J. R. Statist. Soc. B (1999)
61, *Part* 1, *pp.* 191–212

Double-blind deconvolution: the analysis of post-synaptic currents in nerve cells

D. S. Poskitt, K. Doğançay and Shin-Ho Chung
Australian National University, Canberra, Australia

[Received April 1996. Final revision February 1998]

Summary. This paper is concerned with the analysis of observations made on a system that is being stimulated at fixed time intervals but where the precise nature and effect of any individual stimulus is unknown. The realized values are modelled as a stochastic process consisting of a random signal embedded in noise. The aim of the analysis is to use the data to unravel the unknown structure of the system and to ascertain the probabilistic behaviour of the stimuli. A method of parameter estimation based on quasi-profile likelihood is presented and the statistical properties of the estimates are established while recognizing that there will be a discrepancy between the model and the true data-generating mechanism. A method of model validation and determination is also advanced and kernel smoothing techniques are proposed as a basis for identifying the amplitude distribution of the stimuli. The data processing techniques described have a direct application to the investigation of excitatory post-synaptic currents recorded from nerve cells in the central nervous system and their use in quantal analysis of such data is illustrated.

Keywords: Blind deconvolution; Excitatory post-synaptic currents; Finite algorithm; Gaussian estimator; Gauss–Newton recursions; Initial estimates; Kernel smoothing; Quantal analysis; Quasi-profile likelihood

1. Introduction

Consider a situation where T data values are derived from a stochastic process $\{y(t)\}$ defined by the signal plus noise relationship

$$y(t) = s(t, u) + \epsilon(t), \qquad t = 0, \ldots, T-1, \tag{1}$$

where $\{s(t, u)\}$ denotes an unknown signal that is generated by passing an input process $\{u(t)\}$ through an observed system and $\{\epsilon(t)\}$ denotes a zero-mean, white Gaussian noise disturbance with variance σ_ϵ^2, independent of the signal, that represents such features as background noise and measurement error. The input $\{u(t)\}$, like the individual signal and noise components, is not directly observable. It is made up of a periodic pulse train where the time interval between pulses is given but the amplitude of each pulse is random. We express this algebraically as

$$u(t) = \sum_{r=1}^{N} \delta\{t - (r-1)L\} a_r \tag{2}$$

where $\delta(t)$ is the Kronecker delta function,

Address for correspondence: D. S. Poskitt, Department of Statistics, Australian National University, Canberra, ACT 0200, Australia.
E-mail: Don.Poskitt@anu.edu.au

 1369–7412/99/61191

$$\delta(t) = \begin{cases} 1, & \text{if } t = 0, \\ 0, & \text{if } t \neq 0, \end{cases}$$

$L = T/N$ with L, N and T known integers and a_r, $r = 1, \ldots, N$, are independent and identically distributed scalar random variables drawn from a probability distribution $\mathbb{P}_A$ with support contained in $\{0\} \cup \mathbb{R}^+$, $\mathbb{R}^+ = \{x: x > 0\}$. This formulation is employed since in the applications that we have in mind, namely the analysis of observations on evoked synaptic responses, the data are obtained by stimulating a neuron at fixed points in time that are subject to experimental control. Exactly the same stimulus is applied at each of these time points but the observed response of the neuron varies from one stimulus to the next in such a way that the precise effect of each stimulus is not known beforehand and is thus outside the experimenters' control. The objective of the analysis is to use $y(t)$, $t = 0, \ldots, T-1$, to model the structure of the system and to ascertain the values and probabilistic properties of the amplitude variations of the stimuli.

Problems that are similar to that considered here have been discussed in the literature, under the assumptions that the signal is observed without error, that the system is linear and time invariant and that the input either satisfies some form of stationarity condition or has a lattice distribution; see Donoho (1981), Lii and Rosenblatt (1982), Cheng (1990), Davis and Rosenblatt (1991), Li (1995) and Gamboa and Gassiat (1996), for example. The aim, in general, is to determine the unknown structure of the system and to reconstruct the unobserved input on the basis only of the observed data, and the term blind deconvolution has been adopted to describe this process. A common feature of these references is that the techniques discussed determine a method of deconvolution whose rationale depends on having partial knowledge of the probabilistic structure of the input process and they are motivated by a variety of applications, such as engineering, geophysics and digital communications, where such information may be available. With recordings made on post-synaptic currents, however, not only is the signal not observable without error but also nothing is known about the probability characteristics of the amplitude variations in the evoked responses. These two features have led us to use the nomenclature double-blind deconvolution to describe the procedures that we are about to consider.

Our work is directly relevant to the solution of some outstanding problems in neurobiology. Fig. 1 is a schematic diagram of a post-synaptic neuron. As shown in Fig. 1, a small number of nerve terminals a form excitatory synaptic contacts with the dendrite d of the nerve cell. When a nerve fibre is activated an electrical impulse invades the terminals causing the release of packets b of neurotransmitter molecules. The released molecules diffuse across the narrow gap causing ion channels c located on the dendrite of the post-synaptic neuron to open. When a channel opens, typically for only a few milliseconds, approximately 10^7–10^8 sodium ions per second flow from outside to the inside of the dendrite. Because the duration that each channel stays open is governed by a first-order Markov process the total current flux brought about by the activation of the different channels rises nearly instantaneously and decays exponentially. The nerve signal generated in this way passes along the dendrite to the cell body e and is propagated to the next nerve cell via the axon f. Owing to the physical dimensions of the dendrite the measurement of the current that flows across the membrane ion channels, known as the excitatory post-synaptic current, can only be made from the cell body, which is some distance from the synaptic site. The original signal is therefore attenuated and its shape distorted owing to the cable properties of the dendrite. Moreover, the signal evoked is contaminated and obscured by the amplifier and electrode noise of the recording device.

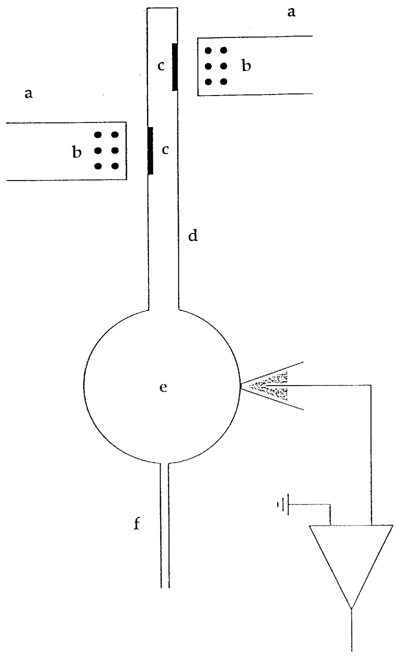

Fig. 1. Schematic diagram of a post-synaptic neuron together with two presynaptic terminals: the current evoked is recorded at the cell body with a voltage-to-current converter (see the text for detailed particulars of the annotation)

Observations of this type made on evoked responses obtained from nerve cells in the central nervous system allow many questions of fundamental scientific significance to be addressed. For example, it is important to determine whether each packet contains approximately equal numbers of neurotransmitter molecules and, if so, whether or not the number contained in each packet increases when the synaptic link between the nerve terminal and cell body is used more frequently. It is also important to ascertain whether the probability that the terminal releases one or more packets of neurotransmitter when stimulated can be characterized as having certain statistical properties. Such quantitative analysis is currently not possible because of the low signal-to-noise ratios that are encountered in practice. As will be demonstrated later, the sensitivity of the processing scheme described here now makes it possible to answer such questions by using a limited set of excitatory post-synaptic current recordings.

In the light of the previous discussion we shall consider modelling process (1) by using a parametric specification of the form

$$y(t) = \sum_{j \geqslant 0} h_\alpha(j)\, u(t-j) + e(t) \qquad t = 0, \ldots, T-1,$$

where

(a) the input process $u(t)$ follows the regime as specified in equation (2) (assumption M1) and
(b) the impulse response function

$$H_\alpha(z) = \sum_{j\geqslant 0} h_\alpha(j)z^{-j}$$

corresponds to a stable, causal low pass filter given by $H_\alpha(z) = \alpha(z)^{-1}$ where

$$\alpha(z) = \prod_{j=1}^{p}(1 - \xi_j z^{-1})$$

(assumption M2). The order p of the operator $\alpha(z) = 1 + \alpha_1 z^{-1} + \ldots + \alpha_p z^{-p}$ is constant and $0 < |\xi_j| < 1$, $j = 1, \ldots, p$.

Notice that model assumption M2 implies that the normalization rule $h_\alpha(0) = 1$ is being adopted to remove the scale ambiguity that is inherent to the solution of the problem. The restriction placed on the ξ_j, $j = 1, \ldots, p$, reflects the presumption that the system being observed is both causal, in that $y(t)$ depends only on $u(s)$, $s \leqslant t$, and stable, i.e. bounded input values produce a bounded output. The biophysical interpretation of this condition is that the nerve cell is passive until stimulated and that the total energy flowing through the neuron is dissipative. Assumptions M1 and M2 lead to the representation

$$y\{t + (r-1)L\} = a_r\, h_\alpha(t) + \rho_{r,L}(t) + e\{t + (r-1)L\},$$

for $t = 0, \ldots, L-1$ and $r = 1, \ldots, N$ where

$$\rho_{r,L}(t) = \sum_{j=0}^{r-1} a_j\, h_\alpha\{t + (r-j)L\}, \qquad a_0 = 0\,,$$

represents the truncation error that arises from neglecting the effect of all except the most recent stimulus or evoked response. By the triangular inequality

$$|\rho_{r,L}(t)| \leqslant \sum_{j=0}^{r-1} |a_j||h_\alpha\{t + (r-j)L\}|,$$

which is bounded above by

$$\rho_{r,L} = \sum_{j=0}^{r-1} |a_j| \sum_{s=0}^{L-1} |h_\alpha\{s + (r-j)L\}|$$

for all t. But

$$\rho_{r,L} \leqslant \sum_{j=1}^{N} |a_j| \sum_{s=L}^{\infty} |h_\alpha(s)|$$

uniformly in r and assumption M2 implies that the factor

$$\sum_{s\geqslant L} |h_\alpha(s)| \leqslant \text{ constant} \times \lambda^L/(1-\lambda), \qquad \lambda = \max_j |\xi_j| < 1.$$

It follows that by choosing L sufficiently large we can make $\rho_{r,L}$ arbitrarily small. Thus it seems reasonable to suppose that for any process $\{y(t)\}$ we can, by making L sufficiently large, ensure that the truncation error is sufficiently small in relation to the size of the residual

$$e_r(t) = y\{t + (r-1)L\} - a_r\, h_\alpha(t), \qquad t = 0, \ldots, L-1, \quad r = 1, \ldots, N,$$

that it can almost surely be neglected. This assumption is formalized as follows (assumption M3).

Let $e_r(t) = y\{t + (r-1)L\} - a_r\, h_\alpha(t)$ for $t = 0, \ldots, L-1$ and $r = 1, \ldots, N$. Then

$$\begin{aligned} e_r(t) &= \rho_{r,L}(t) + e\{t + (r-1)L\} \\ &= \epsilon\{t + (r-1)L\} + s\{t + (r-1)L, u\} - a_r\, h_\alpha(t) \end{aligned}$$

and for every $\eta > 0$ there is an L_η such that, for all r and t, $\mathbb{P}[|e_r(t)| > |\rho_{r,L}(t)|] > 1 - \eta$ whenever $L > L_\eta$ uniformly in $(\alpha_1, \ldots, \alpha_p)$.

(Here and throughout the paper statements involving probability calculations will be assumed to be evaluated using the true probability distribution induced by the data-generating mechanism given in equations (1) and (2) although this will not be made explicit in the notation.) Heuristically assumption M3 implies that the realization $y(t)$, $t = 0, \ldots, T-1$, can be divided into N disjoint segments of length L in each of which the observed values are modelled as the sum of two components: the product of the impulse response of a linear, time-invariant system with the amplitude evoked by the most recent stimulus, plus a residual that encapsulates features such as measurement error, background noise and the modelling approximation error. A block diagram depicting our basic model and its relationship to the stylized post-synaptic neural transmission process is shown in Fig. 2.

Before closing this section let us define some additional notation and assumptions. The quantities that we wish to estimate are σ^2, $\mathbf{a} = (a_1, \ldots, a_N)'$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$, which we place together into a parameter vector $\boldsymbol{\theta} = (\sigma^2, \mathbf{a}', \boldsymbol{\alpha}')'$. The parameter space $\Theta = \mathbb{R}^+ \times (\{0\} \cup \mathbb{R}^+)^N \times \mathbb{R}^p(C)$ where $\mathbb{R}^p(C)$ indicates the subset of $\mathbb{R}^p$ such that $\alpha(z) \neq 0$ for $|z| \geqslant 1$, $\boldsymbol{\alpha} \in \mathbb{R}^p$. For any $\boldsymbol{\alpha} \in \mathbb{R}^p$ we shall use $\|\boldsymbol{\alpha}\|$ to denote the Euclidean norm and $\|\mathbf{A}\|$ will denote the corresponding Frobenius norm of the matrix $\mathbf{A} = (a_{i,j})$. For any $\eta > 0$, $N_\eta(\dot{\boldsymbol{\alpha}}) = \{\boldsymbol{\alpha} |\ \|\boldsymbol{\alpha} - \dot{\boldsymbol{\alpha}}\| < \eta\}$ where, in the usual set builder notation, the radius may be determined such that $N_\eta(\dot{\boldsymbol{\alpha}}) \subseteq \{\boldsymbol{\alpha} | \mathcal{P}(\boldsymbol{\alpha})\}$ for some property $\mathcal{P}(\boldsymbol{\alpha})$. The true structure of the process as

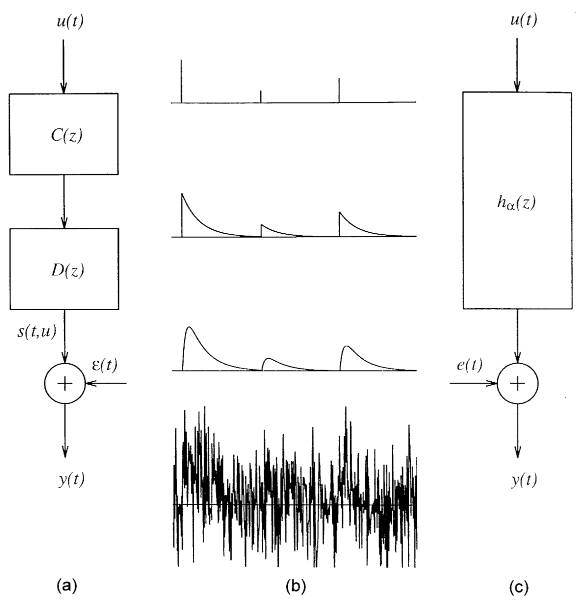

Fig. 2. Signal model: (a) hypothetical transformation of the input $u(t)$ to the observed sequence $y(t)$ ($C(z)$, current flux across the ion channels; $D(z)$, distortion of the current as it flows along the dendrite towards the cell body); (b) recording made at the cell body comprising signal contaminated by noise; (c) statistical specification ($C(z)$ and $D(z)$ are jointly modelled as $H_\alpha(z)$)

described in the discussion surrounding equations (1) and (2) will be labelled assumption P1, to which we append the following conditions.

(a) For all t, $|s(t, u)| < m(a)$ where

$$\int m(a)^{\beta+\beta'} \, \mathrm{d}\mathbb{P}_A < \infty, \qquad \beta > 2, \quad \beta' > 0,$$

and $\{s(t, u)\}$ is alpha mixing with mixing coefficients c_n such that $\Sigma_{n \geqslant 1} c_{n+L}^{(1-2/\beta)} < \infty$ for $L > L'$ (assumption P2).

(b) The approximation error

$$\Delta_\theta(r, t, L) = s\{t + (r-1)L, u\} - a_r h_\alpha(t),$$

$r = 1, \ldots, N$, $t = 0, \ldots, L-1$, is such that

$$(NL)^{-1} \sum_{r=1}^{N} \sum_{t=0}^{L-1} \Delta_\theta(r, t, L)^2$$

converges almost surely to a finite limit. Moreover, there exists a p_L, non-decreasing in L, $p_L = O(L^q)$, $0 < q < \frac{1}{2}$, with associated parameter values $(a_1, \ldots, a_N)$ and $(\alpha_1, \ldots, \alpha_p)$ such that the mean-squared approximation error is bounded by $2\sigma_\epsilon^2 \log\{\log(L)\}/L$ almost surely (assumption P3).

The first part of assumption P2 states that oscillations in the signal induced by amplitude variations in the evoked response will be bounded in probability. The second part recognizes that although $\{u(t)\}$ is an independent process the signal is likely to exhibit some form of strong dependence. If $\{s(t, u)\}$ is alpha mixing, however, then, for any events $\mathcal{A}_t \in \mathbb{F}\{s(\tau, u)$: $\tau \leqslant t\}$, the σ-field generated by $s(\tau, u)$, $\tau \leqslant t$, and $\mathcal{B}_{t+n} \in \mathbb{F}\{s(\tau, u)$: $\tau > t + n\}$ the mixing coefficients

$$c_n = \sup_t |\mathbb{P}(\mathcal{A}_t \cap \mathcal{B}_{t+n}) - \mathbb{P}(\mathcal{A}_t)\,\mathbb{P}(\mathcal{B}_{t+n})|$$

are such that $c_n \to 0$ as $n \to \infty$; see Billingsley (1986). This means that members of the sequence $\{s(t, u)\}$ will become independent as they become more separated in time and ensures that the effects of the different evoked responses can be disentangled provided that the time interval between successive stimuli is sufficiently large. Assumption P3 is motivated by the Stone–Weierstrass approximation theorem and establishes a bound on the rate of growth of the approximation error as the order of $\alpha(z)$ increases with the length of the interval between successive stimuli. The import of assumption P3 is that, although we recognize that the true data-generating mechanism and the model are unlikely to coincide, the discrepancy between the true signal and that given by the model cannot be so large as to make the mean-squared error unbounded for otherwise there would be no way of ascertaining whether the following estimation techniques, which are based on least-squares-type calculations, have any desirable qualities.

2. Quasi-profile likelihood and the Gaussian estimator

To estimate the parameter vector $\boldsymbol{\theta}$ we shall consider minimizing the criterion function

$$G_{N,L}(\boldsymbol{\theta}) = \frac{NL}{2}\log(\sigma^2) + \sum_{r=1}^{N}\sum_{t=0}^{L-1}\frac{e_r(t)^2}{2\sigma^2} \tag{3}$$

with respect to $\boldsymbol{\theta}$. The function $G_{N,L}(\boldsymbol{\theta})$ is motivated by noting from assumption P1 that the conditional distribution of $y(t)$, $t = 0, \ldots, T-1$, given the unobserved impulse values a_j, $j = 1, \ldots, N$, is Gaussian and $2\pi^{-NL/2}\exp\{-G_{N,L}(\boldsymbol{\theta})\}$ yields the conditional likelihood when the model obtains. Note that the complete likelihood is unavailable because $\mathbb{P}_A$ is unknown and therefore application of the EM algorithm (Dempster *et al.*, 1977), for example, is not feasible. Following McCullagh (1991) and White (1994), we shall therefore refer to $G_{N,L}(\boldsymbol{\theta})$ as the quasi-likelihood and the value minimizing $G_{N,L}(\boldsymbol{\theta})$, $\hat{\boldsymbol{\theta}}_{N,L}$, will be called the Gaussian estimator.

The natural way to obtain $\hat{\boldsymbol{\theta}}_{N,L}$ is to solve the first-order conditions $\partial G_{N,L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$. It is simple to verify that $G_{N,L}(\boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}$ and it follows from the extreme value theorem for continuous functions and the relationship between extreme values and critical points (see White (1994), section 1.2, for example) that, if $\hat{\boldsymbol{\theta}}_{N,L}$ has a neighbourhood $N_\eta(\hat{\boldsymbol{\theta}}_{N,L}) \subset \Theta$, then $\partial G_{N,L}(\hat{\boldsymbol{\theta}}_{N,L})/\partial\boldsymbol{\theta} = \mathbf{0}$ almost surely. The determination of this solution is not straightforward, however, because of the highly non-linear manner in which the parameter $\boldsymbol{\alpha}$ enters the calculations. Nevertheless, given a value of $\boldsymbol{\alpha}$, $\dot{\boldsymbol{\alpha}} = (\dot{\alpha}_1, \ldots, \dot{\alpha}_p)'$, say, closed form analytic expressions for the associated optimizing values of σ^2 and $\mathbf{a}$, $\dot{\sigma}^2$ and $\dot{\mathbf{a}}$, are readily obtained. From the first-order conditions

$$\frac{\partial G_{N,L}(\cdot)}{\partial\sigma^2} = \frac{NL}{2\sigma^2} - \sum_{r=1}^{N}\sum_{t=0}^{L-1}\frac{e_r(t)^2}{2\sigma^4} = 0$$

and

$$\frac{\partial G_{N,L}(\cdot)}{\partial a_r} = -\sum_{t=0}^{L-1}\frac{e_r(t)\,h_\alpha(t)}{\sigma^2} = 0, \qquad r = 1, \ldots, N,$$

it is easily deduced that

$$\dot{\sigma}^2 = (NL)^{-1}\sum_{r=1}^{N}\sum_{t=0}^{L-1}\dot{e}_r(t)^2$$

where

$$\dot{e}_r(t) = y\{t + (r-1)L\} - \dot{a}_r\,h_{\dot{\alpha}}(t), \qquad t = 0, \ldots, L-1, \quad r = 1, \ldots, N,$$

and

$$\dot{a}_r = \sum_{t=0}^{L-1} y\{t + (r-1)L\}\,h_{\dot{\alpha}}(t) \bigg/ \sum_{t=0}^{L-1} h_{\dot{\alpha}}(t)^2, \qquad r = 1, \ldots, N.$$

Substituting these quantities into $G_{N,L}(\cdot)$ and evaluating at the point $\dot{\boldsymbol{\theta}}$ where $\dot{\boldsymbol{\theta}} = \mathbf{f}_{N,L}(\dot{\boldsymbol{\alpha}}) = (\dot{\sigma}^2_\epsilon, \dot{\mathbf{a}}', \dot{\boldsymbol{\alpha}}')'$ gives the quasi-profile likelihood function value

$$\mathcal{G}_{N,L}(\dot{\boldsymbol{\alpha}}) = \frac{NL}{2}\{\log(\dot{\sigma}^2) + 1\}$$

and it is clear that $\hat{\boldsymbol{\theta}}_{N,L}$ is obtained by selecting $\dot{\boldsymbol{\alpha}}$ such that $\mathcal{G}_{N,L}(\dot{\boldsymbol{\alpha}}) \leqslant \mathcal{G}_{N,L}(\boldsymbol{\alpha})$ for all $\boldsymbol{\alpha} \in \mathbb{R}^p(C)$. Reverting to generic rather than specific values, it follows from the construction of the mapping $\mathbf{f}_{N,L}$: $\mathbb{R}^p(C) \to \Theta$ and the chain rule that

$$\frac{\partial \mathcal{G}_{N,L}(\cdot)}{\partial \alpha_j} = \frac{\partial G_{N,L}\{\mathbf{f}_{N,L}(\cdot)\}}{\partial \alpha_j} = -\sum_{r=1}^{N}\sum_{t=0}^{L-1} \frac{e_r(t) a_r\, v_\alpha(t-j)}{\sigma^2} \qquad j = 1, \ldots, N \tag{4}$$

where the derivative process $v_\alpha(t-j) = \partial h_\alpha(t)/\partial \alpha_j$ is obtained by differentiating $H_\alpha(z)\,\alpha(z) = 1$ and satisfies the difference equation

$$v_\alpha(t) + \alpha_1 v_\alpha(t-1) + \ldots + \alpha_p v_\alpha(t-p) = -h_\alpha(t), \qquad t \geqslant 0,$$

with initial conditions $v_\alpha(t) = h_\alpha(t) = 0,\ t < 0$. The equation system $\partial \mathcal{G}_{N,L}(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha} = \mathbf{0}$, which is analogous to the normal equations from a non-linear least squares problem, does not have an explicit closed form solution and must be solved recursively. The determination of this solution forms the focus of attention of the following two sections.

At this point we wish to describe the statistical properties of $\hat{\boldsymbol{\theta}}_{N,L}$. Let $\bar{G}_{N,L}(\boldsymbol{\theta}) = \mathbb{E}_{y|a}[G_{N,L}(\boldsymbol{\theta})]$. Then a simple calculation shows that

$$\bar{G}_{N,L}(\dot{\boldsymbol{\theta}}) = \frac{NL}{2}\log(\dot{\sigma}^2) + \frac{1}{2\dot{\sigma}^2}\left\{ NL\sigma_\epsilon^2 + \sum_{r=1}^{N}\sum_{t=0}^{L-1} \Delta_{\dot{\theta}}(r,\ t,\ L)^2 \right\}.$$

Since $\log(x) + d/x,\ d > 0$, is minimized at $x = d$ it follows that

$$\bar{G}_{N,L}(\boldsymbol{\theta}) \geqslant \frac{NL}{2}\{\log(\overset{*}{\sigma}{}^2) + 1\} = \bar{G}_{N,L}(\overset{*}{\boldsymbol{\theta}}_{N,L})$$

where

$$\overset{*}{\sigma}{}^2 = \sigma_\epsilon^2 + (NL)^{-1} \sum_{r=1}^{N}\sum_{t=0}^{L-1} \Delta_{\overset{*}{\theta}}(r,\ t,\ L)^2,$$

$$\Delta_{\overset{*}{\theta}}(r,\ t,\ L) = s\{t + (r-1)L,\ u\} - \overset{*}{a}_r\, h_{\overset{*}{\alpha}_{N,L}}(t),$$

$$\overset{*}{a}_r = \sum_{t=0}^{L-1} s\{t + (r-1)L,\ u\}\, h_{\overset{*}{\alpha}_{N,L}}(t) \Big/ \sum_{t=0}^{L-1} h_{\overset{*}{\alpha}_{N,L}}(t)^2,$$

and $\overset{*}{\boldsymbol{\alpha}}_{N,L}$ is the value that minimizes $\Sigma_r\, \Sigma_t\, \Delta_{\overset{*}{\theta}}(r,\ t,\ L)^2$. This yields the parameter value $\overset{*}{\boldsymbol{\theta}}_{N,L} = (\overset{*}{\sigma}{}^2,\ \overset{*}{\mathbf{a}}{}',\ \overset{*}{\boldsymbol{\alpha}}{}'_{N,L})'$ that provides the minimum mean-squared discrepancy between the actual response and that predicted by the model and the behaviour of $\hat{\boldsymbol{\theta}}_{N,L}$ is expressed in terms of its relationship to the pseudo-true parameter $\overset{*}{\boldsymbol{\theta}}_{N,L}$. It can be shown that $\hat{\boldsymbol{\theta}}_{N,L} - \overset{*}{\boldsymbol{\theta}}_{N,L} \to \mathbf{0}$ almost surely as $N,\ L \to \infty$ and, if $\{\overset{*}{\mathbf{V}}_{N,L}\}$ denotes a sequence of positive definite matrices such that

$$(NL\overset{*}{\mathbf{V}}_{N,L})^{-1/2}\, \partial G_{N,L}(\overset{*}{\boldsymbol{\theta}}_{N,L})/\partial \boldsymbol{\theta} \overset{\mathbb{D}}{\rightsquigarrow} N(\mathbf{0},\ \mathbf{I}_d),$$

then $(NL)^{1/2}(\hat{\boldsymbol{\theta}}_{N,L} - \overset{*}{\boldsymbol{\theta}}_{N,L})$ has an approximate asymptotic normal distribution with zero mean and covariance matrix $\Sigma(\boldsymbol{\theta}_{N,L}) = (\bar{\mathbf{H}}^*_{N,L} \overset{*}{\mathbf{V}}{}^{-1}_{N,L} \bar{\mathbf{H}}^*_{N,L})^{-1}$ where

$$\bar{\mathbf{H}}^*_{N,L} = (NL)^{-1}\, \partial^2 \bar{G}_{N,L}(\overset{*}{\boldsymbol{\theta}}_{N,L})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' > 0 \qquad \text{almost surely.}$$

For a detailed discussion of different regularity conditions and a rigorous exposition of the type of derivation that will lead to such results the interested reader is referred to White (1994), particularly the statement and proof of theorem 3.12 and theorem 6.2. For current purposes it is sufficient to observe that the behaviour of $\hat{\boldsymbol{\theta}}_{N,L}$ relative to $\overset{*}{\boldsymbol{\theta}}_{N,L}$ parallels that of the maximum likelihood estimator in more traditional settings and these properties provide a basis for the model estimation and inferential procedures developed now.

3. A Gauss–Newton-type algorithm

Since there is no closed form solution to $\partial \mathcal{G}_{N,L}(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha} = \mathbf{0}$ it is natural to consider the use of an iterative scheme in which the solution is placed in a fixed point framework. Applying Newton–Raphson iterations directly gives the sequence of recursive calculations

$$\hat{\boldsymbol{\alpha}}_{N,L}^{(r)} = \hat{\boldsymbol{\alpha}}_{N,L}^{(r-1)} - \left\{ \frac{\partial^2 \mathcal{G}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(r-1)})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \right\}^{-1} \frac{\partial \mathcal{G}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(r-1)})}{\partial \boldsymbol{\alpha}}, \qquad r = 1, 2, \ldots \tag{5}$$

with $\hat{\boldsymbol{\alpha}}_{N,L}^{(0)}$ set before the recursions. The calculations in equation (5) are asymptotically equivalent, however, to the Gauss–Newton-type scheme

$$\hat{\boldsymbol{\alpha}}_{N,L}^{(i)} = \hat{\boldsymbol{\alpha}}_{N,L}^{(i-1)} + \{\mathbf{H}_{\mathcal{G}}^{-1} \mathbf{D}_{\mathcal{G}}\}|_{\boldsymbol{\theta} = \mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i-1)})}, \qquad i = 1, 2, \ldots, \tag{6}$$

where

$$\mathbf{H}_{\mathcal{G}} = \left[\sum_{t=0}^{L-1} \upsilon_\alpha(t-j)\, \upsilon_\alpha(t-k) - \frac{4 \sum_{t=0}^{L-1} \upsilon_\alpha(t-j)\, h_\alpha(t) \sum_{t=0}^{L-1} \upsilon_\alpha(t-k)\, h_\alpha(t)}{\sum_{t=1}^{L-1} h_\alpha(t)^2} \right]_{j,k=1,\ldots,p}$$

and

$$\mathbf{D}_{\mathcal{G}} = \left[\sum_{r=1}^{N} \omega_r \sum_{t=0}^{L-1} e_r(t)\, \upsilon_\alpha(t-j) \right]_{j=1,\ldots,p},$$

$\omega_r = a_r / \Sigma_{s=1}^N a_s^2$. To see this, note that by construction $\partial \mathcal{G}_{N,L}\{\mathbf{f}_{N,L}(\boldsymbol{\alpha})\}/\partial \sigma^2 = \mathbf{0}$ and $\partial \mathcal{G}_{N,L}\{\mathbf{f}_{N,L}(\boldsymbol{\alpha})\}/\partial \mathbf{a} = \mathbf{0}$. The implicit function theorem (see Dieudonné (1960), theorem 10.2.1) therefore implies that

$$\frac{\partial \mathbf{f}'_{N,L}(\cdot)}{\partial \boldsymbol{\alpha}} = \left(- \frac{\partial^2 G_{N,L}\{\mathbf{f}_{N,L}(\cdot)\}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\theta}'_1} \left[\frac{\partial^2 G_{N,L}\{\mathbf{f}_{N,L}(\cdot)\}}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}'_1} \right]^{-1} \vdots \mathbf{I}_p \right), \tag{7}$$

where $\boldsymbol{\theta}_1 = (\sigma^2, \mathbf{a}')'$, and hence that

$$\frac{\partial^2 \mathcal{G}_{N,L}(\cdot)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} = \frac{\partial^2 G_{N,L}\{\mathbf{f}_{N,L}(\cdot)\}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \frac{\partial^2 G_{N,L}\{\mathbf{f}_{N,L}(\cdot)\}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\theta}'_1} \left[\frac{\partial^2 G_{N,L}\{\mathbf{f}_{N,L}(\cdot)\}}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}'_1} \right]^{-1} \frac{\partial^2 G_{N,L}\{\mathbf{f}_{N,L}(\cdot)\}}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\alpha}'}. \tag{8}$$

But

$$\begin{aligned} \frac{\partial G_{N,L}(\cdot)}{\partial \alpha_j \partial \sigma^2} &= \sigma^{-4} \sum_{r=1}^{N} \sum_{t=0}^{L-1} e_r(t) a_r\, \upsilon_\alpha(t-j) \\ &= \sigma^{-4} \sum_{r=1}^{N} \sum_{t=0}^{L-1} [\epsilon\{t + (r-1)L\} + \Delta_\alpha(r, t, L)] a_r\, \upsilon_\alpha(t-j) \end{aligned}$$

for $j = 1, \ldots, p$ and if $p \geqslant p_L$ it follows from the Cauchy–Schwarz inequality, assumption P2 and Jennrich (1969), theorem 2, that the right-hand side is $o(NL)$. In a similar manner we also find that

$$\frac{\partial^2 G_{N,L}(\cdot)}{\partial \alpha_j \partial a_r} = -2\sigma^{-2} a_r \sum_{t=0}^{L-1} \upsilon_\alpha(t-j)\, h_\alpha(t) + o(NL) \qquad \text{almost surely}$$

for $j = 1, \ldots, p$, $r = 1, \ldots, N$,

$$\frac{\partial^2 G_{N,L}(\cdot)}{\partial\alpha_j \partial\alpha_k} = \sigma^{-2} \sum_{r=1}^{N} a_r^2 \sum_{t=0}^{L-1} v_\alpha(t-j)\, v_\alpha(t-k) + o(NL) \qquad \text{almost surely}$$

with j, $k = 1, \ldots, p$ and

$$\frac{\partial^2 G_{N,L}(\cdot)}{\partial a_r \partial a_s} = \delta(r-s)\sigma^{-2} \sum_{t=0}^{L-1} h_\alpha(t)^2 \qquad r, s = 1, \ldots, N.$$

Substituting these values into expression (8) and using equation (4) we conclude, after some straightforward but tedious algebra, that the recursions in equation (5) are the same up to terms $o(1)$ almost surely as those in equation (6). Proofs of the next and subsequent theorems are given in Appendix A.

Theorem 1. Suppose that assumptions P1, P2 and M1–M3 hold and that $p \geqslant p_L$. Let $\mathbf{R}_{N,L}$: $\mathbb{R}^p \to \mathbb{R}^p$ denote the mapping given by the right-hand side of equation (6). Then there is an $\eta > 0$ such that $\hat{\boldsymbol{\alpha}}_{N,L}$ is, with probability 1 as $N, L \to \infty$, the unique fixed point of $\mathbf{R}_{N,L}$ in $N_\eta(\hat{\boldsymbol{\alpha}}_{N,L})$ and if $\hat{\boldsymbol{\alpha}}_{N,L}^{(0)} \in N_\eta(\hat{\boldsymbol{\alpha}}_{N,L})$ then $\hat{\boldsymbol{\alpha}}_{N,L}^{(i)} \to \hat{\boldsymbol{\alpha}}_{N,L}$ almost surely as $i \to \infty$.

Theorem 2. Suppose that the assumptions of theorem 1 obtain and that the iteration scheme is initialized at $\hat{\boldsymbol{\alpha}}_{N,L}^{(0)}$ where $\hat{\boldsymbol{\alpha}}_{N,L}^{(0)} - \overset{*}{\boldsymbol{\alpha}}_{N,L} \to \mathbf{0}$ almost surely as N, $L \to \infty$. Then, with probability 1, the iterates $\hat{\boldsymbol{\alpha}}_{N,L}^{(i)}$ will converge to $\hat{\boldsymbol{\alpha}}_{N,L}$ as $i \to \infty$ and $\|\hat{\boldsymbol{\alpha}}_{N,L}^{(i)} - \overset{*}{\boldsymbol{\alpha}}_{N,L}\| < \eta$, $\eta > 0$, for all $i \geqslant 1$, for N and L sufficiently large. If, in addition, $(NL)^{1/2}(\hat{\boldsymbol{\alpha}}_{N,L}^{(0)} - \overset{*}{\boldsymbol{\alpha}}_{N,L}) = O_p(1)$ then

$$\overset{*}{\mathbf{V}}{}_{N,L}^{-1/2} \bar{\mathbf{H}}_{N,L}^{*} (NL)^{1/2} (\hat{\boldsymbol{\theta}}_{N,L}^{(i)} - \overset{*}{\boldsymbol{\theta}}_{N,L}) \overset{\mathbb{D}}{\rightsquigarrow} N(0, \mathbf{I}_p),$$

where $\hat{\boldsymbol{\theta}}_{N,L}^{(i)} = \mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i)})$ for all $i \geqslant 1$.

The previous results indicate that when appropriately initialized the iterative scheme given in equation (6) will generate a sequence of estimates $\hat{\boldsymbol{\theta}}_{N,L}^{(i)} = \mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i)})$ that will converge to $\hat{\boldsymbol{\theta}}_{N,L}$ and, moreover, the iterates $\hat{\boldsymbol{\theta}}_{N,L}^{(i)}$ will be statistically equivalent to the Gaussian estimator for all $i \geqslant 1$. This suggests that a finite algorithm may be constructed by determining a suitable initial estimate and then conducting only a single iteration of algorithm (6). We shall return to a discussion of this issue in the following section.

4. Construction of initial estimates

Consider dividing the data $y(t)$, $t = 0, \ldots, T-1$, into N disjoint segments of length L, in each of which the observations are taken as the noise-corrupted output produced by a single evoked response. Averaging the different observed responses obtained after each stimulus is applied at the time points $t = 0, L, \ldots, (N-1)/L$ we obtain

$$\bar{y}(t) = \bar{a}\, h_\alpha(t) + \bar{\Delta}_\theta(t) + \bar{\epsilon}(t), \qquad t = 0, \ldots, L-1, \tag{9}$$

where

$$\bar{y}(t) = N^{-1} \sum_{r=1}^{N} y\{t + (r-1)L\}$$

and $\bar{\Delta}_\theta(t)$ and $\bar{\epsilon}(t)$ are similarly defined in an obvious manner, and

$$\bar{a} = N^{-1} \sum_{r=1}^{N} a_r,$$

the average amplitude of the evoked responses. From the strong law of large numbers, however, $\bar{\epsilon}(t) = N^{-1} \Sigma\, \epsilon\{t + (r-1)L\}$ converges to 0 almost surely uniformly in t as $N \to \infty$. Furthermore, since by definition $\alpha(z)\, h_\alpha(z) = 1$ it follows that

$$\alpha(z)\, \bar{y}(t) = \bar{a}\, \delta(t) + \alpha(z)\{\bar{\Delta}_\theta(t) + \bar{\epsilon}(t)\}, \tag{10}$$

in which z^{-1} is interpreted as the unit time delay operator, so that $z^{-1}\, \bar{y}(t) = \bar{y}(t-1)$ etc. This implies that for N sufficiently large

$$\alpha(z)\, \bar{y}(t) = \alpha(z)\, \bar{\Delta}_\theta(t) + o(1) \qquad \text{almost surely, } t = 1, \ldots, L-1,$$

and suggests that we choose as a very first estimate of $\boldsymbol{\alpha}$ the value that minimizes the mean square of $\alpha(z)\, \bar{y}(t)$, $t = 1, \ldots, L-1$, $\bar{\boldsymbol{\alpha}}_{N,L}$, say. It is well known (Rao (1965), sections 1.f.2 and 8.g.2) that this is achieved by setting $\bar{\boldsymbol{\alpha}}_{N,L}$ equal to the appropriately normalized eigenvector associated with the smallest eigenvalue of the matrix $\bar{\mathbf{M}}_p$ of mean squares and cross-products of the variables $\bar{y}(t-j)$, $j = 0, 1, \ldots, p$, i.e. $\bar{\boldsymbol{\alpha}}_{N,L}$ is obtained by scaling the principal component loading vector associated with the smallest principal component of the variables in $(\bar{y}(t), \ldots, \bar{y}(t-p))'$ to have 1 as its first value.

Theorem 3. Assume that the conditions of theorem 1 obtain. Then for any $\eta > 0$ there are an N'_η and L'_η such that for all $N \geqslant N'_\eta$ and $L \geqslant L'_\eta$ the probability that $\|\overset{*}{\boldsymbol{\alpha}}_{N,L} - \bar{\boldsymbol{\alpha}}_{N,L}\| \leqslant \eta$ is arbitrarily large.

In the light of theorem 2 we might therefore consider using $\bar{\boldsymbol{\alpha}}_{N,L}$ as our initial estimate in the recursive algorithm described above. It is possible, however, to exploit the assumed structure of the observed process and the model a little further.

By analogy with equation (9), the sum-of-squares term in the quasi-likelihood function

$$\sum_{r=1}^{N} \sum_{t=0}^{L-1} e_r(t)^2 = \sum_{r=1}^{N} \sum_{t=0}^{L-1} \{e_r(t) - \bar{e}(t)\}^2 + N \sum_{t=0}^{L-1} \bar{e}(t)^2$$

for any value of $\boldsymbol{\theta}$ where

$$\bar{e}(t) = N^{-1} \sum_{r=1}^{N} e_r(t) = \bar{y}(t) - \bar{a}\, h_\alpha(t).$$

Substituting into equation (3) we obtain

$$G_{N,L}(\boldsymbol{\theta}) = \frac{NL}{2} \log(\sigma^2) + \sum_{r=1}^{N} \sum_{t=0}^{L-1} \frac{\{e_r(t) - \bar{e}(t)\}^2}{2\sigma^2} + N \sum_{t=0}^{L-1} \frac{\bar{e}(t)^2}{2\sigma^2}.$$

Now observe that if the amplitudes of the evoked responses are constant $a_r = a = \bar{a}$, $r = 1, \ldots, N$, and the second term simplifies to

$$(2\sigma^2)^{-1} \sum_{r=1}^{N} \sum_{t=0}^{L-1} [y\{t + (r-1)L\} - \bar{y}(t)]^2,$$

which is independent of $\mathbf{a}$ and $\boldsymbol{\alpha}$. In this case, minimization of $G_{N,L}(\boldsymbol{\theta})$ leads to a consideration of the average quasi-profile likelihood function

$$\bar{\mathcal{G}}_{N,L}(\boldsymbol{\alpha}) = \frac{NL}{2}\{\log(\bar{\sigma}^2) + 1\},$$

where $\bar{\sigma}^2 = L^{-1}\,\Sigma_t\,\{\bar{y}(t) - \bar{\bar{a}}\,h_\alpha(t)\}^2$ and $\bar{\bar{a}} = \Sigma_t\,\bar{y}(t)\,h_\alpha(t)/\Sigma_t\,h_\alpha(t)^2$.

Determination of the value $\bar{\bar{\boldsymbol{\alpha}}}_{N,L}$ such that $\bar{\mathcal{G}}_{N,L}(\bar{\bar{\boldsymbol{\alpha}}}_{N,L}) = \min_{\mathbb{R}^p(C)}\{\bar{\mathcal{G}}_{N,L}(\boldsymbol{\alpha})\}$ is, once again, a highly non-linear optimization problem. A similar argument to that developed in Section 3 leads, nevertheless, to a Gauss–Newton-type recursion equivalent to that presented in equation (6) in which the weights $w_r = 1/N\bar{a}$, $r = 1, \ldots, N$. Using $\bar{\boldsymbol{\alpha}}_{N,L}$ in a single iteration of the recursion so derived yields

$$\hat{\boldsymbol{\alpha}}^{(0)}_{N,L} = \bar{\boldsymbol{\alpha}}_{N,L} + \{\mathbf{H}_{\mathcal{G}}^{-1}\bar{\mathbf{D}}_{\mathcal{G}}\}|_{\boldsymbol{\theta}=\mathbf{f}_{N,L}(\bar{\alpha}_{N,L})} \tag{11}$$

where

$$\bar{\mathbf{D}}_{\mathcal{G}} = \left[\bar{a}^{-1}\sum_{t=0}^{L-1}\bar{e}(t)\,v_\alpha(t-j)\right]_{j=1,\ldots,p}.$$

It is this quantity that can now be taken as the initial value. We have continued to use the same notation for the initial estimate despite the fact that it previously indicated a generic quantity whereas now it is being used for the specific value computed as shown. Although somewhat ambiguous this should cause no confusion. The choice of value and notation is justified by the following theorem.

Theorem 4. If the conditions of theorem 3 hold, then as $N \to \infty$ and $L \to \infty$ $\|\hat{\boldsymbol{\alpha}}^{(0)}_{N,L} - \overset{*}{\boldsymbol{\alpha}}_{N,L}\| \to 0$ almost surely.

As previously intimated, a combination of the theoretical results given in the previous section with theorems 3 and 4 indicates that estimates equivalent to $\hat{\boldsymbol{\theta}}_{N,L}$ can be obtained from the following finite algorithm.

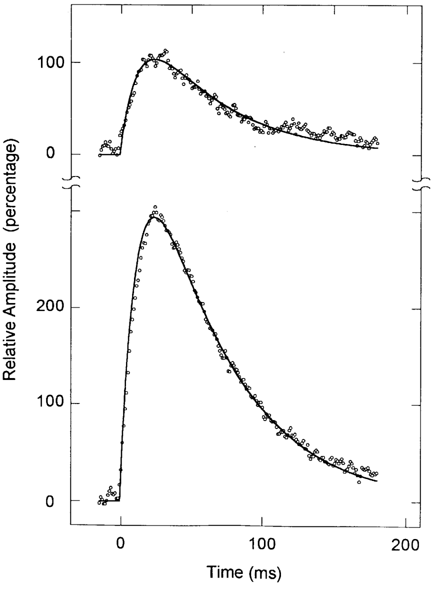

Fig. 3. Impulse response analysis: time profiles of $\bar{y}(t)$ (○) and $h_{\hat{\alpha}^{(1)}_{N,L}}(t)$ (———), $N = 20$ and $L = 400$, for rat hippocampal pyramidal cell post-synaptic currents (the top element relates to observations obtained *ex ante*, and the bottom element to observations obtained *ex post*, long-term potentiation)

Step 1: determine the preliminary estimate $\bar{\boldsymbol{\alpha}}_{N,L}$.
Step 2: calculate the initial value $\hat{\boldsymbol{\alpha}}_{N,L}^{(0)}$ as in equation (11).
Step 3: evaluate the estimate $\hat{\boldsymbol{\alpha}}_{N,L}^{(1)}$ from a single iteration of algorithm (6).
Step 4: take $\hat{\boldsymbol{\theta}}_{N,L}^{(1)} = \mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(1)})$ as the final estimate of $\boldsymbol{\theta}$.

It is this estimate that, for most practical purposes, we would recommend for the basis of subsequent analysis. An application of the algorithm to recordings made on a hippocampal pyramidal neuron from a laboratory rat is illustrated in Fig. 3. The data analysed consisted of two sets of $T = 8000$ observations obtained by sampling the post-synaptic currents at 2 kHz taking the first 20 segments of 200 ms both before and after the induction of long-term potentiation, a process designed to increase the sensitivity of the neuron to the stimulus. Further details are given in Stricker *et al.* (1996). For simplicity the data values are summarized by graphing $\bar{y}(t)$, which is plotted against $h_{\hat{\alpha}_{N,L}^{(1)}}(t)$, $t = 0, \ldots, L-1$. In both cases $\hat{\alpha}_{N,L}^{(1)}(z) = 1.0 - 1.907z^{-1} + 0.908z^{-2} = (1 - 0.982z^{-1})(1 - 0.925z^{-1})$, to three decimal places, and the estimated standard error of $|\hat{\alpha}_{N,L}^{(1)}(z) - \overset{*}{\alpha}_{N,L}(z)|$, $|z| \geqslant 1$, was less than 0.027. A second-order system was chosen using the methodology described in the following section. The ability of the modelling procedure to capture the basic features of the neurobiological transmission mechanism in the two different experimental situations is readily apparent.

5. Model specification

In the analysis of recordings of post-synaptic currents often sufficient information is available about the biophysical process being observed to be able to characterize some of the gross features of the evoked response function generated by a stimulus. This implies that prior knowledge about values of p that will most likely yield an impulse response function that closely approximates the true signal will be at hand. Nevertheless, it seems desirable to be able to test the model specification and to ascertain whether the value of p employed in practice is sufficiently large to ensure that the theoretical results presented above are applicable. In particular, we would like to test the validity of assumption P3 concerning the magnitude of the model approximation error.

As previously, let the moment matrix

$$\bar{\mathbf{M}}_{P_L} = \left[L^{-1} \sum_{t=0}^{L-1} \bar{y}(t-j)\, \bar{y}(t-k) \right] \qquad j, k = 0, 1, \ldots, P_L$$

where $\bar{y}(t)$ is defined as in equation (9) and $P_L = O(p_L)$, $P_L > p_L$, is an upper bound of the order of $\alpha(z)$ chosen by the practitioner. The value p_L is assumed to be such that the bound on the mean-squared model approximation error given in assumption P3 is applicable. Let $r_1 \geqslant \ldots \geqslant r_{P_L}$ denote the eigenvalues of $\bar{\mathbf{M}}_{P_L}$.

Theorem 5. Suppose that assumptions M1–M3 obtain and that assumptions P1–P3 hold. Then under the null hypothesis H_0: $p \geqslant p_L$ the test statistic

$$L \left\{ \sum_{i=p}^{P_L} (r_i - \gamma_{N,L}) \right\}^2 \Big/ 2 \sum_{i=p}^{P_L} r_i^2$$

converges in distribution to a χ^2-variate with 1 degree of freedom, $\chi^2(1)$, as $N, L \to \infty$, $\gamma_{N,L} = 2\hat{\sigma}_\epsilon^2 p([\log\{\log(L)\}/L]^{1/2} + [\log\{\log(N)\}/N]^{1/2})^2$.

An important point to note concerning this result is that the statistic can not only be employed to test a specific value of p but may also be used to determine an adequate yet parsimonious specification. Thus, if significantly large values of the statistic lead to rejection of H_0: $p \geqslant p_L$, then we may consider testing the sequence of values $p = 1, 2, \ldots$, and selecting for p the smallest value such that the hypothesis H_0: $p \geqslant p_L$ is not rejected.

6. Inference

In the analysis of evoked synaptic currents many questions of scientific interest relate to the properties of the transmission mechanism and the unobserved amplitudes. If such questions can be expressed in the form of J explicit functional relationships or restrictions as $\phi(\boldsymbol{\theta}) = \mathbf{0}$ then the application of the Wald test principle in conjunction with theorem 2 would lead to a consideration of the test statistic

$$\frac{\partial\phi(\hat{\boldsymbol{\theta}}_{N,L})}{\partial\boldsymbol{\theta}}\{\Sigma(\boldsymbol{\theta}_{N,L})\}^{-1}\frac{\partial\phi(\hat{\boldsymbol{\theta}}_{N,L})}{\partial\boldsymbol{\theta}}.$$

Values of the statistic greater than $\chi^2_\alpha(J)$, the $100(1-\alpha)\%$ quantile point of a $\chi^2(J)$ random variable, would provide a locally most powerful critical region for testing the null hypothesis H_0: $\phi(\boldsymbol{\theta}) = \mathbf{0}$. We shall not go into details here but refer the interested reader to White (1994), chapter 8. We have in mind at this point parametric hypotheses that relate to the structure of $H_\alpha(z)$ and specify the behaviour of the current flow and/or the filter properties of the dendrite.

Alternatively, if the focus of attention centres on $\mathbb{P}_A$, the estimates $\hat{a}_{N,L(r)}$, $r = 1, \ldots, N$, may be taken as noisy observations on the unknown a_r, $r = 1, \ldots, N$, and examined directly. Given the degree of uncertainty surrounding $\mathbb{P}_A$ it is desirable to impose as little *a priori* structure as possible and the approach that we consider here is to use kernel smoothing methods (see Silverman (1986) and Wand and Jones (1995)) as a basis for making inferences about the amplitude distribution. More specifically, taking account of the fact that $\mathbb{P}_A$ will be convolved with the error distribution to give the distribution $\mathbb{P}_{\hat{A}}$ of the amplitude estimates $\hat{a}$, the deconvolving kernel density estimate due to Stefanski and Carroll (1990) can be employed:

$$\hat{p}_A(a) = (N\omega_N)^{-1}\sum_{r=1}^{N} K_{dc}\left(\frac{\hat{a}_{N,L(r)} - a}{\omega_N}, \omega_N\right),$$

$$K_{dc}(x, b) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-\mathrm{i}tx)\frac{\varphi_K(t)}{\varphi_{(\hat{a}-a)}(t/b)}\,\mathrm{d}t,$$

where $\varphi_K(t)$ is the characteristic function of the basic kernel $K(x) \geqslant 0$, $\int K(x)\,\mathrm{d}x = 1$, $\varphi_{(\hat{a}-a)}(t)$ is the characteristic function of the error process $\hat{a}_{N,L(r)} - a_r$, $r = 1, \ldots, N$, and ω_N is a bandwidth parameter chosen such that $\omega_N \to 0$ as $N \to \infty$, $N\omega_N \to \infty$. From theorem 2 it follows that for N and L large the errors will be approximately Gaussian with mean 0 and variance $O\{(NL)^{-1}\}$ and theorem 5 of Fan (1992) implies that $\mathbb{P}_A$ can therefore be estimated to a level of accuracy of the order $O(N^{-1/2})$, despite the fact that convergence rates for deconvolving densities are known in general to be $\{\log(N)\}^{-1}$ at best; see Carroll and Hall (1988). Consequently, the estimation of $\mathbb{P}_A$ may not be appreciably more difficult than ordinary density estimation. Indeed, the continuity property of characteristic functions (Billingsley (1986), theorem 26.3) implies that

$$\varphi_{(\hat{a}-a)}(t) = \exp[-t^2\,O\{(NL)^{-1}\}] + o(1)$$

and it is easily deduced that

$$K_{dc}(x, \omega_N) = K(x)[1 + O\{(NL\omega_N^2)^{-1}\} + o(1)].$$

Thus, for N and L large little, if anything, may be lost by applying the kernel $K(x)$ directly.

Suppose that the support of $\mathbb{P}_A$ is discrete with n_a quantal values $a(i) < a(i+1)$ with probability of occurrence $p(i) = \Pr\{a = a(i)\}$, $i = 1, \ldots, n_a$. As just indicated, for N and L sufficiently large the errors $\Delta a = \hat{a} - a$ may be treated as $N(0, \sigma^2_{\Delta a})$ random variates with variance $\sigma^2_{\Delta a} \propto 1/NL$. The convolution integral then implies that $\mathbb{P}_{\hat{A}}$ can be closely approximated by

$$\sum_{i=1}^{n_a} p(i)\, \Phi\left\{\frac{a - a(i)}{\sigma_{\Delta a}}\right\}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. Hence if $\mathbb{P}_A$ is discrete and NL is sufficiently large to ensure that $\sigma^2_{\Delta a}$ is small in relation to the quantal separation $\min_{1 \leqslant i < n_a} a(i+1) - a(i)\}$ then we can expect $\hat{p}_A(a)$ to resolve into a mixture of n_a normal densities concentrated at the quantal values $a(i)$ with ordinates proportional to $p(i)$. Such behaviour is clearly observed in the results presented immediately below.

7. Empirical results

In this section we shall illustrate the application of the methodology in the context of a simulated example designed to indicate the type of data processing that might be encountered in the analysis of post-synaptic neural currents. Our purpose is to ascertain the behaviour and properties of the techniques in a known situation thereby avoiding the necessity to disentangle questions concerning the performance and practical relevance of the methods from the scientific uncertainties that are likely to be present with real world data.

Figs 4 and 5 display results obtained by applying the techniques to two data-generating mechanisms that conform with the structure of the model. Namely, $\{s(t, u)\}$ was generated by passing an input process of the type given in equation (2) through the transfer function $\alpha(z)^{-1}$, $\alpha(z) = (1 - 0.97z^{-1})(1 - 0.81z^{-1})$. A second-order system was used to mimic the convolution of exponential decay in the total current flux and the low pass filter properties of the dendrite, the parameter values being chosen to approximate those of real signals recorded from hippocampal pyramidal cells. The probability measure $\mathbb{P}_A$ associated with $\{u(t)\}$ was chosen to represent the two extreme cases of interest in the quantal analysis of post-synaptic currents. The discrete distribution

$$\mathbb{P}_A(a = a_n) = \exp(-2.1)(2.1)^n \Bigg/ \frac{n! \sum_{r=0}^{5} \exp(-2.1)(2.1)^r}{r!}$$

$$a_n = (0)(0.771)(3.855), \qquad n = 0, \ldots, 5,$$

for the first data-generating mechanism and the mixed mass density function

$$\mathbb{P}_A(a = 0) = 0.2,$$

$$\mathbb{P}_A(a \leqslant a') = 0.2 + 0.8\{1 - \exp(-a'^2)\}, \qquad a' > 0,$$

for the second. These two choices of $\mathbb{P}_A$ are motivated by the neurobiological question of whether amplitude fluctuations brought about by variations in the packets of neurotrans-

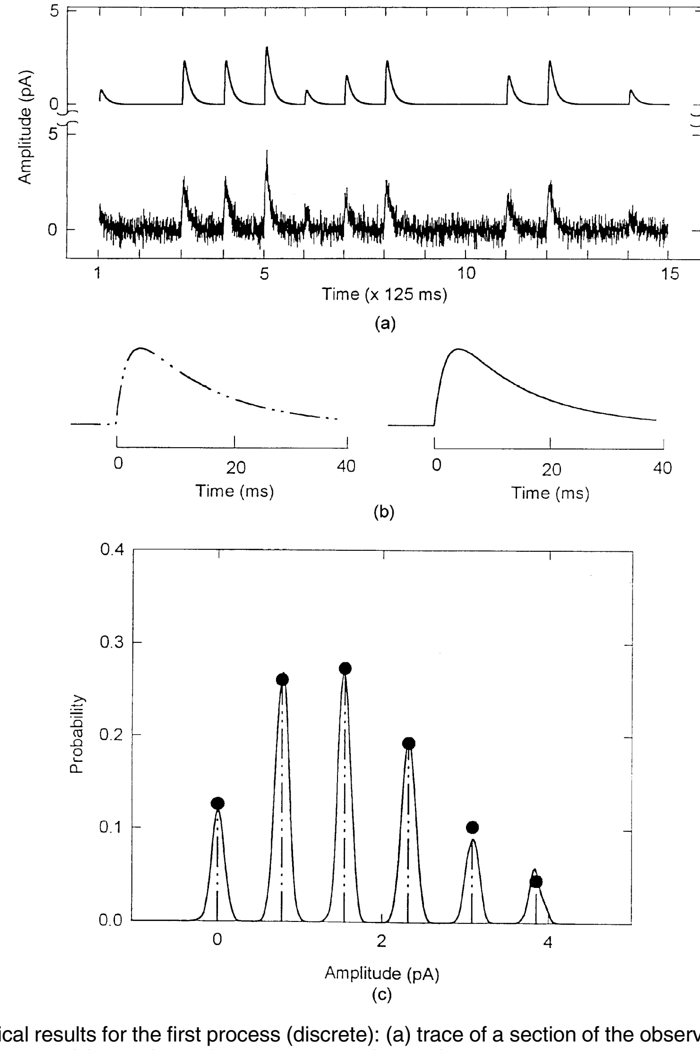

Fig. 4. Empirical results for the first process (discrete): (a) trace of a section of the observed series preceded by the underlying signal; (b) true (·· — ··) and estimated (———) impulse response function; (c) true (·· — ··) and estimated (———) amplitude distribution (•, probability mass) (see the text for further details)

mitter molecules are characterized by a probability distribution with countable support, as in the classical quantal model of synaptic currents, or are continuous; see Clements (1991). The observations are constructed by adding independent and identically distributed $N(0, \sigma_\epsilon^2)$ random disturbances to $\{s(t, u)\}$ where $\sigma_\epsilon = 0.35$ for the first process and 0.7 for the second; σ_ϵ is a little under half the quantal separation of $\mathbb{P}_A$ in the discrete case and just over a quarter of the modal value of the positive responses (non-failures) in the continuous case. These values yield similar signal-to-noise ratios and correspond to magnitudes of the noise variance at which the resolution of current methods of quantal analysis will break down; see Stricker and Redman (1994).

Figs 4(a) and 5(a) present a trace of a section of the observed series preceded by the underlying signal over a sequence of 14 stimuli. Figs 4(b) and 5(b) plot the actual and estimated impulse response function. The estimates are based on $T = 250000$ data points

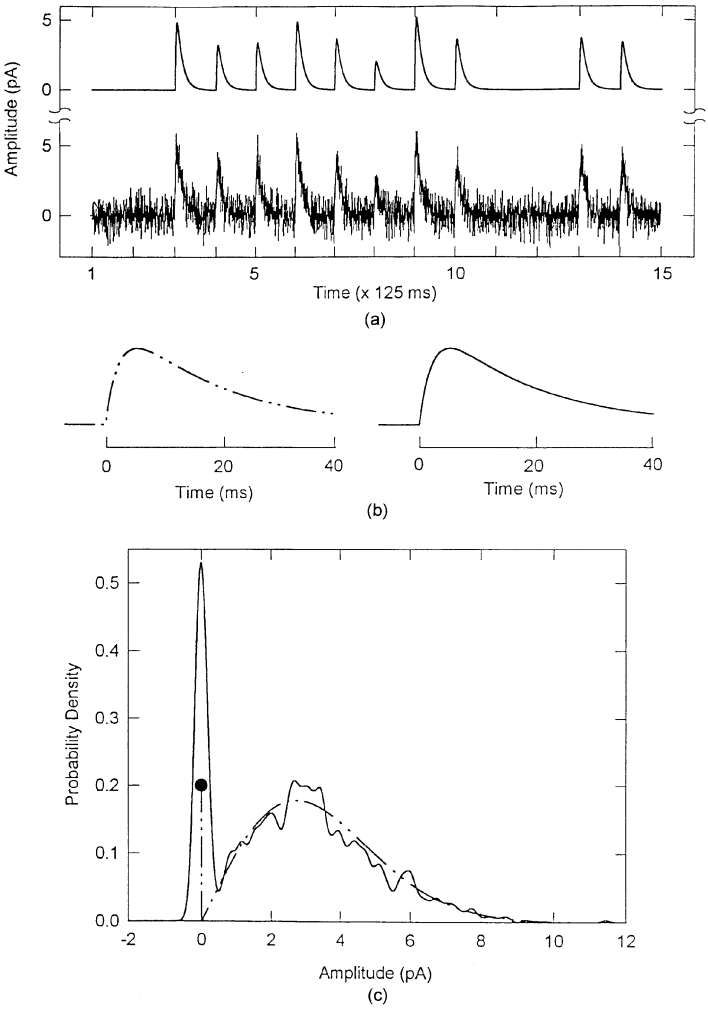

Fig. 5. Empirical results for the second process (continuous): (a) trace of a section of the observed series preceded by the underlying signal; (b) true (·· — ··) and estimated (———) impulse response function; (c) true (·· — ··) and estimated (———) amplitude distribution (see the text for further details)

with $L = 250$ and $N = 1000$. In the analysis of post-synaptic currents, data values are available at rates exceeding 2 kHz; at this rate the figures for L and N correspond to observing the post-synaptic response of a cell that is being stimulated every 125 ms for 125 s $\approx$ 2 min. Figs 4(c) and 5(c) show a graph of $\mathbb{P}_A$ with the kernel density estimate

$$(N\omega_N)^{-1} \sum_{i=1}^{N} K\left(\frac{\hat{a}^{(1)}_{N,L(i)} - a}{\omega_N}\right),$$

where $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ and ω_N is given by the data-based optimal bandwidth selection rule of Sheather and Jones (1991), superimposed. For Fig. 4, $\omega_N = 0.041$ and, for Fig. 5, $\omega_N = 0.1169$.

The ability of the technique to detect the salient features of the data-generating process is

Table 1. Parameter estimates†

	$\bar{\boldsymbol{\alpha}}_{N,L}$	$\hat{\boldsymbol{\alpha}}^{(0)}_{N,L}$	$\hat{\boldsymbol{\alpha}}^{(1)}_{N,L}$
Process one (discrete)			
$\alpha_1 = -1.78$	−1.7889	−1.7841	−1.7841
$\alpha_2 = 0.7857$	0.7946	0.7897	0.7897
$\sigma_\epsilon = 0.35$		$\hat{\sigma}^{(1)}_{N,L} = 0.3491$	
Process two (continuous)			
$\alpha_1 = -1.78$	−1.769	−1.7743	−1.7742
$\alpha_2 = 0.7857$	0.7747	0.7801	0.78
$\sigma_\epsilon = 0.7$		$\hat{\sigma}^{(1)}_{N,L} = 0.6975$	

† $N = 1000,\ L = 250$.

Table 2. Mean-squared error $\hat{\boldsymbol{\alpha}}^{(1)}_{N,L}$†

	NL(bias)(bias)′		*NL(variance)*	
Theoretical	0	0	2.8748	−2.9536
	0	0	−2.9536	3.0346
Empirical	0.0534	−0.0521	2.8447	−2.7805
	−0.0521	0.0508	−2.7805	2.7165

† $N = 1000,\ L = 250$.

clear and is further illustrated by the comparison between the actual and estimated parameter values given in Table 1, which indicates that the convergence properties presented above hold reasonably well with sample sizes of this magnitude. The standard errors of all entries in Table 1 are less than 3.4×10^{-3}. Table 2 presents the empirical mean-squared error of $\hat{\boldsymbol{\alpha}}^{(1)}_{N,L}$ together with the asymptotic values derived from theorem 2. Once again the proximity of the realized values to the asymptotic quantities given in our theoretical results is apparent.

In the examples presented here the presence or absence of quanta in $\mathbb{P}_A$ is readily ascertained with little ambiguity by simple visual inspection. There may be cases, however, where it may be unclear whether the appearance of a local mode (or bump) in $\hat{p}_A(a)$ is due to sampling fluctuation or is to be attributed to a genuine quantal level. One approach to this problem might be to employ the procedures for detecting the number of underlying modes in a distribution suggested in Silverman (1986). An alternative would be to use current techniques of quantal analysis based on mixture deconvolution with unknown number and type of parent distribution; see Stricker and Redman (1994) once again and the references contained therein. Another possibility, and one that is currently under investigation by the authors, is to adapt the method of spacings as discussed by Roeder (1992). What is the most effective approach to take in such situations seems, however, to be an open question.

The figures and tables given above constitute a small sample from a much wider range of simulations designed to investigate the effect of various forms of model and assumption violation, different signal-to-noise ratios etc. that we present elsewhere; see Chung *et al.* (1998). Experience with the analysis of real data gained at the time of writing also indicates that it may be necessary to adjust our methodology to take on board such features as the time delay between stimulation of the nerve fibre and activation of the excitatory synaptic contacts and more complicated impulse response functions than those considered above. Nevertheless, the outcomes reported here are representative and indicate the efficacy of the techniques described in the paper for analysing data obtained from evoked synaptic response experiments.

Acknowledgements

This work was supported in part by a grant from the National Health and Medical Research Council of Australia. We wish to thank Professor Steve Redman and Dr Christian Stricker for allowing us to examine their unpublished synaptic current recordings obtained from hippocampal pathways. We are also grateful to the Joint Editor and the referees for helpful and constructive comments on previous versions of this paper; these have resulted in the correction of errors, the clarification of ideas and significant improvements in presentation.

Appendix A

A.1. Proof of theorem 1

Since the quadratic form

$$\mathbf{x}'\{\partial^2 \bar{G}_{N,L}(\overset{*}{\boldsymbol{\theta}}_{N,L})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\}\mathbf{x} > 0$$

for all $\mathbf{x}$ almost surely, $(NL)^{-1}\partial^2 G_{N,L}(\overset{*}{\boldsymbol{\theta}}_{N,L})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ will be positive definite for N and L sufficiently large since

$$(NL)^{-1}\partial^2\{G_{N,L}(\boldsymbol{\theta}) - \bar{G}_{N,L}(\boldsymbol{\theta})\}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' \rightarrow \mathbf{0} \qquad \text{almost surely}$$

uniformly on $N_\eta(\overset{*}{\boldsymbol{\theta}}_{N,L})$. Hence $\mathbf{H}_\mathcal{G}$ is non-singular in $N_{2\eta}(\overset{*}{\boldsymbol{\alpha}}_{N,L})$ for some $\eta > 0$ and since $\|\hat{\boldsymbol{\alpha}}_{N,L} - \overset{*}{\boldsymbol{\alpha}}_{N,L}\| \rightarrow 0$ almost surely $N_\eta(\hat{\boldsymbol{\alpha}}_{N,L}) \subset N_{2\eta}(\overset{*}{\boldsymbol{\alpha}}_{N,L})$, so $\mathbf{R}_{N,L}$ is well defined for N and L sufficiently large. Now suppose that $\mathbf{R}_{N,L}$ satisfies the Lipschitz condition $\|\mathbf{R}_{N,L}(\boldsymbol{\alpha}_1) - \mathbf{R}_{N,L}(\boldsymbol{\alpha}_2)\| \leqslant \mu\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|$, $0 < \mu < 1$, almost surely as $N, L \rightarrow \infty$, on $N_\eta(\overset{*}{\boldsymbol{\alpha}}_{N,L})$. By definition $\hat{\boldsymbol{\theta}}_{N,L} = \arg\min\{G_{N,L}(\boldsymbol{\theta})\}$ and $\partial G_{N,L}(\hat{\boldsymbol{\theta}}_{N,L})/\partial\boldsymbol{\theta} = \mathbf{0}$ almost surely, from which we can conclude that $\hat{\boldsymbol{\alpha}}_{N,L} = \mathbf{R}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L})$ and the result of the theorem follows from the contraction mapping principle and application of the method of successive approximations (Dieudonné (1960), pages 260–261).

To verify that the supposition is valid observe that $\mathbf{H}_\mathcal{G}$ and $\mathbf{D}_\mathcal{G}$ are both continuously differentiable with respect to $\boldsymbol{\alpha}$ and therefore, from the mean value theorem,

$$\|\mathbf{R}_{N,L}(\boldsymbol{\alpha}_1) - \mathbf{R}_{N,L}(\boldsymbol{\alpha}_2)\| \leqslant \|\frac{\partial\mathbf{R}_{N,L}(\bar{\boldsymbol{\alpha}})}{\partial\boldsymbol{\alpha}'}\|\,\|(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)\|,$$

$\bar{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_1 + \lambda(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)$, $0 \leqslant \lambda \leqslant 1$. By definition of $\mathbf{R}_{N,L}$

$$\frac{\partial\mathbf{R}_{N,L}}{\partial\boldsymbol{\alpha}'} = \mathbf{I}_p + \mathbf{H}_\mathcal{G}^{-1}\,\frac{\partial\mathbf{D}_\mathcal{G}}{\partial\boldsymbol{\alpha}'} - (\mathbf{H}_\mathcal{G}^{-1}\mathbf{D}_\mathcal{G} \otimes \mathbf{H}_\mathcal{G}^{-1})'\,\frac{\partial\{\text{vec}(\mathbf{H}_\mathcal{G})\}}{\partial\boldsymbol{\alpha}'}.$$

Employing a similar argument to that leading to the definition of $\mathbf{R}_{N,L}$ in equation (6), however, we find that $\partial\mathbf{D}_\mathcal{G}/\partial\boldsymbol{\alpha}' \rightarrow -\mathbf{H}_\mathcal{G}$ almost surely for $N, L \rightarrow \infty$ and by continuity $\|\mathbf{D}_\mathcal{G}\|$ can almost surely be made arbitrarily small in $N_\eta(\hat{\boldsymbol{\alpha}}_{N,L})$ by taking η sufficiently small since $\partial G_{N,L}(\hat{\boldsymbol{\theta}}_{N,L})/\partial\boldsymbol{\theta} = \mathbf{0}$ almost surely. Thus we can conclude that $0 \leqslant \|\partial\mathbf{R}_{N,L}(\bar{\alpha})/\partial\boldsymbol{\alpha}'\| < 1$ almost surely for N and L sufficiently large when $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in N_\eta(\hat{\boldsymbol{\alpha}}_{N,L})$ and the supposition is verified, thereby completing the proof of the theorem.

A.2. Proof of theorem 2

By assumption, for any $\eta' > 0$, $\hat{\boldsymbol{\alpha}}^{(0)}_{N,L} \in N_{\eta'}(\overset{*}{\boldsymbol{\alpha}}_{N,L})$ almost surely for all $N > N_{\eta'}$ and $L > L_{\eta'}$. The triangular inequality and the convergence of $\hat{\boldsymbol{\alpha}}_{N,L} - \overset{*}{\boldsymbol{\alpha}}_{N,L}$ to 0 then imply that $\hat{\boldsymbol{\alpha}}^{(0)}_{N,L} \in N_\eta(\hat{\boldsymbol{\alpha}}_{N,L})$ almost surely, $N > N_{\eta'}$ and $L > L_{\eta'}$, $\eta' = \eta/2$, and therefore $\hat{\boldsymbol{\alpha}}^{(i)}_{N,L} \rightarrow \hat{\boldsymbol{\alpha}}_{N,L}$ almost surely as $i \rightarrow \infty$ by theorem 1. Similarly, since for all N and L sufficiently large $\mathbf{R}_{N,L}$ is a contraction with fixed point $\hat{\boldsymbol{\alpha}}_{N,L}$,

$$\begin{aligned}\|\hat{\boldsymbol{\alpha}}_{N,L}^{(i)} - \overset{*}{\boldsymbol{\alpha}}_{N,L}\| &\leqslant \|\hat{\boldsymbol{\alpha}}_{N,L}^{(i)} - \hat{\boldsymbol{\alpha}}_{N,L}^{(i-1)}\| + \|\hat{\boldsymbol{\alpha}}_{N,L}^{(i-1)} - \overset{*}{\boldsymbol{\alpha}}_{N,L}\| \\ &\leqslant (\mu^{(i-1)} + \ldots + \mu + 1)\|\hat{\boldsymbol{\alpha}}_{N,L}^{(1)} - \hat{\boldsymbol{\alpha}}_{N,L}^{(0)}\| + \|\hat{\boldsymbol{\alpha}}_{N,L}^{(0)} - \overset{*}{\boldsymbol{\alpha}}_{N,L}\| \\ &\leqslant \{1/(1-\mu)\}(\|\hat{\boldsymbol{\alpha}}_{N,L}^{(1)} - \hat{\boldsymbol{\alpha}}_{N,L}\| + \|\hat{\boldsymbol{\alpha}}_{N,L} - \hat{\boldsymbol{\alpha}}_{N,L}^{(0)}\|) + \|\hat{\boldsymbol{\alpha}}_{N,L}^{(0)} - \overset{*}{\boldsymbol{\alpha}}_{N,L}\| \\ &\leqslant \{(1+\mu)/(1-\mu)\}\|\hat{\boldsymbol{\alpha}}_{N,L}^{(0)} - \hat{\boldsymbol{\alpha}}_{N,L}\| + \|\hat{\boldsymbol{\alpha}}_{N,L}^{(0)} + \overset{*}{\boldsymbol{\alpha}}_{N,L}\| \\ &\leqslant \eta \qquad \text{almost surely}\end{aligned}$$

for $\eta = \eta(1-\mu)/2$, $0 \leqslant \mu < 1$, which completes the proof of the first statement of the theorem.
Now suppose that $(NL)^{1/2}(\hat{\boldsymbol{\theta}}_{N,L}^{(i)} - \overset{*}{\boldsymbol{\theta}}_{N,L}) = O_p(1)$. Then

$$\hat{\boldsymbol{\theta}}_{N,L}^{(i+1)} - \hat{\boldsymbol{\theta}}_{N,L}^{(i)} = \mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i+1)}) - \mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i)})$$

and

$$\mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i+1)}) = \mathbf{f}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i)}) + \frac{\partial \mathbf{f}_{N,L}(\bar{\boldsymbol{\alpha}}_{N,L}^{(i)})}{\partial \boldsymbol{\alpha}'}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i+1)} - \hat{\boldsymbol{\alpha}}_{N,L}^{(i)})$$

where $\bar{\boldsymbol{\alpha}}_{N,L}^{(i)} = \hat{\boldsymbol{\alpha}}_{N,L}^{(i+1)} + \lambda(\hat{\boldsymbol{\alpha}}_{N,L}^{(i+1)} - \hat{\boldsymbol{\alpha}}_{N,L}^{(i)})$, $0 \leqslant \lambda \leqslant 1$. Substituting for $\hat{\boldsymbol{\alpha}}_{N,L}^{(i+1)} - \hat{\boldsymbol{\alpha}}_{N,L}^{(i)}$ from expression (5) and for $\partial \mathbf{f}_{N,L}(\bar{\boldsymbol{\alpha}}_{N,L}^{(i)})/\partial \boldsymbol{\alpha}'$ from equation (7) we find that

$$\hat{\boldsymbol{\theta}}_{N,L}^{(i+1)} - \hat{\boldsymbol{\theta}}_{N,L}^{(i)} = \begin{pmatrix} \left(\dfrac{\partial^2 G_{N,L}(\bar{\boldsymbol{\theta}}_{N,L}^{(i)})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'}\right)^{-1} \left(\dfrac{\partial^2 G_{N,L}(\bar{\boldsymbol{\theta}}_{N,L}^{(i)})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\alpha}}\right)\left(\dfrac{\partial^2 \mathcal{G}_{N,L}(\bar{\boldsymbol{\alpha}}_{N,L}^{(i)})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}\right)^{-1} \\ -\left(\dfrac{\partial^2 \mathcal{G}_{N,L}(\bar{\boldsymbol{\alpha}}_{N,L}^{(i)})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}\right)^{-1} \end{pmatrix} \frac{\partial \mathcal{G}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i)})}{\partial \boldsymbol{\alpha}}$$

where

$$\bar{\boldsymbol{\theta}}_{N,L}^{(i)} = \mathbf{f}_{N,L}(\bar{\boldsymbol{\alpha}}_{N,L}^{(i)}).$$

Since $G_{N,L}(\cdot)$ has continuous first and second derivatives, however, $\partial \mathbf{f}_{N,L}(\bar{\boldsymbol{\alpha}}_{N,L}^{(i)})/\partial \boldsymbol{\alpha}' = \partial \mathbf{f}_{N,L}(\overset{*}{\boldsymbol{\alpha}}_{N,L})/\partial \boldsymbol{\alpha}'$ $+O\{(NL)^{-1/2}\}$ on $N_\eta(\overset{*}{\boldsymbol{\alpha}}_{N,L})$. Similarly, if $\hat{\boldsymbol{\alpha}}_{N,L}^{(i)} \in N_\eta(\overset{*}{\boldsymbol{\alpha}}_{N,L})$ then

$$\frac{\partial^2 \mathcal{G}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i)})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} = \frac{\partial^2 \mathcal{G}_{N,L}(\overset{*}{\boldsymbol{\alpha}}_{N,L})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} + O\{(NL)^{-1/2}\}.$$

Using standard formulae for partitioned inversion in conjunction with equation (8), recognizing that by construction

$$\partial G_{N,L}(\hat{\boldsymbol{\theta}}_{N,L}^{(i)})/\partial \boldsymbol{\theta} = \begin{pmatrix} \mathbf{0} \\ \partial \mathcal{G}_{N,L}(\hat{\boldsymbol{\alpha}}_{N,L}^{(i)})/\partial \boldsymbol{\alpha} \end{pmatrix},$$

leads to the conclusion that

$$\hat{\boldsymbol{\theta}}_{N,L}^{(i+1)} - \hat{\boldsymbol{\theta}}_{N,L}^{(i)} = -(\mathbf{H}_{N,L}^{*-1})(NL)^{-1} \frac{\partial G_{N,L}(\hat{\boldsymbol{\theta}}_{N,L}^{(i)})}{\partial \boldsymbol{\theta}} + o\{(NL)^{-1/2}\}$$

where $\mathbf{H}_{N,L}^* = \partial^2 G_{N,L}(\overset{*}{\boldsymbol{\alpha}}_{N,L})/\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'$. Finally, expanding $\partial G_{N,L}(\hat{\boldsymbol{\theta}}_{N,L}^{(i)})/\partial \hat{\boldsymbol{\theta}}$ about $\partial G_{N,L}(\overset{*}{\boldsymbol{\theta}}_{N,L})/\partial \boldsymbol{\theta}$ and rearranging terms then yields

$$\begin{aligned}(NL)^{1/2}(\hat{\boldsymbol{\theta}}_{N,L}^{(i+1)} - \overset{*}{\boldsymbol{\theta}}_{N,L}) &= (\mathbf{I} - \mathbf{H}_{N,L}^{*-1}[\mathbf{H}_{N,L}^* + O\{(NL)^{-3/2}\}])(NL)^{1/2}(\hat{\boldsymbol{\theta}}_{N,L}^{(i)} - \overset{*}{\boldsymbol{\theta}}_{N,L}) \\ &\quad + \mathbf{H}_{N,L}^{*-1}(NL)^{-1/2} \frac{\partial G_{N,L}(\overset{*}{\boldsymbol{\theta}}_{N,L})}{\partial \theta} + o(1),\end{aligned}$$

and hence $(NL)^{1/2}(\hat{\boldsymbol{\theta}}_{N,L}^{(i+1)} - \overset{*}{\boldsymbol{\theta}}_{N,L})$ converges in distribution when appropriately scaled because the first term on the right-hand side is $o(1)$. The last statement of the theorem now follows by induction.

A.3. Proof of theorem 3

From expression (10) and the Cauchy–Schwarz inequality

$$L^{-1}\sum_{t=1}^{L-1}\{\alpha(z)\,\bar{y}(t)\}^2 \leqslant L^{-1}\sum_{t=1}^{L-1}\left(\sum_{j=0}^{p}\alpha_j^2\right)p\{\bar{\Delta}_\theta(t)+\bar{\epsilon}(t)\}^2$$
$$\leqslant \left(\sum_{j=0}^{p}\alpha_j^2\right)p(NL)^{-1}\left[\left\{\sum_{t=0}^{L-1}\sum_{r=1}^{N}\Delta_\theta(t,\,r,\,L)^2\right\}^{1/2}+\left\{N\sum_{t=0}^{L-1}\bar{\epsilon}(t)^2\right\}^{1/2}\right]^2. \qquad (12)$$

Assumption P3 and the law of the iterated logarithm applied to $\bar{\epsilon}(t)$ then imply that the smallest principal component of $\bar{\mathbf{M}}_p$ is bounded above almost surely by the value

$$2\sigma_\epsilon^2 p([\log\{\log(L)\}/L]^{1/2}+[\log\{\log(N)\}/N]^{1/2})^2$$

as $L,\,N\to\infty$. Suppose, therefore, that N and L are sufficiently large to guarantee that the bound is less than η' and let $N_\eta(\bar{\boldsymbol{\alpha}}_{N,L})$ denote the neighbourhood of $\bar{\boldsymbol{\alpha}}_{N,L}$ such that

$$\|\boldsymbol{\alpha}\|^{-2}L^{-1}\sum\{\alpha(z)\,\bar{y}(t)\}^2<\eta'.$$

If $N_\eta(\overset{*}{\boldsymbol{\alpha}}_{N,L})$ denotes the set of values of $\boldsymbol{\alpha}$ such that the corresponding factor

$$p(NL)^{-1}\left[\left\{\sum_{t=0}^{L-1}\sum_{r=1}^{N}\Delta_\theta(t,\,r,\,L)^2\right\}^{1/2}+\left\{N\sum_{t=0}^{L-1}\bar{\epsilon}(t)^2\right\}^{1/2}\right]^2$$

is less than η' then the above inequality implies that $N_\eta(\overset{*}{\boldsymbol{\alpha}}_{N,L})\subseteq N_\eta(\bar{\boldsymbol{\alpha}}_{N,L})$, giving the required result.

A.4. Proof of theorem 4

Theorem 4 is established by using a combination of the arguments employed in the derivations of theorems 1 and 2. The detailed steps, which closely parallel those given above, are omitted to save unnecessary repetition.

A.5. Proof of theorem 5

It follows from inequality (12) and Rao (1965), section 8g.2, that

$$r_p \leqslant \|\bar{\boldsymbol{\alpha}}_{N,L}\|^{-2}L^{-1}\sum_{t=0}^{L-1}\{\bar{\alpha}_{N,L}(z)\,\bar{y}(t)\}^2$$
$$\leqslant 2\sigma_\epsilon^2 p_L([\log\{\log(L)\}/L]^{1/2}+[\log\{\log(N)\}/N]^{1/2})^2 \qquad \text{almost surely}$$

for $p_L\leqslant p\leqslant P_L$ when N and L are sufficiently large. By definition $\bar{y}(t)=\bar{s}(t,\,u)+\bar{\epsilon}(t)$ where

$$\bar{s}(t,\,u)=N^{-1}\sum_r s\{t+(r-1)L,\,u\}$$

and

$$\bar{\epsilon}(t)=N^{-1}\sum_r \epsilon\{t+(r-1)L\}.$$

Standard central limit theorems for independent and identically distributed random variables and alpha mixing processes applied to $\bar{\epsilon}(t)$ and $\bar{s}(t,\,u)$ respectively, using assumptions P1 and P2, therefore imply that for N and L sufficiently large $\bar{y}(t)$ will be approximately normally distributed. The result now follows by invoking theorem (13.5.1) of Anderson (1984).

References

Anderson, T. W. (1984) *An Introduction to Multivariate Analysis*, 2nd edn. New York: Wiley.
Billingsley, P. (1986) *Probability and Measure*, 2nd edn. New York: Wiley.

Carroll, R. J. and Hall, P. (1988) Optimal rates of convergence for deconvolving a density. *J. Am. Statist. Ass.*, **83**, 1184–1186.

Cheng, O. (1990) Maximum standardized cumulant deconvolution of non-Gaussian linear processes. *Ann. Statist.*, **18**, 1774–1783.

Chung, S.-H., Doğançay, K. and Poskitt, D. S. (1998) A new analytical method of studying synaptic transmission in the brain. *Technical Report*. Australian National University, Canberra.

Clements, J. (1991) Quantal synaptic transmission? *Nature*, **353**, 396.

Davies, R. A. and Rosenblatt, M. (1991) Parameter estimation for some time series models without contiguity. *Statist. Probab. Lett.*, **11**, 515–521.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Dieudonné, J. (1960) *Foundations of Modern Analysis*. New York: Academic Press.

Donoho, D. (1981) On minimum entropy deconvolution. In *Applied Time Series Analysis* (ed. D. F. Findlay), vol. II, pp. 564–608. New York: Academic Press.

Fan, J. (1992) Deconvolution with supersmooth distributions. *Can. J. Statist.*, **20**, 155–169.

Gamboa, F. and Gassiat, E. (1996) Blind deconvolution of discrete linear systems. *Ann. Statist.*, **24**, 1964–1981.

Jennrich, R. I. (1969) Asymptotic properties of nonlinear least square estimators. *Ann. Math. Statist.*, **40**, 633–643.

Li, T. H. (1995) Blind deconvolution of linear system with multilevel non-stationary input. *Ann. Statist.*, **23**, 690–704.

Lii, K. S. and Rosenblatt, M. (1982) Deconvolution and estimation of transfer function phase and coefficient for non-Gaussian linear processes. *Ann. Statist.*, **10**, 1195–1208.

McCullagh, P. (1991) Quasi-likelihood and estimation functions. In *Statistical Theory and Modelling: in Honour of Sir David Cox* (eds D. V. Hinkley, N. Reid and E. J. Snell). London: Chapman and Hall.

Rao, C. R. (1965) *Linear Statistical Inference and Its Applications*. New York: Wiley.

Roeder, K. (1992) Semiparametric estimation of normal mixture densities. *Ann. Statist.*, **20**, 929–943.

Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc.* B, **53**, 683–690.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Stefanski, L. and Carroll, R. J. (1990) Deconvolving kernel density estimators. *Statistics*, **21**, 169–184.

Stricker, C., Field, A. C. and Redman, S. J. (1996) Statistical analysis of amplitude fluctuations in EPSCs evoked in rat CA1 pyramidal neurons *in vitro*. *Lond. J. Physiol.*, **490**, 419–441.

Stricker, C. and Redman, S. (1994) Statistical models of synaptic transmission evaluated using the expectation-maximization algorithm. *Biophys. J.*, **67**, 656–670.

Wand, M. P. and Jones, M. C. (1995) *Kernel Smoothing*. London: Chapman and Hall.

White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.