Journal of Neuroscience Methods 179 (2009) 131–141

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Denoising neural data with state-space smoothing: Method and application

Hariharan Nalatore[a], Mingzhou Ding[b], Govindan Rangarajan[c,*]

[a] Applied Research International, New Delhi 110016, India
[b] The J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611, USA
[c] Department of Mathematics and Centre for Neuroscience, Indian Institute of Science, Bangalore 560012, India

ARTICLE INFO

Article history:
Received 1 October 2008
Received in revised form 12 January 2009
Accepted 14 January 2009

Keywords:
EM algorithm
Kalman smoothing

ABSTRACT

Neural data are inevitably contaminated by noise. When such noisy data are subjected to statistical analysis, misleading conclusions can be reached. Here we attempt to address this problem by applying a state-space smoothing method, based on the combined use of the Kalman filter theory and the Expectation–Maximization algorithm, to denoise two datasets of local field potentials recorded from monkeys performing a visuomotor task. For the first dataset, it was found that the analysis of the high gamma band (60–90 Hz) neural activity in the prefrontal cortex is highly susceptible to the effect of noise, and denoising leads to markedly improved results that were physiologically interpretable. For the second dataset, Granger causality between primary motor and primary somatosensory cortices was not consistent across two monkeys and the effect of noise was suspected. After denoising, the discrepancy between the two subjects was significantly reduced.

1. Introduction

Experimental measurements are noisy. For neural recordings, the noise may arise from a multitude of sources, both intrinsic and extrinsic to the nervous system. Operationally, supposing that recorded data are composed of two parts, signal of interest and other processes unrelated to the experimental conditions, the latter can be collectively referred to as noise. The presence of noise can adversely impact the statistical analysis performed on the data (Albo et al., 2004). Consider two possibilities. First, if noise has a broadband spectrum, its deleterious effect is thus expected to become progressively more severe in higher frequencies (e.g. gamma band), as the power of neural signals typically decreases with frequency in a $1/f$ fashion (Buzsaki and Draguhn, 2004). Thus far, this problem has not received much research interest, despite the fact that high frequency neural activity is hypothesized to have a significant role in normal brain functions and in pathology (Keil et al., 1999; Tallon-Baudry and Bertrand, 1999; Buzsaki and Draguhn, 2004; Schnitzler and Gross, 2005). Second, for multivariate neural data, Granger causality has become a useful tool in revealing directions of neuronal interactions among different brain regions, both in the time and in the frequency domain (Bernasconi and Konig, 1999; Bernasconi et al., 2000; Hesse et al., 2003; Brovelli et al., 2004; Bollimunta et al., 2008; Dhamala et al., 2008a,b; Guo et al., 2008a,b; Marinazzo et al., 2008). Theoretical derivations and numerical simulations have shown that noise, depending on the signal-to-noise ratio, can give rise to false directions while masking true directions (Nalatore et al., 2007). The manifestation of this problem in neural data analysis has not been studied.

Analyzing two datasets of local field potential recordings from monkeys performing a visuomotor task, we wish to accomplish two objectives. The first objective is to demonstrate the adverse effects of noise in two specific problems: (1) correlation between prefrontal high gamma activity prior to stimulus onset and response time and (2) beta band Granger causality between primary motor and primary somatosensory cortex during motor maintenance. First, a positive correlation between the level of prestimulus high gamma oscillation (60–90 Hz) and the response time (RT) was found in one monkey (TI). This result contradicts the known properties of gamma oscillations, and the effect of noise is suspected. Second, for the interaction between the primary motor and primary somatosensory cortex in the beta band (15–30 Hz), Granger causality analysis revealed apparent discrepancies between two monkeys (GE and LU), and the effect of noise was again suspected. The second objective is to evaluate the effectiveness of a statistically principled method to separate signal from noise. The method, formulated in state space, combines Kalman filter smoothing with the Expectation and Maximization (EM) algorithm, and has been proven effective in a number of previous studies (Dempster et al., 1977; Digalakis et al., 1993; Gahramani and Hinton, 1996; Weinstein et al., 1994; Shumway and Stoffer, 1982; Smith and Brown, 2003; Smith et al., 2004, 2005; Nalatore et al., 2007). Our results show that, after denoising, (1) the correlation between prestimulus prefrontal high

* Corresponding author. Tel.: +91 80 23600373; fax: +91 80 23600146.
E-mail address: rangaraj@math.iisc.ernet.in (G. Rangarajan).

0165-0270/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2009.01.013

gamma activity and response time became significantly negative, meaning that higher levels of gamma activity immediately prior to stimulus onset lead to faster response times, an observation consistent with the putative role of gamma activity in mediating top-down attentional control (Engel et al., 2001) and (2) the Granger causal influences in the beta band between primary motor and primary somatosensory cortices become more consistent between the two monkey subjects. Both results can be seen as providing evidence for the effectiveness of the proposed denoising approach.

2. Methods

2.1. The denoising algorithm

Kalman filtering (Haykin, 2001) is a standard method for removing noise from noisy data, a process called denoising. Let $y_t=(y_{1t}, y_{2t}, \ldots, y_{Nt})'$ denote the data from N recording channels at time t (where $'$ denotes matrix transpose). We will model this using the following noisy multivariate autoregressive (MVAR) model:

$$z_t = A_1 z_{t-1} + A_2 z_{t-2} + \cdots + A_p z_{t-p} + \varepsilon_t, \tag{1}$$

$$y_t = z_t + v_t, \tag{2}$$

where z_t is an $N\times 1$ vector giving the true state of the system, A_i is an $N\times N$ coefficient matrix, p is the order of the MVAR process, ε_t is an $N\times 1$ Gaussian error vector with zero mean and covariance matrix S, and v_t is an $N\times 1$ noise vector with zero mean and covariance matrix R. Thus the observed time series $\{y_t\}$ is being viewed as composed of two parts: signal of interest (z_t) and other unrelated processes collectively referred to as noise (v_t). The signal is recovered via an iteration process to be detailed below, which is initiated by estimating the parameters in the above equations through fitting an MVAR model directly to the noisy data using standard procedures, including model order determination by the AIC criterion (Ding et al., 2000). The source of noise in neural data can be manifold, including noise that is intrinsic to the nervous system, as well as environmental and instrumental noise. For more discussions on this, see Section 4.

To apply the Kalman filter algorithm, we first need to rewrite the above model in a state-space form. This can be accomplished by introducing an $M\times 1$ state vector $x_t = (z'_t, z'_{t-1}, \ldots, z'_{t-p+1})'$ where $M=Np$. In terms of this vector, it can be easily shown (Shumway and Stoffer, 2000) that the noisy MVAR(p) model can be written as

$$x_t = Ax_{t-1} + w_t, \tag{3}$$

$$y_t = Cx_t + v_t. \tag{4}$$

Here x_t is the unobserved (or "hidden") signal vector of dimension $M\times 1$ and y_t is the $N\times 1$ observed data vector that is the signal contaminated by noise v_t. The objective of Kalman filtering is to recover x_t based on y_t. In the above equation, A is an $M\times M$ state transition matrix given in terms of the unknown coefficient matrices A_i (Shumway and Stoffer, 2000), C is a trivial $N\times M$ observation matrix given by $(I, 0, \ldots, 0)$ comprising one $N\times N$ identity matrix and $p-1$ $N\times N$ zero matrices, and $w_t = (\varepsilon'_t, 0', \ldots, 0' \ldots)'$ is an $M\times 1$ zero-mean Gaussian independent and identically distributed vector random variable with the $M\times M$ covariance matrix Q (which has S in the upper right-hand corner and zeros elsewhere).

This formulation of Kalman filter is not directly applicable to experimental data as it assumes the knowledge of the model describing the state-space dynamics. That is, A, C, Q, R are assumed to be known. In our case, this knowledge is not available except for C (which is a fixed constant matrix for MVAR models as described above). This problem is overcome by combining the Kalman filter formulation with the Expectation and Maximization algorithm (Dempster et al., 1977; Digalakis et al., 1993; Gahramani and Hinton, 1996; Weinstein et al., 1994). A similar approach has been used by Smith et al. to estimate state-space parameters from neural spike trains and behavioral data (Smith and Brown, 2003; Smith et al., 2004, 2005). A review of other applications of this method appears in Roweis and Ghahramani (1999).

The denoising algorithm includes the following steps (Nalatore et al., 2007). Let $\{x\}$ and $\{y\}$ denote the set $\{x_t\}$ and $\{y_t\}$, respectively, for all time. Other than the actually observed vector $\{y\}$, if we were able to observe the hidden state vector $\{x\}$, then we could consider $\{x, y\}$ as the complete data with the joint density (Shumway and Stoffer, 1982, 2000):

$$p(\{x\},\{y\}) = \prod_{t=1}^{T} p(y_t|x_1) \prod_{t=1}^{T} p(x_t|x_{t-1}) p(x_1). \tag{5}$$

Under the Gaussian assumption, the joint log likelihood (given by $\log P(\{x\},\{y\})$) can be written as

$$\begin{aligned}\log P(\{x\},\{y\}) = &-\sum_{t=1}^{T}\left(\frac{1}{2}[y_t - Cx_t]'R^{-1}[y_t - Cx_t]\right) - \frac{T}{2}\log|R| \\ &-\sum_{t=1}^{T}\left(\frac{1}{2}[x_t - Ax_{t-1}]'Q^{-1}[x_t - Ax_{t-1}]\right) \\ &-\frac{(T-1)}{2}\log|Q| - \frac{1}{2}\sum_{t=1}^{T}[x_1 - \mu_1]'V_1^{-1}[x_1 - \mu_1] \\ &-\frac{1}{2}\log|V_1| - \frac{T(N+m)}{2}\log 2\pi \end{aligned} \tag{6}$$

where $'$ again denotes matrix transpose. We have assumed that $x_1 \sim N(\mu_1, V_1)$ where μ_1, V_1 are fixed. The unknown parameters are $\theta=\{A, Q, R\}$. If we could observe $\{x\}$, we could obtain the maximum likelihood estimates (MLEs) of these parameters by maximizing the above joint likelihood function with respect to these parameters. Since $\{x\}$ is unobserved, we need to use the EM algorithm in conjunction with the Kalman smoother to obtain estimates of θ and of course $\{x\}$. These are obtained by iteratively maximizing the conditional expectation of the joint likelihood function given by:

$$O = E[\log P(\{x\},\{y\})|\{y\}, \theta^{(j-1)}], \tag{7}$$

For $j=1, 2, \ldots$.

We start the iteration with the initial guess $\theta^{(0)}$ for the parameter values. We obtain these by applying the standard AR model estimation procedures to the noisy data (Ding et al., 2000), yielding A_is and S, which can then be put in their respective state-space forms $A^{(0)}$ and $Q^{(0)}$. The initial guess $R^{(0)}$ for R is usually taken to be a fractional multiple of the identity matrix. Then O depends on the following three conditional expectations: $\hat{x}_{t|T} \equiv E[x_t|\{y\}, \theta^{(j-1)}]$, $P_{t|T} \equiv E[x_t x'_t\,|\{y\}\,, \theta^{(j-1)}]$, $P_{t,t-1|T} \equiv E[x_t x'_{t-1}|\{y\}, \theta^{(j-1)}]$.

These quantities can be calculated using the Kalman smoother (see Appendix A for the smoother equations). This completes the E step.

Next, we go to the maximization (M) step. Each of the parameters A, Q, R is re-estimated by maximizing O. The expressions obtained for these parameters are:

$$A^{(1)} = \left(\sum_{t=2}^{T} P_{t,t-1|T}\right)\left(\sum_{t=2}^{T} P_{t-1|T}\right)^{-1}, \tag{8}$$

$$Q^{(1)} = \frac{1}{T-1}\left(\sum_{t=2}^{T} P_t - A^{(1)}\sum_{t=2}^{T} P_{t-1,t|T}\right), \tag{9}$$

$$R^{(1)} = \frac{1}{T}(y_t y'_t - C\hat{x}_{t|T} y'_t). \tag{10}$$