ will be reported in Section 4.4.

4.4. Numerical experiments. In this subsection we shall study the performance of the estimator $\hat{\nu}$ of Section 4.3 via 8 simulated spike train experiments. Each experiment is divided into 5 scenarios depending on the number of neurons generating the spike train (i.e., $\nu = 1, \dots, 5$). A total of at most 5 neurons are considered. The spike shapes of these neurons are given in Figure 3. As there are many ways of selecting ν neurons from 5 neurons if $\nu < 5$, we shall choose the ν neurons in a “least favorable” manner for estimating ν : if $\nu = 1$, we take the neuron with the smallest peak height in Figure 3 to be the one generating the spike train; if $\nu = 2$, then we take the 2 neurons with the two smallest peak heights in Figure 3 to be the ones generating the spike train; and so on until we reach $\nu = 5$. In all the experiments the estimator $\hat{\nu}$ is used after rescaling the data so that $\sigma = 0.1$ and setting $p = p_{\text{max}}$ and $\gamma_{\text{threshold}} = 1$. The spike shape vectors each have $d = 45$ components. Table 1 summarizes the 8 experiments as a factorial design.

EXPERIMENT 1. In this experiment the noise is i.i.d. Gaussian with mean 0 and variance $\sigma^2 = 1$. We assume that the spikes are detected from the spike train without error (or, equivalently, there exists an oracle spike detector):

- The proportion of overlapping spikes $\pi_{\text{cont}} \approx 0.1$ (or 10%).

TABLE 1
Factorial experimental design

| Experiment | Noise distribution | Spike detection algorithm | Proportion of overlapping spikes |
|------------|--------------------|---------------------------|----------------------------------|
| 1 | Gaussian | Oracle | 10% |
| 2 | Gaussian | SpikeOMatic | 10% |
| 3 | Gaussian | Oracle | 0 |
| 4 | Gaussian | SpikeOMatic | 0 |
| 5 | Student t_5 | Oracle | 10% |
| 6 | Student t_5 | SpikeOMatic | 10% |
| 7 | Student t_5 | Oracle | 0 |
| 8 | Student t_5 | SpikeOMatic | 0 |

TABLE 2
Approximate values of μ_1, \dots, μ_ν for Experiments 1, 3, 5 and 7

| | | μ_1 | μ_2 | μ_3 | μ_4 | μ_5 |
|--------------|-----------|---------|---------|---------|---------|---------|
| Experiment 1 | $\nu = 1$ | 11.1 | | | | |
| | $\nu = 2$ | 9.7 | 13.7 | | | |
| | $\nu = 3$ | 8.4 | 11.3 | 15.8 | | |
| | $\nu = 4$ | 5.7 | 9.5 | 12.3 | 20.4 | |
| | $\nu = 5$ | 11.4 | 14.3 | 16.9 | 19.5 | 58.9 |
| Experiment 3 | $\nu = 1$ | 11.7 | | | | |
| | $\nu = 2$ | 8.1 | 12.4 | | | |
| | $\nu = 3$ | 9.2 | 12.2 | 16.6 | | |
| | $\nu = 4$ | 5.5 | 9.3 | 12.0 | 20.2 | |
| | $\nu = 5$ | 11.4 | 14.3 | 17.0 | 19.5 | 58.9 |
| Experiment 5 | $\nu = 1$ | 11.2 | | | | |
| | $\nu = 2$ | 10.3 | 14.0 | | | |
| | $\nu = 3$ | 9.9 | 13.0 | 17.2 | | |
| | $\nu = 4$ | 5.5 | 9.2 | 12.0 | 20.1 | |
| | $\nu = 5$ | 11.4 | 14.3 | 16.9 | 19.4 | 58.9 |
| Experiment 7 | $\nu = 1$ | 11.6 | | | | |
| | $\nu = 2$ | 8.5 | 12.7 | | | |
| | $\nu = 3$ | 9.1 | 12.2 | 16.5 | | |
| | $\nu = 4$ | 5.5 | 9.2 | 12.0 | 20.2 | |
| | $\nu = 5$ | 11.4 | 14.3 | 17.0 | 19.6 | 58.9 |

- $n = 1000$ (or 500), that is, there are 1000 (or 500) spikes detected from the spike train.
- $m = 2n$, where m is the number of pure noise observations Y_j 's [as in (5)] obtained from the silent region of the spike train.
- The approximate μ_i 's (i.e., the projections of the spike shape vector onto the first principal component $\alpha \in \mathbb{R}^d$) of these neurons are presented in Table 2.

Table 3 gives the percentage of the time (out of 100 repetitions) that $\hat{\nu} = \nu$, the true number of neurons producing the spike train. In the case $n = 1000, m = 2000$, $\hat{\nu}$ does well, for example, for $\nu = 2$, the percentage of the time $\hat{\nu} = \nu$ is 97%. In the case $n = 500$ and $m = 1000$, $\hat{\nu}$ does well for the scenarios $1 \leq \nu \leq 4$ but does poorly for $\nu = 5$.

REMARK. The Editor raised the following question: “would detection improve if m/n were larger than 2/1?” The nature in which the sample sizes m, n affect the accuracy of $\hat{\nu}$ is not a straightforward one. If n is held fixed and m is allowed to increase, then the accuracy of $\hat{\nu}$ should improve. However, if n is also allowed to increase, then p_{\max} would likewise increase. This implies from (13) and Condition (I) that $\psi_Z(\sigma p_{\max})$ will be harder to estimate because the latter is further

TABLE 3
Frequency (%) of $\hat{\nu} = \nu$ with standard error in parentheses for Experiments 1–8

| | n | m | $\nu = 1$ | $\nu = 2$ | $\nu = 3$ | $\nu = 4$ | $\nu = 5$ |
|--------------|------|------|-----------|-----------|-----------|-----------|-----------|
| Experiment 1 | 1000 | 2000 | 81 (3.9) | 97 (1.7) | 100 (0.0) | 100 (0.0) | 100 (0.0) |
| | 500 | 1000 | 97 (1.7) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 10 (3.0) |
| Experiment 2 | 1000 | 2000 | 51 (5.0) | 79 (4.1) | 98 (1.4) | 100 (0.0) | 100 (0.0) |
| | 500 | 1000 | 65 (4.8) | 99 (1.0) | 100 (0.0) | 100 (0.0) | 22 (4.1) |
| Experiment 3 | 1000 | 2000 | 89 (3.1) | 98 (1.4) | 100 (0.0) | 100 (0.0) | 100 (0.0) |
| | 500 | 1000 | 91 (2.9) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 5 (2.2) |
| Experiment 4 | 1000 | 2000 | 67 (4.7) | 98 (1.4) | 100 (0.0) | 100 (0.0) | 100 (0.0) |
| | 500 | 1000 | 59 (4.9) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 21 (4.1) |
| Experiment 5 | 1000 | 2000 | 82 (3.8) | 97 (1.7) | 100 (0.0) | 100 (0.0) | 100 (0.0) |
| | 500 | 1000 | 94 (2.4) | 99 (1.0) | 100 (0.0) | 100 (0.0) | 6 (2.4) |
| Experiment 6 | 1000 | 2000 | 0 (0.0) | 0 (0.0) | 99 (1.0) | 100 (0.0) | 100 (0.0) |
| Experiment 7 | 1000 | 2000 | 88 (3.2) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 100 (0.0) |
| | 500 | 1000 | 88 (3.2) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 8 (2.7) |
| Experiment 8 | 1000 | 2000 | 0 (0.0) | 0 (0.0) | 100 (0.0) | 100 (0.0) | 100 (0.0) |

out in the tail of the characteristic function. Thus, in order to estimate $\psi_Z(\sigma p_{\max})$ accurately, a much larger sample of Y_1, \dots, Y_m is needed. Thus, doubling both sample sizes m and n may not necessarily increase the accuracy of $\hat{\nu}$. However, for sufficiently large m (depending on the value of n), the resulting $\hat{\nu}$ will improve in accuracy.

EXPERIMENT 2. This experiment is identical to Experiment 1 except that, more realistically, the oracle spike detector is not used. Instead, as described in Section 4.2, we use the function `find.spikes.with.template` (with spike detection threshold 2.25σ) provided by SpikeOMatic to detect the spikes of the spike train. Table 3 gives a summary of the percentage of the time that $\hat{\nu} = \nu$, the true number of neurons producing the spike train. We observe that for moderately large sample sizes $n = 1000$ and $m = 2000$, the accuracy of $\hat{\nu}$ is still reasonable though not as high as in Experiment 1. This is due to the fact that spike detection is now not error free. We note that the scenario $\nu = 1$ with $\pi_{\text{cont}} \approx 0.1$ is unlikely to occur in practice due to the refractory period of a neuron which prevents the occurrence of overlapping spikes.

As a comparison, we shall now compute the SpikeOMatic estimate $\hat{\nu}_1$ of ν . Briefly, the SpikeOMatic software uses an EM algorithm to compute $\hat{\nu}_1$ based on a penalized likelihood function. The likelihood is a finite mixture of multivariate normal distributions and the penalty is derived from the Bayesian information criterion (BIC). This implies the assumption of Gaussian noise and no overlapping spikes. Another point of note is that the SpikeOMatic software assumes at

TABLE 4
Frequency (%) of $\hat{\nu}_1$ for Experiments 2 and 3 with $n = 1000$ and $m = 2000$

| | | $\hat{\nu}_1 = 1$ | $\hat{\nu}_1 = 2$ | $\hat{\nu}_1 = 3$ | $\hat{\nu}_1 = 4$ | $\hat{\nu}_1 = 5$ |
|--------------|-----------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Experiment 2 | $\nu = 1$ | 0 | 100 | 0 | 0 | 0 |
| | $\nu = 2$ | 0 | 100 | 0 | 0 | 0 |
| | $\nu = 3$ | 0 | 100 | 0 | 0 | 0 |
| | $\nu = 4$ | 0 | 20 | 80 | 0 | 0 |
| | $\nu = 5$ | 0 | 0 | 80 | 20 | 0 |
| Experiment 3 | $\nu = 1$ | 0 | 100 | 0 | 0 | 0 |
| | $\nu = 2$ | 0 | 100 | 0 | 0 | 0 |
| | $\nu = 3$ | 0 | 70 | 30 | 0 | 0 |
| | $\nu = 4$ | 0 | 0 | 100 | 0 | 0 |
| | $\nu = 5$ | 0 | 0 | 100 | 0 | 0 |

least 2 neurons generating the spike train and, hence, $\hat{\nu}_1$ is always ≥ 2 . There are 2 major tuning parameters for the SpikeOMatic software. We take these to be `nb.samples.per.site = 3` and `tolerance.factor = 3.5`. The other parameters are set as in the software tutorial 1 distributed with SpikeOMatic. The performance of $\hat{\nu}_1$ appears to be rather robust to the choice of the 2 tuning parameters. Table 4 gives the frequency (%) of $\hat{\nu}_1$ for 10 repetitions of the spike train. In particular, when $\nu = 2$, the frequency of $\hat{\nu}_1 = \nu$ is 100% and when $\nu = 3, 4$ or 5, the frequency of $\hat{\nu}_1 = \nu$ is 0%.

EXPERIMENT 3. In this experiment the noise is i.i.d. $N(0, 1)$. We assume that the spikes are detected from the spike train without error:

- The proportion of overlapping spikes $\pi_{\text{cont}} = 0$ or, equivalently, that there are no overlapping spikes.
- $n = 1000$ (or 500) and $m = 2n$.
- The approximate μ_i 's of these neurons are presented in Table 2.

Table 3 gives the percentage of the time (out of 100 repetitions) that $\hat{\nu} = \nu$.

As a comparison, we shall now compute the SpikeOMatic estimate $\hat{\nu}_1$ for ν . We take the major 2 tuning parameters to be `nb.samples.per.site = 3` and `tolerance.factor = 3.5`. The other parameters are set as in the software tutorial 1 distributed with SpikeOMatic. Table 4 gives the frequency (%) of $\hat{\nu}_1$ for 10 repetitions of the spike train. In particular, when $\nu = 2$, the frequency of $\hat{\nu}_1 = \nu$ is 100%, when $\nu = 3$, the frequency of $\hat{\nu}_1 = \nu$ is 30% and when $\nu = 4$ or 5, the frequency of $\hat{\nu}_1 = \nu$ is 0%.

EXPERIMENT 4. This experiment is identical to Experiment 3 except that we use the function `find.spikes.with.template` (with spike detection threshold 2.25σ)

provided by SpikeOMatic to detect the spikes of the spike train. Table 3 gives the percentage of the time (out of 100 repetitions) that $\hat{\nu} = \nu$.

EXPERIMENT 5. In this experiment the noise has the t_5 distribution multiplied by $\sqrt{3/5}$ (so as to have mean 0 and variance 1). We assume that the spikes are detected from the spike train without error:

- The proportion of overlapping spikes $\pi_{\text{cont}} \approx 0.1$ (or 10%).
- $n = 1000$ (or 500) and $m = 2n$.
- The approximate μ_i 's (i.e., the projections of the spike shape vector onto the first principal component $\alpha \in \mathbb{R}^d$) of these neurons are presented in Table 2.

Table 3 gives the percentage of the time (out of 100 repetitions) that $\hat{\nu} = \nu$.

EXPERIMENT 6. This experiment is identical to Experiment 5 except that we use the function `find.spikes.with.template` (with spike detection threshold 2.25σ) provided by SpikeOMatic to detect the spikes of the spike train. Table 3 gives the percentage of the time (out of 100 repetitions) that $\hat{\nu} = \nu$. Here the accuracy of $\hat{\nu}$ is poor for $\nu = 1$ or 2. This is almost certainly due to the heavy-tailed t_5 distribution presenting severe difficulties to the SpikeOMatic spike detection algorithm with threshold set to 2.25σ .

EXPERIMENT 7. In this experiment the noise has the t_5 distribution multiplied by $\sqrt{3/5}$ (so as to have mean 0 and variance 1). We assume that the spikes are detected from the spike train without error:

- The proportion of overlapping spikes $\pi_{\text{cont}} = 0.0$, that is, there are no overlapping spikes.
- $n = 1000$ (or 500) and $m = 2n$.
- The approximate μ_i 's of these neurons are presented in Table 2.

Table 3 gives the percentage of the time (out of 100 repetitions) that $\hat{\nu} = \nu$.

EXPERIMENT 8. This experiment is identical to Experiment 7 except that we use the function `find.spikes.with.template` (with spike detection threshold 2.25σ) provided by SpikeOMatic to detect the spikes of the spike train. Table 3 gives the percentage of the time (out of 100 repetitions) that $\hat{\nu} = \nu$. As in Experiment 6, the accuracy of $\hat{\nu}$ here is poor for $\nu = 1$ or 2 and is due to the heavy-tailed t_5 distribution presenting severe difficulties to the SpikeOMatic spike detection algorithm with threshold set to 2.25σ .

4.5. *Study summary.* The 8 experiments indicate that for moderately large sample sizes such as $n = 1000$ and $m = 2000$, $\hat{\nu}$ is capable of very good performance if the noise is i.i.d. Gaussian. If the noise has a heavy-tailed distribution such as the t_5 distribution, $\hat{\nu}$ is still capable of very good performance if an oracle (error free) spike detector is used. One reason for the good performance of $\hat{\nu}$ for moderately large samples is that in all the experiments the first principal component separates the spike shapes from one another as well as from pure noise very well.

However, for smaller sample sizes such as $n = 500$ and $m = 1000$, $\hat{\nu}$ performs reasonably well for $1 \leq \nu \leq 4$ if the noise is i.i.d. Gaussian but performs poorly for $\nu = 5$. This implies that the sample size $n = 500$ is too small for $\hat{\nu}$ to handle the latter case well.

5. A real data example. This section considers two spike train data sets. Both of these data sets are taken from Lewicki (1994). One data set is an actual 40-second spike train recording. The other data set is a synthesized recording using 6 spike shapes estimated from the former data set. For each of the 2 data sets, we shall compute and compare the estimates of the number of neurons generating the spike train by applying Lewicki's software SUN Solaris OS version 1.1.8 [cf. Lewicki (1994)], SpikeOMatic version 0.6-1 [cf. Pouzat, Mazor and Laurent (2002)] and our proposed method-of-moments estimator $\hat{\nu}$. For simplicity, let $\hat{\nu}_1$ denote the estimate of ν given by SpikeOMatic and $\hat{\nu}_2$ be corresponding to the estimate given by the Lewicki software.

5.1. *Synthesized data set.* The spike shapes, the number of spikes from each neuron and the number of overlapping spikes for the synthesized recording can be found in Section 7 of Lewicki (1994). The true number of neurons ν for the synthesized recording is 6. However, all three methods underestimate ν . In particular, $\hat{\nu} = 4$ while $\hat{\nu}_1 = \hat{\nu}_2 = 5$. A likely reason is that the two smallest spike shapes are too close to each other for them to be identified as two distinct spike shapes (and not one) by the 3 estimators. A graph of the 6 spike shapes is given in Figure 7 of Lewicki (1994).

With reference to $\hat{\nu}$, we have set the tuning parameters to $\gamma_{\text{threshold}} = 1$, $\sigma = 0.1$ and $p = p_{\text{max}}$ as in Section 4. p_{max} turns out to be 18. SpikeOMatic is used for spike detection and $n = 746$ spikes are detected. Table 5 lists the eigenvalues $\lambda_1(\hat{M}_p) \geq \dots \geq \lambda_{19}(\hat{M}_p)$ with $\lambda_5(\hat{M}_p) = 0.96 < \gamma_{\text{threshold}} = 1$. Thus, as it stands, our estimate $\hat{\nu}$ misses returning the value 5 by a very narrow margin. Also, we observe from Lewicki (1994) that $\pi_{\nu-1} + \pi_\nu \leq 32/895 \approx 0.036$, which is rather small. In fact, $(p_{\text{max}} + 1)0.036 \approx 0.68$, which is way below $\gamma_{\text{threshold}}$. Thus, if we suspect that π_ν is this small, we should lower the value of $\gamma_{\text{threshold}}$ to below 1. For example, if $\gamma_{\text{threshold}}$ is set to be 0.8 say, we shall obtain $\hat{\nu} = 5$ for the synthesized data set. However, reducing the value of $\gamma_{\text{threshold}}$ would, of course, increase the chance of false positives too.

TABLE 5
The eigenvalues in decreasing order computed from our algorithm for the two data sets

| | | | | | | | | | | |
|-----------------------|------|------|------|------|-------|-------|-------|-------|-------|------|
| Synthesized recording | 7.86 | 4.73 | 4.52 | 2.02 | 0.96 | 0.25 | 0.10 | 0.05 | 0.03 | 0.01 |
| | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.02 | -0.09 | -1.42 | |
| Actual recording | 6.64 | 3.83 | 1.59 | 1.17 | 0.86 | 0.67 | 0.48 | 0.33 | 0.20 | 0.14 |
| | 0.11 | 0.07 | 0.05 | 0.01 | -0.04 | -0.11 | | | | |

5.2. *Actual spike train data.* When the 3 methods are applied to the actual spike train data set, we obtain $\hat{\nu} = 4$ while $\hat{\nu}_1 = 12$ and $\hat{\nu}_2 = 9$. Lewicki [(1994), page 1020] inferred that the number of neurons ν generating the actual spike train recording is still 6. Thus, it would appear that $\hat{\nu}_1$ and $\hat{\nu}_2$ have overestimated ν and that they give rather different values of ν for the synthesized and actual data sets. On the other hand, $\hat{\nu}$, with the tuning parameters as in Section 4, gives $\hat{\nu} = 4$ as in the synthesized data set. This shows that $\hat{\nu}$ is rather stable and is probably less variable than $\hat{\nu}_1$ or $\hat{\nu}_2$. As in the discussion of Section 5.1, if we suspect that π_ν is very small such that $(p_{\max} + 1)\pi_\nu < 1$, we should lower the value of $\gamma_{\text{threshold}}$ in order to detect this spike shape. If we reduce the value of $\gamma_{\text{threshold}}$ to 0.8, we obtain $\hat{\nu} = 5$ (from the eigenvalues of Table 5).

We conclude this section by presenting the parameter settings for the three methods. The parameters for Lewicki's software are set as his default values for the analysis of the synthesized recording. Here a spike is defined as the window of measurements that are within 1.0 millisecond prior to the occurrence of the peak and 4.0 milliseconds after the occurrence of the peak. Since the sampling frequency is 20 kHz, there are 20 measurements before the peak and 80 measurements after the peak.

The parameters for SpikeOMatic software are given in Table 6. For the detailed explanation of the meaning of the parameters, we refer the reader to the SpikeOMatic manual which comes along with the software. We further note that the spike length is chosen to be $d = 100$, the same as for Lewicki's software, and $n = 1447$ spikes are detected with $m = 2000$.

With respect to $\hat{\nu}$, the spike detection procedure of SpikeOMatic is used. This part of the parameter setting is the same as in Table 6. Here $d = 100$, spike detection threshold = 3.00σ and the other tuning parameters for $\hat{\nu}$ are the same as in Section 4.

6. Concluding remarks. In conclusion, this article proposes a new estimator $\hat{\nu}$ for estimating the number of neurons ν in a multi-neuronal spike train. $\hat{\nu}$ has a number of advantages over alternative estimators for ν in the existing literature. First, it is a method-of-moments estimator and uses trigonometric moment matrices in its construction (unlike maximum likelihood estimators). As a result, the assumptions needed for $\hat{\nu}$ are minimal. Indeed, the model (4) on which it is based

TABLE 6
Parameter settings for SpikeOMatic

| Parameter | Value |
|---------------------|-------|
| template.length | 160 |
| threshold | 3.0 |
| sweep.length | 100 |
| peak.location | 20 |
| nb.noise.evt | 2000 |
| nb.samples.per.site | 6 |
| tolerance.factor | 3.5 |
| nb.clusters.min | 2 |
| nb.cluster.max | 12 |
| nb.iterations | 25 |
| nb.tries | 20 |

is a nonparametric mixture distribution. (4) takes explicitly into account the possibility of overlapping spikes while no parametric assumptions are made on the noise distribution or the overlapping spike distribution.

Second, we have managed to develop a rigorous nonasymptotic theory in support of $\hat{\nu}$. This theory is reasonably simple and transparent. In particular, it shows that $\hat{\nu}$ is a strongly consistent estimator of ν under mild conditions. Also, perhaps more importantly, the nonasymptotic error bounds of Theorems 1 and 3 provide us with a way of setting the tuning parameters $\gamma_{\text{threshold}}$, p and σ so as to ensure that $\hat{\nu}$ performs well in practice. The latter is further justified by applying $\hat{\nu}$ to a number of spike train simulations in Section 4 and to an actual spike train data set in Section 5.

Finally, we have assumed independent noise (i.e., the $\eta_{i,k}$'s and $\eta_{l,j}^*$'s of Section 1) throughout this article. If the noise is a stationary and weakly dependent process, Theorem 2 still holds (i.e., $\hat{\nu}$ is strongly consistent) as long as

$$(14) \quad \begin{aligned} \frac{1}{n} \sum_{i=1}^n e^{-i(j-k)X_i} &\rightarrow E e^{-i(j-k)X_1}, \\ \frac{1}{m} \sum_{l=1}^m e^{-i(j-k)Y_l} &\rightarrow E e^{-i(j-k)Y_1} \quad \forall 1 \leq j, k \leq p+1, \end{aligned}$$

almost surely as $\min(m, n) \rightarrow \infty$. We observe that (14) is a rather mild condition and is satisfied by many weakly dependent processes.

APPENDIX

PROOF OF THEOREM 1. First suppose that $i \in \{1, \dots, \nu\}$. We observe from Lemma 3 below that $\lambda_i(M_{p,\text{disc}}) = \lambda_i(B)$. Let $B^\dagger = \text{diag}((p+1)\pi_1, \dots, (p+1)\pi_\nu)$.

1) π_v). Using Theorem A.37 of Bai and Silverstein (2009), we have

$$\begin{aligned} \sum_{i=1}^v [\lambda_i(B) - \lambda_i(B^\dagger)]^2 &\leq \sum_{j=1}^v \sum_{k=1}^v |B_{jk} - B_{jk}^\dagger|^2 \\ &= 2 \sum_{1 \leq j < k \leq v} \pi_j \pi_k \left| \frac{1 - e^{i(p+1)(\mu_j - \mu_k)}}{1 - e^{i(\mu_j - \mu_k)}} \right|^2. \end{aligned}$$

Thus,

$$|\lambda_i(M_{p,\text{disc}}) - (p + 1)\pi_i| \leq \sqrt{2 \sum_{1 \leq j < k \leq v} \pi_j \pi_k \left| \frac{1 - e^{i(p+1)(\mu_j - \mu_k)}}{1 - e^{i(\mu_j - \mu_k)}} \right|^2}.$$

Since $M_p = M_{p,\text{disc}} + \pi_{\text{cont}} M_{p,\text{cont}}$, we observe from Corollary 4.9 of Stewart and Sun (1990) and Lemma 4 that

$$\begin{aligned} \lambda_i(M_p) &\geq \lambda_i(M_{p,\text{disc}}) + \pi_{\text{cont}} \lambda_{p+1}(M_{p,\text{cont}}) \\ &\geq (p + 1)\pi_i + 2\pi \pi_{\text{cont}} \left\{ \min_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j) \right\} \\ &\quad - \sqrt{2 \sum_{1 \leq j < k \leq v} \pi_j \pi_k \left| \frac{1 - e^{i(p+1)(\mu_j - \mu_k)}}{1 - e^{i(\mu_j - \mu_k)}} \right|^2}, \\ \lambda_i(M_p) &\leq \lambda_i(M_{p,\text{disc}}) + \pi_{\text{cont}} \lambda_1(M_{p,\text{cont}}) \\ &\leq (p + 1)\pi_i + 2\pi \pi_{\text{cont}} \left\{ \max_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j) \right\} \\ &\quad + \sqrt{2 \sum_{1 \leq j < k \leq v} \pi_j \pi_k \left| \frac{1 - e^{i(p+1)(\mu_j - \mu_k)}}{1 - e^{i(\mu_j - \mu_k)}} \right|^2}. \end{aligned}$$

This proves the first statement of Theorem 1. Next, suppose that $i \in \{v + 1, \dots, p + 1\}$. Then $\lambda_i(M_{p,\text{disc}}) = 0$. Using Corollary 4.9 of Stewart and Sun (1990) and Lemma 4 again, we obtain the second statement of Theorem 1. \square

LEMMA 2. *Let $m > n$ be positive integers and A be a $m \times n$ matrix with complex-valued entries. Then the eigenvalues of AA^* are the eigenvalues of A^*A and $(m - n)$ zeros where A^* denotes the conjugate transpose of A .*

PROOF. We observe from the singular value decomposition of A that $A = UDV^*$ where U is a $m \times m$ unitary matrix, D a $m \times n$ diagonal matrix with nonnegative real numbers on the diagonal and V a $n \times n$ unitary matrix. Then $A^*A = VD^*DV^*$, and $AA^* = UDD^*U^*$. Lemma 2 follows since D^*D and DD^* are both diagonal matrices. \square

PROOF. Let $a = (a_1, \dots, a_{p+1})' \in \mathbb{C}^{p+1}$. Then

$$\begin{aligned} \frac{a^* M_{p,\text{cont}} a}{a^* a} &= \frac{\sum_{k=1}^{p+1} \sum_{j=1}^{p+1} [\int_0^{2\pi} \sum_{l=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi l) e^{i(k-j)\mu} d\mu] a_k a_j^*}{a^* a} \\ &= \frac{\int_0^{2\pi} |\sum_{k=1}^{p+1} a_k e^{ik\mu}|^2 \sum_{l=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi l) d\mu}{\sum_k |a_k|^2} \\ &= \frac{2\pi \int_0^{2\pi} |\sum_{k=1}^{p+1} a_k e^{ik\mu}|^2 \sum_{l=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi l) d\mu}{\int_0^{2\pi} |\sum_k a_k e^{ik\mu}|^2 d\mu}. \end{aligned}$$

Thus, for an arbitrary $a \in \mathbb{C}^{p+1}$ such that $a^* a = 1$,

$$\begin{aligned} &2\pi \min_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j) \\ &\leq a^* M_{p,\text{cont}} a \\ &\leq 2\pi \max_{0 \leq \mu < 2\pi} \sum_{j=-\infty}^{\infty} f_{\text{cont}}(\mu + 2\pi j). \end{aligned}$$

Since $\lambda_1(M_{p,\text{cont}}) = \sup_{a \in \mathbb{C}^{p+1}, a^* a = 1} a^* M_{p,\text{cont}} a$ and $\lambda_{p+1}(M_{p,\text{cont}}) = \inf_{a \in \mathbb{C}^{p+1}, a^* a = 1} a^* M_{p,\text{cont}} a$, Lemma 4 is proved. \square

PROOF OF THEOREM 2. We observe from the strong law of large numbers that $\hat{M}_p \rightarrow M_p$ almost surely as $\min(m, n) \rightarrow \infty$. This implies that $\lambda_{\nu+1}(\hat{M}_p) < \gamma \sqrt{p+1} < \lambda_{\nu}(\hat{M}_p)$ almost surely as $\min(m, n) \rightarrow \infty$. \square

PROOF OF LEMMA 1. Using Markov's inequality, we observe for $\ell = 1, \dots, p$ that

$$\begin{aligned} P(\Omega_{\ell, \varepsilon}) &= P\left(\left|\frac{1}{m} \sum_{j=1}^m (e^{-i\ell Y_j} - E e^{-i\ell Y_1})\right| \geq \varepsilon |E(e^{-i\ell Y_1})|\right) \\ &\leq \frac{E[|m^{-1} \sum_{j=1}^m (e^{-i\ell Y_j} - E e^{-i\ell Y_1})|^k]}{[\varepsilon |E(e^{-i\ell Y_1})|]^k} \quad \forall k \in \mathbb{Z}^+. \end{aligned}$$

Taking $k = 3, 4$, we obtain

$$\begin{aligned} P(\Omega_{\ell, \varepsilon}) &\leq \min \left\{ \frac{6}{m^2 [\varepsilon |E(e^{-i\ell Y_1})|]^4} \left(1 + O\left(\frac{1}{m}\right)\right), \right. \\ &\quad \left. \frac{42}{m^3 [\varepsilon |E(e^{-i\ell Y_1})|]^6} \left(1 + O\left(\frac{1}{m}\right)\right) \right\}. \quad \square \end{aligned}$$

PROOF OF THEOREM 3. Using Theorem A.37 of Bai and Silverstein (2009), we have

$$\begin{aligned}
 & E^{\Omega^c} \sum_{i=1}^{p+1} [\lambda_i(\hat{M}_p) - \lambda_i(M_p)]^2 \\
 & \leq E^{\Omega^c} \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} |(\hat{M}_p)_{j,k} - (M_p)_{j,k}|^2 \\
 & = 2 \sum_{j=1}^p (p-j+1) E^{\Omega^c} \left| \frac{n^{-1} \sum_{i=1}^n e^{-ijX_i}}{m^{-1} \sum_{l=1}^m e^{-ijY_l}} - \frac{E(e^{-ijX_1})}{E(e^{-ijY_1})} \right|^2 \\
 & = 2 \sum_{j=1}^p (p-j+1) \left[E^{\Omega^c} \left| \frac{n^{-1} \sum_{i=1}^n e^{-ijX_i}}{m^{-1} \sum_{l=1}^m e^{-ijY_l}} - \frac{E(e^{-ijX_1})}{m^{-1} \sum_{l=1}^m e^{-ijY_l}} \right|^2 \right. \\
 & \qquad \qquad \qquad \left. + E^{\Omega^c} \left| \frac{E(e^{-ijX_1})}{m^{-1} \sum_{l=1}^m e^{-ijY_l}} - \frac{E(e^{-ijX_1})}{E(e^{-ijY_1})} \right|^2 \right] \\
 & = 2 \sum_{j=1}^p (p-j+1) \left[E^{\Omega^c} \left| \frac{n^{-1} \sum_{i=1}^n e^{-ijX_i}}{m^{-1} \sum_{l=1}^m e^{-ijY_l}} - \frac{E(e^{-ijX_1})}{m^{-1} \sum_{l=1}^m e^{-ijY_l}} \right|^2 \right. \\
 & \qquad \qquad \qquad \left. + |E(e^{-ijX_1})|^2 E^{\Omega^c} \left| \frac{1}{m^{-1} \sum_{l=1}^m e^{-ijY_l}} - \frac{1}{E(e^{-ijY_1})} \right|^2 \right] \\
 & \leq 2 \sum_{j=1}^p (p-j+1) \left[\frac{E|n^{-1} \sum_{i=1}^n e^{-ijX_i} - E(e^{-ijX_1})|^2}{(1-\varepsilon)^2 |E(e^{-ijY_1})|^2} \right. \\
 & \qquad \qquad \qquad \left. + \frac{\varepsilon^2 |E(e^{-ijX_1})|^2}{(1-\varepsilon)^2 |E(e^{-ijY_1})|^2} \right] \\
 & \leq \frac{2}{(1-\varepsilon)^2} \sum_{j=1}^p (p-j+1) \left[\frac{1}{n |E(e^{-ijY_1})|^2} + \varepsilon^2 \right]. \quad \square
 \end{aligned}$$

Acknowledgments. Wei-Liem Loh would like to thank Professors Rob Kass, Shinsuke Koyama and Matt Harrison for the many discussions on neural spikes trains when he visited the Department of Statistics, Carnegie Mellon University in Fall 2006 and Summer 2007. We would also like to thank the Editor, Professor Karen Kafadar and an Associate Editor for their suggestions and comments that motivated Sections 5 and 4, respectively.

REFERENCES

BAI, Z. D. and SILVERSTEIN, J. W. (2009). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

- BRILLINGER, D. R. (1988). Some statistical methods for random process data from seismology and neurophysiology. *Ann. Statist.* **16** 1–54. [MR0924855](#)
- BROWN, E. N., KASS, R. E. and MITRA, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature Neurosci.* **7** 456–461.
- DACUNHA-CASTELLE, D. and GASSIAT, E. (1997). The estimation of the order of a mixture model. *Bernoulli* **3** 279–299. [MR1468306](#)
- DAYAN, P. and ABBOTT, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, Cambridge, MA. [MR1985615](#)
- GARCÍA-ESCUADERO, L. A., GORDALIZA, A., MATRÁN, C. and MAYO-ISCAR, A. (2008). A general trimming approach to robust cluster analysis. *Ann. Statist.* **36** 1324–1345. [MR2418659](#)
- GALLEGOS, M. T. and RITTER, G. (2005). A robust method for cluster analysis. *Ann. Statist.* **33** 347–380. [MR2157806](#)
- LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York. [MR0702834](#)
- LEWICKI, M. S. (1994). Bayesian modeling and classification of neural signals. *Neural Comput.* **6** 1005–1030.
- LEWICKI, M. S. (1998). A review of methods for spike sorting: The detection and classification of neural action potentials. *Network Comput. Neural Syst.* **9** R53–R78.
- LINDSAY, B. G. (1989a). On the determinants of moment matrices. *Ann. Statist.* **17** 711–721. [MR0994262](#)
- LINDSAY, B. G. (1989b). Moment matrices: application in mixtures. *Ann. Statist.* **17** 722–740. [MR0994263](#)
- MOKRI, Y. and YEN, S. C. (2008). Spike overlap resolution of electrode and tetrode data from primary visual cortex. *BMC Neurosci.* **9**(Suppl. 1) 94.
- NGUYEN, D. P., FRANK, L. M. and BROWN, E. N. (2003). An application of reversible-jump MCMC to spike classification of multiunit extracellular recordings. *Network Comput. Neural Syst.* **14** 61–82. [MR2040807](#)
- POUZAT, C., MAZOR, O. and LAURENT, G. (2002). Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *J. Neurosci. Methods* **122** 43–57.
- POUZAT, C., DELESCLUSE, M., VOIT, P. and DIEBOLT, J. (2004). Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: A Markov chain Monte Carlo approach. *J. Neurophysiol.* **91** 2910–2928.
- SAHANI, M. (1999). *Latent Variable Models for Neural Data Analysis*. PhD thesis, Caltech, Pasadena, CA. [MR2699221](#)
- SHOHAM, S., FELLOWS, M. R. and NORMANN, R. A. (2003). Robust, automatic spike sorting using mixtures of multivariate t -distributions. *J. Neurosci. Methods* **127** 111–122.
- STEWART, G. W. and SUN, J. (1990). *Matrix Perturbation Theory*. Academic Press, Boston. [MR1061154](#)
- VENTURA, V. (2009). Traditional waveform based spike sorting yields biased rate code estimates. *Proc. Natl. Acad. Sci.* **106** 6921–6926.
- WOOD, F. and BLACK, M. J. (2008). A nonparametric Bayesian alternative to spike sorting. *J. Neurosci. Methods* **173** 1–12.

GENOME INSTITUTE OF SINGAPORE
60 BIOPOLIS STREET
#02-01, GENOME
SINGAPORE 138672
REPUBLIC OF SINGAPORE
E-MAIL: doadvan@gmail.com

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117543
REPUBLIC OF SINGAPORE
E-MAIL: stalohwl@nus.edu.sg