

A Signal Estimation Approach to Functional MRI

V. Solo*, *Fellow, IEEE*, P. Purdon, R. Weisskoff, and E. Brown

Abstract—In the last half decade, fast methods of magnetic resonance imaging have led to the possibility, for the first time, of non-invasive dynamic brain imaging. This has led to an explosion of work in the Neurosciences. From a signal processing viewpoint the problems are those of nonlinear spatio-temporal system identification. In this paper, we develop new methods of identification using novel spatial regularization. We also develop a new model comparison technique and use that to compare our method with existing techniques on some experimental data.

Index Terms—Functional MRI, image processing, inverse problem, MRI.

I. INTRODUCTION

A. General

RECENT developments in fast imaging with magnetic resonance imaging (MRI) have allowed researchers to study the function of the human brain with a unique combination of high spatial and temporal resolution that is unmatched by other imaging modalities [1], [2]. The MR signal is sensitive to local changes in blood oxygenation and blood volume brought about by neural activity. The signal changes observed through this blood oxygen level-dependent (BOLD) contrast mechanism occur on a time scale of seconds, allowing observation of an interesting range of dynamical patterns within and among brain regions. A brief, readable survey of brain Mapping with this functional MRI (fMRI) is available in [3].

In a typical block design fMRI experiment, a subject is presented with a stimulus or cognitive task, in a periodic “off”–“on” pattern, while images of the brain are taken in rapid succession. An example might be a flickering checkerboard visual stimulus which is “on” for 10 s and “off” for 15 s (i.e., a “square-wave” stimulus). Such experiments are designed to determine and analyze regions of functional specialization within the brain that are related to the stimulus presented. A typical fMRI experiment lasts several minutes, with a sampling interval between images ranging from hundreds of milliseconds to several seconds and with a spatial resolution of a few millimeters. More recently, so-called event related fMRI experiments have been developed

[4], [5] in which multiple brief stimuli are applied during a single experiment. Although our approach supports analysis of this kind of data, we have not pursued such examples here.

The reconstructed fMRI signal $x_{t,P}$ at time t and pixel P is thought to consist of two factors: 1) the BOLD hemodynamic response $s_{t,P}$ brought about by some stimulus c_t and 2) the “brain noise” $v_{t,P}$. This “brain noise” consists of hemodynamic fluctuations of unknown origin, possibly related to “background” processes in the brain, as well as cardiac and respiratory fluctuations. In most cases, the cardiac and respiratory influences account for only a small portion of the observed variability.

B. Review and Preview

Methods of analysis of fMRI data are in a rapid state of development. With rare exceptions ([6] and our own conference papers [7], [8]) previous modeling work has been pixel-wise with no notion of using any spatial continuity to improve signal and parameter estimation efficiency. Also, many approaches use hypothesis testing based on differences between averages in ON and OFF periods. In this paper, we are interested in taking a signal estimation approach that is not tied to a particular stimulus signal paradigm and that allows the possibility of increased statistical estimation efficiency. This is particularly important since more complex stimulus signals are coming into use [4]. Also scientific questions may go unanswered because of low signal-to-noise ratio (SNR) in estimation. The most widely used current approach is that of [9] which uses a linear convolution model to relate stimulus to response and deals with colored noise by prefiltering. They make no use of spatial regularization. Because of its common use, we will compare our results to those obtained by the modeling technique of [9]. In [6], a framework similar to that of [9] is used but spatially varying delays are allowed as well as colored noise. Also spatial smoothing is employed and a more flexible impulse response model is used. Our work has some similarities to this but was developed independently. Other work of note includes [10] who use singular value decomposition (SVD). This approach does not impose temporal or spatial continuity. And it does not directly incorporate estimation of the stimulus to response transfer function. This can be done after the SVD is computed but the introduction of this constraint late in the signal processing procedure leads to bias. In [11], wavelets have been applied to fMRI. The method assumes white noise and does not give an explicit transfer function estimate. In [12], Markov random field (MRF) methods have been applied to the spatio-temporal modeling of fMRI. There are several stages. First, an overall MRF prior is applied in space and time. This is not attractive since the data is much rougher space-wise than time-wise. It would be interesting to develop more suitable priors that recognized the distinction between time and space. An impulse response is estimated from

Manuscript received June 9, 1999; revised October 27, 2000. This work was supported by the National Institutes of Health (NIH) under Grant NCR P41RR14075 at the NMR Center, MGH. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was V. Johnson. *Asterisk indicates corresponding author.*

*V. Solo is with the School of Electrical Engineering and Telecommunications, University of New South Wales, 2052 Sydney, Australia. He is also with the MGH-NMR Center, Department of Radiology, Harvard University Medical School, Cambridge, MA 02138 USA.

P. Purdon and R. Weisskoff are with the MGH-NMR Center, Department of Radiology, Harvard University Medical School, Cambridge, MA 02138 USA.

E. Brown is with the Neuroscience Statistics Research Lab, Department of Anaesthesia and Critical Care, Massachusetts General Hospital, Harvard University Medical School, Cambridge, MA 02138 USA.

Publisher Item Identifier S 0278-0062(01)00952-1.

the first step reconstruction and then another MRF step follows. The failure to apply all constraints at the beginning leads to a loss of statistical efficiency for this procedure.

We develop a weakly nonlinear spatio-temporal colored noise model based partly on empirical work and fit the model using a local spatial regularization method that is computationally efficient. We also develop a new model comparison method. The local regularization and the model comparison are the main contributions of the work, but the emphasis on system identification rather than detection is also important. Section II describes the model, Section III discusses the model fitting based on cyclic descent and the model comparison is developed in Section IV. An example of the method applied to real data including comparison to the generalized linear method (GLM) is described in Section V while conclusions are offered in Section VI. More technical material is placed in Appendices A and B.

C. Notation

For convenience, we collect notation and definitions here. In places a superscript zero denotes no regularization. P or Q denotes a pixel, t denotes a time. f_P is an activation amplitude at pixel P and in our model below will have three components, $f_P = [f_{aP}, f_{bP}, f_{cP}]^T$. θ_P denotes a general parameter at pixel P and is split into $\theta_P = [\beta_P, \alpha_P]^T$ where β_P is a signal parameter and α_P is a noise parameter. We denote d_β, d_α to be the per pixel dimensions of β_P, α_P . Below we will meet some nuisance parameters (motion parameters) which we will lump in with the signal parameters. $x_{t,P}$ is a signal at time t , pixel P and $\tilde{x}_{k,P}$ denotes its temporal Fourier transform where k is a Fourier frequency

$$\tilde{x}_{k,P} = \sum_0^{n-1} x_{t,P} e^{-j\omega_k t}, \quad w_k = \frac{2\pi k}{n}.$$

$F_{k,P} = F_k(\alpha_P)$ is a power spectrum at frequency k . Also we will use several covariances, each of the general form

$$\sigma_{uxP} = n^{-2} \Re(\sum_1^{n-1} \tilde{x}_{k,P} \tilde{u}_k^H / F_{k,P})$$

where superscript H denotes Hermetian i.e., complex conjugate transform. Also we will use a spatial weighting kernel or mask, K_Q^h that sums to one and is supported on a finite neighborhood. This could be e.g., cone or pyramid shaped, e.g.,

$$K_Q^h = K\left(\frac{Q_1}{M_1 h}\right) K\left(\frac{Q_2}{M_2 h}\right) \frac{c_M}{M h^2}, \quad (M = M_1 M_2).$$

e.g., $K(u) = 0.75(1 - u^2)_+$ or $K(u) = (1 - |u|)_+$ and $0 < h < 1$; h controls the effective mask size. The support region corresponding to h is of size $(2M_1 h - 1)(2M_2 h - 1)$. Also c_M is a correction term for the discreteness of the pixels; $c_M = (1 - (1/4M_1^2))(1 - (1/4M_2^2))$. In the sequel, we use “dc” to denote a mean or baseline or constant value.

II. TIME SERIES DECOMPOSITION

Our approach builds on earlier work of several researchers; [13] who used a convolution with a Poisson shaped impulse response to relate stimulus to response; [14] who showed that there is important low frequency colored noise; [15] who investigated the validity of linearity and convolution; and work

of [16] which attempts to inject some physiology into the modeling. More details of the construction of the model are given in [17]; here, we concentrate on statistical signal processing issues. There are other approaches which do not, however, explicitly model the relation between stimulus and response. These are discussed further below.

Our model then is a pixel by pixel construction. Here, we only deal with single slice data and so treat motion effects crudely by means of a drift term. We begin by describing the signal model and then follow with a description of the noise model. The signal $s_{t,P}$ is a product of two terms; a term representing oxygen level and a term representing blood volume. Each of these terms is modeled as a convolution of the stimulus with an impulse response; more details are given in [17]. We have then

$$\begin{aligned} s_{t,P} &= H(t - D_P)(\gamma_{oP} + \gamma_{1P} g_P * c_{t-D_P}) \\ &\quad \times (V_{oP} + V_{1P} p_{t-D_P}) \\ p_t &= e^{-t/12} * c_t \end{aligned}$$

where

γ_{oP}, V_{oP}	“dc” levels;
γ_{1P}, V_{1P}	gains;
$g_P(t)$	hemodynamic impulse response;
D_P	spatially varying delay;
$H(\cdot)$	heaviside step function;
*	denotes temporal convolution.

The form of the response p_t is based on empirical studies and has been prespecified here, because as we shall see in the kind of experiments we have analyzed, the stimulus signal has insufficient persistence of excitation to provide reliable estimates of many parameters. The hump shaped $g_P(t)$ impulse response is motivated by empirical studies and also by a similar model used by [6] and given pixel-wise by

$$g_P(t) = (1 - e^{-1/\tau_P})^2 (t + 1) e^{-t/\tau_P}$$

with normalizing constant chosen to make $g_P(t)$ sum to one in t , for each pixel P . Also τ_P is a time constant and then the hump has a maximum at $t = \tau_P + 1$. Note also that $g_P(t)$ is well behaved in t , as $\tau_P \rightarrow 0$ or $\tau_P \rightarrow \infty$.

Expanding the product and reorganizing terms gives

$$\begin{aligned} s_{t,P} &= \gamma_{oP} V_{oP} H(t - D_P) + f_{aP} (g_P * c)(t - D_P) \\ &\quad + f_{bP} p(t - D_P) + f_{cP} p(t - D_P) (g_P * c)(t - D_P). \end{aligned} \quad (2.1)$$

Turning to the noise, based on empirical studies [14] we use a first order autoregressive model for haemodynamic fluctuations and white noise for background “thermal” noise; again see [17] for details. The noise model is then

$$\begin{aligned} v_{t,P} &= \text{AR}(1) \text{ noise} + \text{white noise} \\ &= w_{t,P} + \nu_{t,P} \\ &= \frac{\eta_{t,P}}{1 - a_P z^{-1}} + \nu_{t,P} \end{aligned}$$

where $\nu_{t,P}, \eta_{t,P}$ are independent white noises of variances $\sigma_{\nu P}^2, \sigma_{\eta P}^2$, a_P is the AR(1) parameter and z^{-1} is the backshift operator.

We incorporate the dc term in (2.1) into an overall dc term m_P while a drift term $b_P t$ is added to deal with motion as described above. Our overall model then gives the fMRI signal $x_{t,P}$ as (with $t = 1, \dots, n$, $P = 1 \dots M$)

$$x_{t,P} = m_P + b_P t + s_{t,P} + v_{t,P}. \quad (2.2)$$

III. MODEL FITTING

It proves computationally advantageous to carry out model fitting in the frequency domain. Fourier transforming (2.1), (2.2) and neglecting aliasing (which proves to be no problem) gives

$$\tilde{x}_{k,P} = m_P \delta_{0k} n + b_P u_k + e^{-j\omega_k D_P} \tilde{\xi}_{k,P}^T f_P + \tilde{v}_{k,P} \quad (3.1)$$

where we have set $f_P = [f_{aP}, f_{bP}, f_{cP}]^T$ and

$$\begin{aligned} \tilde{\xi}_{k,P} &= [\tilde{\xi}_{k,P,a}, \tilde{\xi}_{k,P,b}, \tilde{\xi}_{k,P,c}]^T; \tilde{\xi}_{k,P,a} = \tilde{g}_{P,k} \tilde{c}_k \text{ etc.} \\ u_k &= \frac{1}{n^{-1}(1 - e^{-j\omega_k})} \quad k \neq 0; = 0, k = 0 \\ \delta_{0k} &= 0, \quad k \neq 0; = 1, k = 0. \end{aligned}$$

Local Likelihood: The likelihood function is defined pixel-wise and to develop it clearly we first exhibit it without spatial regularization. We use a Gaussian negative log likelihood criterion which has a well known frequency domain approximation [18] as follows. In the absence of spatial regularization, it has the pixel-wise form

$$J_P^0(\theta_P) = \sum \sum \frac{I_{k,P}(\beta_P)}{2MnF_k(\alpha_P)} + \frac{\sum \sum \log F_k(\alpha_P)}{2M} \quad (3.2)$$

$$I_{k,P}(\beta_P) = \frac{|\tilde{x}_{k,P} - m_P \delta_{0k} n - b_P u_k - e^{-j\omega_k D_P} \tilde{\xi}_{k,P}^T f_P|^2}{n} \quad (3.3)$$

$$F_k(\alpha_P) = \sigma_\nu^2 + \frac{\sigma_{\tilde{\xi}_{k,P}}^2}{|1 - \alpha_P e^{-j\omega_k}|^2}. \quad (3.4)$$

A typical approach to recognizing the spatial continuity in some of the parameters is by applying, say, Tikhonov regularization [19] to this criterion. However, that leads to a global nonlinear optimization problem that is computationally intractable. Instead we have applied the local polynomial technique that has become popular in Statistics [20] but which dates back to Econometrics of the 1920s. Similar (but not the same) ideas are used in optical flow estimation in computer vision [21], [22]. At each pixel P we consider a neighborhood of nearby pixels. For the purpose of parameter estimation at pixel P , we set θ_P to be fixed. We use a likelihood function, however, that is a weighted average of the pixel-wise unregularized likelihood above, thus

$$J_P(\theta_P) = \sum_Q K_Q^h J_{P-Q}^0(\theta_P).$$

Or, more carefully

$$J_P(\theta_P) = \frac{1}{2} \sum_Q K_Q^h \left(\sum_k \ln F_k(\alpha_P) + \sum_k \frac{I_{k,P-Q}(\beta_P)}{nF_k(\alpha_P)} \right) \quad (3.5)$$

where K_Q^h is the weight function or mask as described above. We now optimize $J_P(\theta_P)$ to get the estimate of θ_P at pixel P . Then we move through the whole region one pixel at a time repeating the whole computation. This procedure provides non-parametric estimation of θ_P . Our spatio-temporal model is a semiparametric model in the sense that the temporal specification is parametric while the spatial part is nonparametric. We can also allow that θ_P varies as a polynomial in each neighborhood and this is known to provide superior bias and variance properties [20]. We have not implemented that extension here, however. At the edges just use that part of the mask that overlaps the region (with necessary rescaling so weights in mask sum to 1).

In the sequel, we assume σ_ν^2 is known from empirical studies. But below in the EM algorithm we show how to estimate σ_ν^2 also with a minor change.

Computational details are given in Appendix A. The general strategy is to optimize using cyclic descent with a judiciously chosen order that uses parameter linearity where possible and allows the delay to be treated as a one dimensional correlation calculation. We call our method locally regularized spatio-temporal modeling (LRST). Derivations of the computational equations are given in Appendix B.

IV. MODEL COMPARISON

The problem of comparison of statistical models, which includes the problem of model order selection for finitely parameterized models, has a considerable literature. A nice summary of this work can be found in [23]. We need not only to be able to compare models of different structure such as LRST and GLM, but also models of the same structure that depend on a tuning parameter; in our case a mask size h . Despite the considerable work on this problem there are no available methods that cover our case; e.g., criteria like Akaike's information criterion (AIC) are not applicable because the tuning parameter in our case is not a model dimension. We present a new model comparison criterion called NURE (nearly unbiased risk estimation) which is of general applicability. It may be regarded as a generalization of methods such as AIC and has been briefly described on a totally different problem in [8].

To develop a model comparison criterion one needs, for given θ , a discrepancy criterion or risk function $R(\theta)$ that measures the discrepancy between a user chosen feature (such as a spectrum) and the model of it. Models are then compared on the basis of $R = E(R(\hat{\theta}(x)))$ where $\hat{\theta}(x)$ is an estimator of θ based on data x . Of course the model has to be fitted to the data x based on some data fidelity criterion $J(\theta, x)$. $R(\theta)$ and $E(J(\theta, x))$ may or may not be related.

Now there are two problems in carrying this out. Because $R(\theta)$ involves data features it is not computable; so one must find an unbiased estimator of it instead, say $\hat{R}(\theta) = R(\theta, x)$ based on data x ; this is usually not too difficult to construct. The natural surrogate for R will then be $\hat{R}(\hat{\theta}(x)) = R(\hat{\theta}(x), x)$ but this will be biased. So a Taylor series argument is used to provide a correction term (which turns out to measure model complexity) to be added to $\hat{R}(\hat{\theta}(x))$ to yield a nearly unbiased

estimator of R . A brief derivation is given in [8]. We now describe these elements for our problem.

A. Risk Criterion

The risk criterion we use is the Kullback–Liebler (KL) information, because it seems to be almost the only criterion that simultaneously measures discrepancy with respect to both signal and noise parameters in a natural way. The KL information $R_{\text{KL}}(\theta)$ is given by

$$R_{\text{KL}}(\theta) = \int p(x) \ln \left(\frac{p(x)}{p_\theta(x)} \right) dx$$

where $p(x)$ is the true but unknown density of the data and $p_\theta(x)$ is that under the model. Under Gaussian assumptions

$$p(x) = (2\pi)^{-n/2} |\Sigma|^{-(1/2)} e^{-(1/2)(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

and similarly for $p_\theta(x)$. So we find

$$R_{\text{KL}}(\theta) = -\frac{1}{2} \ln |\Sigma_\theta^{-1} \Sigma| - \frac{n}{2} + \frac{1}{2} \Delta_\theta^T \Sigma_\theta^{-1} \Delta_\theta + \frac{1}{2} \text{tr}(\Sigma_\theta^{-1} \Sigma) \\ \Delta_\theta = \mu_\theta - \mu.$$

Under stationary noise assumptions Σ , Σ_θ are both Toeplitz matrices and are nearly diagonalized by Fourier eigenvectors. So we get (dropping the $n/2$) approximately on recognizing also the pixel-wise independence

$$R_{\text{KL}}(\theta) = \Sigma_P R_{\text{KL}}(\theta_P) \\ R_{\text{KL}}(\theta_P) = -\frac{1}{2} \Sigma \ln \frac{F_{k,P,a}}{F_{k,P}} + \frac{1}{2} \Sigma \frac{|\tilde{\Delta}_{k,\theta_P,P}|^2 + F_{k,P,a}}{F_{k,P}}$$

where $F_{k,P,a}$ is the true or actual spectrum of $v_{t,P}$ at temporal frequency $\omega_k = 2\pi k/n$ and $F_{k,P} = F_k(\alpha_P)$ is the spectrum according to the model. Also $\tilde{\Delta}_{k,\theta_P,P}$ is the DFT of $\Delta_{\theta_P,P}$.

Again, we do not know $F_{k,P,a}$, $\tilde{\Delta}_{k,\theta_P,P}$ so we cannot calculate $R_{\text{KL}}(\theta_P)$ and, hence, $R_{\text{KL}}(\theta)$. So as indicted above we construct an unbiased estimator valid for fixed θ . Thus, (dropping the $\ln F_{k,P,a}$ term, which does not depend on θ) we see easily that for fixed θ [with $I_{k,P}(\beta_P) = |\tilde{x}_{k,P} - \tilde{\mu}_{k,P}(\beta_P)|^2 n^{-1}$]

$$\hat{R}_{\text{KL}}(\theta_P) = \frac{1}{2} \Sigma \ln F_{k,P} + \frac{1}{2} \Sigma \frac{|\tilde{x}_{k,P} - \tilde{\mu}_{k,P}(\beta_P)|^2}{n F_{k,P}} \\ = \frac{1}{2} \Sigma \ln F_k(\alpha_P) + \frac{1}{2} \Sigma \frac{I_{k,P}(\beta_P)}{F_k(\alpha_P)} \quad (4.1)$$

is an unbiased estimator of $R_{\text{KL}}(\theta_P)$. Again it is not sufficient to use $\Sigma_P \hat{R}_{\text{KL}}(\hat{\theta}_P)$ as our model comparison criterion since it is biased as an estimator of $\Sigma_P E(R_{\text{KL}}(\hat{\theta}_P))$. Taylor series expansions allow one to calculate the bias terms (see below). But to do so we must specify the model fitting criterion which we now do.

B. Data Fidelity Criterion

We have been using a locally spatially weighted likelihood function. A crucial observation is that the local weighting means that the overall data fidelity criterion is pixelwise separable in

the parameters. This simplifies computation of NURE enormously as compared to what one would get with say Tikhonov regularized criteria, since it makes matrices that appear in the model complexity term to be block diagonal. We have then

$$J(\theta) = \Sigma_P J_P(\theta_P).$$

The NURE is then given by (see [8] for a brief general derivation)

$$\hat{R}_{\text{KL}} = \Sigma_P \hat{R}_{\text{KL}}(\hat{\theta}_P) + \Sigma_P \text{trace}(J_{2,P,e}^{-1} W_{P,e}).$$

The trace term is a bias term or model complexity term and it has both signal and noise components. Also

$$J_{2,P,e} = E \left(\frac{d^2 J_P}{d\theta_{P,e} d\theta_{P,e}^T} \right) \\ W_{P,e} = E \left(\frac{dJ_P}{d\theta_{P,e}} \frac{d\hat{R}_{\text{KL}}(\theta_P)}{d\theta_{P,e}^T} \right) \quad (4.2)$$

where $\theta_{P,e}$ is the ‘‘pseudomean’’ which satisfies

$$\frac{dE(J_P)}{d\theta_{P,e}} = 0.$$

$\theta_{P,e}$ is unknown and we usually replace it by $\hat{\theta}_P$ in the above expressions. This introduces extra bias but it is of lower order.

We can separate $J_{2,P,e}$, $W_{P,e}$ into signal and noise components as follows:

$$J_{2,P,e} = \begin{pmatrix} J_{2,\beta,\beta} & J_{2,\beta,\alpha} \\ J_{2,\alpha,\beta} & J_{2,\alpha,\alpha} \end{pmatrix} \\ W_{P,e} = \begin{pmatrix} W_{\beta,\beta} & W_{\beta,\alpha} \\ W_{\alpha,\beta} & W_{\alpha,\alpha} \end{pmatrix}$$

where e.g.,

$$J_{2,\beta\alpha} = E \left(\frac{d^2 J_P}{d\beta_{P,e} d\alpha_{P,e}^T} \right).$$

In each case below, the NURE has the form

$$\hat{R}_{\text{KL}} = \Sigma_P \hat{R}_{\text{KL}}(\hat{\theta}_P) + \hat{R}_{MC} \\ \hat{R}_{MC} = \text{model complexity} \\ \hat{R}_{MC} = \Sigma_P \text{trace}(J_{2,P,e}^{-1} W_{P,e}).$$

So, below, we just give the model complexity expressions. Explicit expressions for derivatives are straightforward and omitted.

The general NURE criterion does not assume that the data generating process belongs to the model class being fit. If it does belong then the model complexity trace term simplifies a lot and we denote the resulting criterion in lower case as *nure*. Thus, e.g., AIC is actually a *nure* criterion while its NURE equivalent is TIC (see [24] for a review and extensions). Below we develop only *nure* for simplicity.

C. NURE for LRST Model

For simplicity, here, we give only the nure version. Here, the signal and noise components uncouple and we get

$$\begin{aligned}
\hat{R}_{MC} &= \Sigma_P \text{trace}(J_{2,\beta\beta}^{-1} W_{\beta\beta}) + \Sigma_P \text{trace}(J_{2,\alpha\alpha}^{-1} W_{\alpha\alpha}) \\
J_{2,\beta\beta} &= -\Sigma K_Q^h \Sigma_k \frac{\lambda_{k,P-Q}(\beta_{Pe}) \lambda_{k,P-Q}^T(\beta_{Pe})}{F_k(\alpha_{Pe}) n} \\
W_{\beta\beta} &= -K_0^h \Sigma_k \frac{\lambda_{k,P}(\beta_{Pe}) \lambda_{k,P}^T(\beta_{Pe})}{F_k(\alpha_{Pe}) n} \\
\lambda_{k,P}(\beta_{Pe}) &= \frac{d\tilde{\mu}_{k,P}(\beta_{Pe})}{d\beta_{Pe}} \\
J_{2,\alpha\alpha} &= -\frac{1}{2} \Sigma K_Q^h \Sigma_k \frac{\psi_k(\alpha_{Pe}) \psi_k^T(\alpha_{Pe}) F_{k,P-Q,a}}{F_k(\alpha_{Pe})} \\
W_{\alpha\alpha} &= -\frac{1}{4} K_0^h \Sigma_k \psi_k(\alpha_{Pe}) \psi_k^T(\alpha_{Pe}) \\
\psi_k(\alpha_P) &= \frac{d \ln F_k(\alpha)}{d\alpha}. \tag{4.3}
\end{aligned}$$

In these formulas, we replace θ_{Pe} by $\hat{\theta}_P$ and $F_{k,P-Q,a}$ by $I_{k,P-Q}(\hat{\beta}_P)$. These results are relatively straightforward to obtain. For example, to obtain (4.3) use (4.2). So differentiate through (3.5) and (4.1) with respect to α_P . Then take cross covariances between the two expressions using the fact that $\text{cov}(I_{k,P}, I_{l,P-Q}) \approx F_{k,l}^2 \delta_{P-Q} \delta_{k-l}$.

Note that the signal and noise complexity terms are close to (d_β/h^2) , $2(d_\alpha/h^2)$ respectively. Alternatively if we call (approximately) $N = 4h^2M$ the neighborhood size then these penalties are $(4Md_\beta/N)$, $8(Md_\alpha/N)$. We may think of this roughly as # of independent parameters = # of independent parameters per neighborhood \times # of neighborhoods.

D. NURE for GLM

Again we give only the nure version here. The GLM model of [9] is

$$x_{t,P} = \mu_{t,P}(\beta_P) + \epsilon_{t,P}$$

where $\epsilon_{t,P}$ is a white noise. We use μ for a mean and β for a signal parameter but these are different parameters to ours and μ is a different function of β . There should be no confusion in context. Reference [9] smooth the data and estimate β from the smoothed data by ordinary least squares. If we denote their smoothing filter by ϕ_t , then since their calculation is pixel by pixel, their estimator is (superscript F is for [9])

$$\begin{aligned}
\hat{\beta} &= \arg. \min. J^F(\beta) \\
J^F(\beta) &= \Sigma_P J_P^F(\beta_P) \\
J_P^F(\beta_P) &= \frac{1}{2n} \Sigma_k |\tilde{x}_{k,P} - \tilde{\mu}_{k,P}(\beta_P)|^2 |\tilde{\phi}_k|^2 \\
&= \frac{1}{2} \Sigma_k I_{k,P}(\beta_P) |\tilde{\phi}_k|^2
\end{aligned}$$

where $\tilde{\phi}_k$ is the DFT of ϕ_t . If we compare this to our J_P then the simplest interpretation of $\tilde{\phi}_k$ is that it is a whitening filter. Thus,

$|\tilde{\phi}_k|^2$ takes the place of our $1/F_k(\alpha_P)$. Thus, the expression for $\hat{R}_{KL}(\hat{\theta}_P)$ will be

$$\hat{R}_{KL}(\hat{\theta}_P) = \frac{1}{2} \Sigma_k \ln \frac{1}{|\tilde{\phi}_k|^2} + \frac{1}{2} \Sigma_k I_{k,P}(\hat{\beta}_P) |\tilde{\phi}_k|^2.$$

Note that [9] model the drift by a Fourier series of seven terms.

There is only a signal model complexity term to be added namely $\Sigma_P \text{trace}(J_{2,\beta\beta}^{-1} W_{\beta\beta})$ where

$$\begin{aligned}
J_{2,\beta\beta} &= \Sigma_k \lambda_{k,P}(\beta_{Pe}) \lambda_{k,P}^T(\beta_{Pe}) \frac{|\tilde{\phi}_k|^2}{n} \\
W_{\beta\beta} &= \Sigma_k F_{k,P} \lambda_{k,P}(\beta_{Pe}) \lambda_{k,P}^T(\beta_{Pe}) \frac{|\tilde{\phi}_k|^4}{n}.
\end{aligned}$$

To compute $W_{\beta\beta}$ we could either put $1/F_{k,P} = |\tilde{\phi}_k|^2$ in which case the complexity term becomes simply $d_\beta M$. Or else we can use $F_{k,P} = I_{k,P}$ and then the trace will have to be computed.

Explicit expressions for $\lambda_{k,P}(\beta_{Pe})$ are straightforward since the [9] signal model is linear.

V. APPLICATION TO DATA

We applied our LRST method to fMRI data from a combined visual and motor experiment. The data were collected using a GE Signa 1.5-T scanner modified for echo-planar imaging (EPI) by ANMR. The subject was presented with a full-field flickering checkerboard, in a 12.8-s OFF, 12.8-s ON pattern, repeated eight times. A single slice image transecting both the visual and motor cortex was taken once every 800 ms for the duration of the experiment. These data were analyzed with the above methods over a range of regularization smoothing widths. An activation map was created by taking the square-root of the weighted 2-norm of the estimated activation signal as follows:

$$\text{activation} = \Sigma \frac{|\tilde{s}_{k,P}|^2}{F_{k,P}}.$$

This activation metric provides a rough measure of the extent of the activation signal while accounting for the large, spatially varying hemodynamic fluctuations (noise) by weighting with the noise spectrum. In order to compare our regularization technique to other spatial treatments, we also analyzed the data in two other ways: 1) by employing our method without regularization; and 2) by spatially prefiltering the data with a 3×3 separable Hanning kernel and then analyzing it without regularization. In Fig. 1, we show an anatomic image with several regions of interest (roi) corresponding to motor cortex 1), primary visual cortex 2), superior sagittal sinus 3), and white matter 4). Figs. 2–5 show time series of data from these ROIs along with timing diagrams of the stimulus. Figs. 2 and 3 also show plots of the LRST fits superimposed over the actual data, demonstrating a strong correspondence between the observed behavior and the model fit. In Figs. 6 and 7, similar plots are shown for the GLM fits. In each case, residuals are also shown; the residuals are estimates of $v_{t,P}$ in (2.2) and $\epsilon_{t,P}$ in the GLM model. In Fig. 8, we show a frequency domain view of the motion and baseline corrected signal. And by way of comparison in Fig. 9 we show

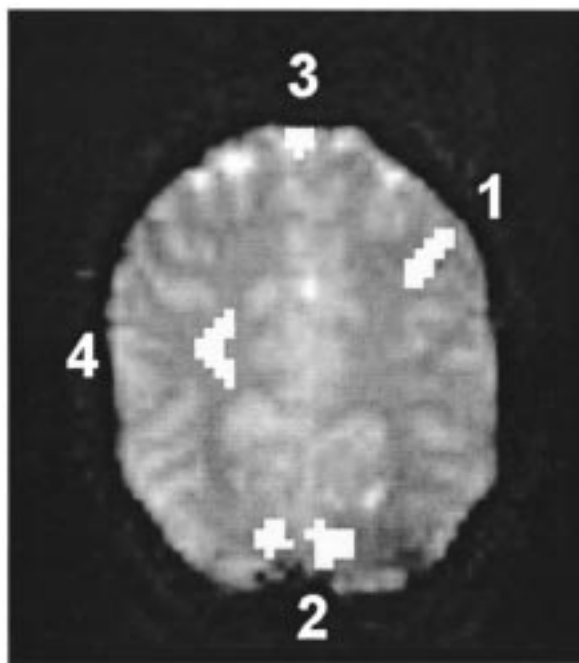


Fig. 1. Brain map showing regions of interest.

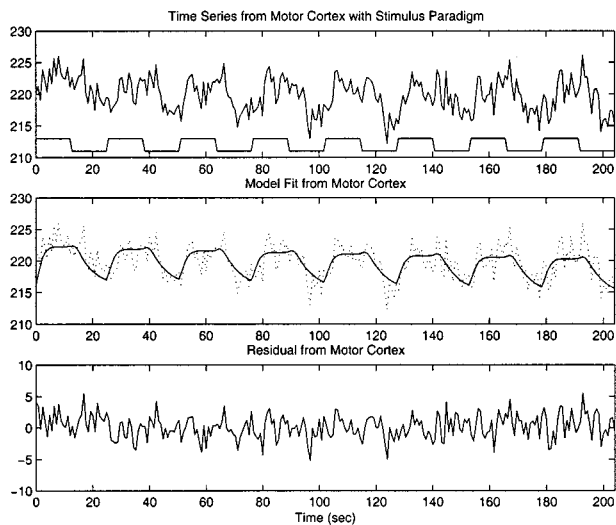


Fig. 2. Primary motor cortex time series, regularized fit and residuals.

the frequency response of the gamma hump and in Fig. 10 the AR(1) noise spectrum. In Fig. 11, we show activation and noise plots for the unregularized, regularized, and prefiltered analysis described above. The noise is plotted as the square-root of the power in the AR(1) portion of the noise signal. The noise estimates for the regularized case appear smoother than either the spatially prefiltered case or the unregularized case. In addition, the noise for the regularized case shows contrast between gray matter and white matter, coinciding with the notion that the noise comes from hemodynamic sources. In the unregularized and spatially prefiltered cases, this contrast is less clear. Because we are regularizing over just the noise, the activation maps for the regularized case have the same (in fact slightly better) high spatial resolution as the unregularized case, compared to the blurred activations in the spatially prefiltered case.

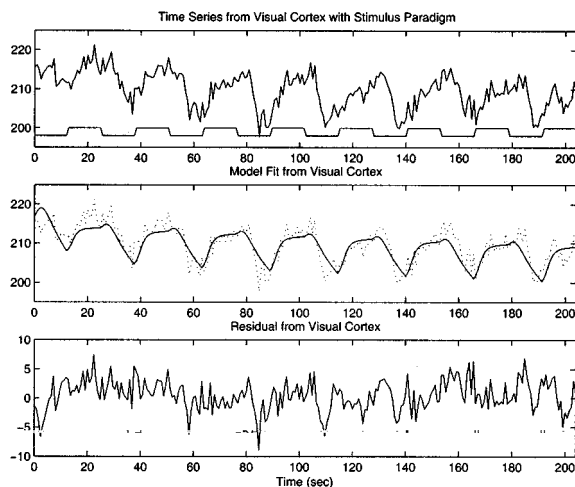


Fig. 3. Primary visual cortex time series, regularized fit and residuals.

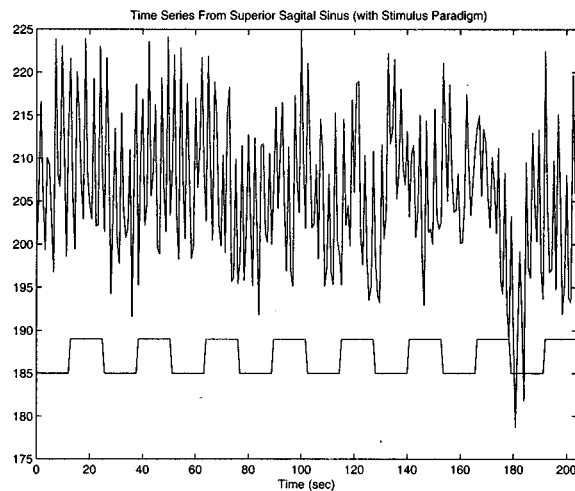


Fig. 4. Superior sagittal sinus time series.

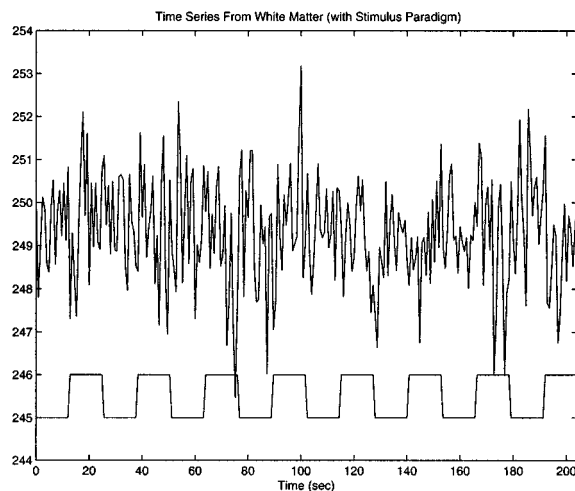


Fig. 5. White matter time series.

In Figs. 12 and 13, we give plots of the nure for the regularized procedure and for the method of [9]. We see that our regularized analysis produces a much better result. Finally, in Fig. 14

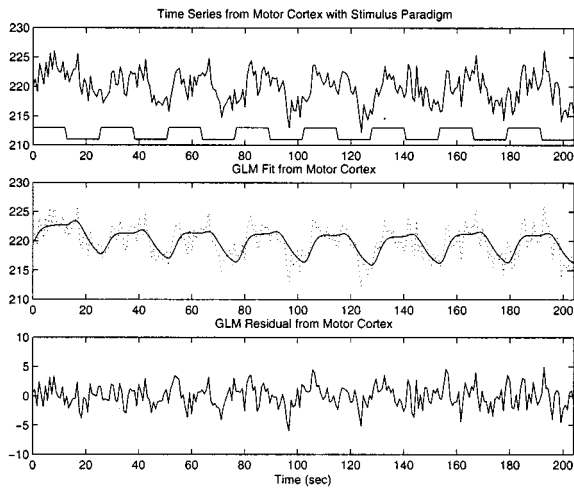


Fig. 6. Primary motor cortex time series, GLM fit and residuals.

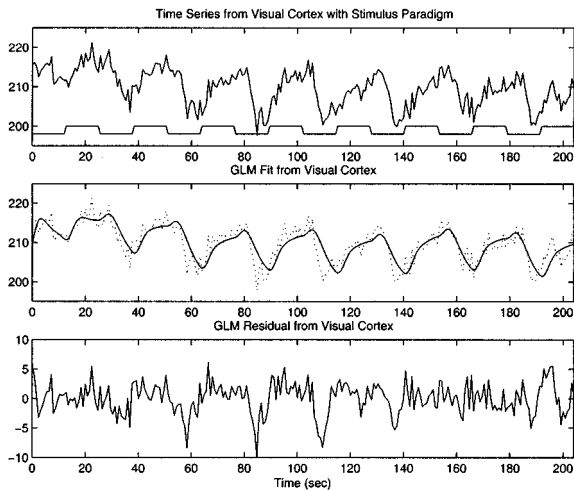


Fig. 7. Primary visual cortex time series, GLM fit and residuals.

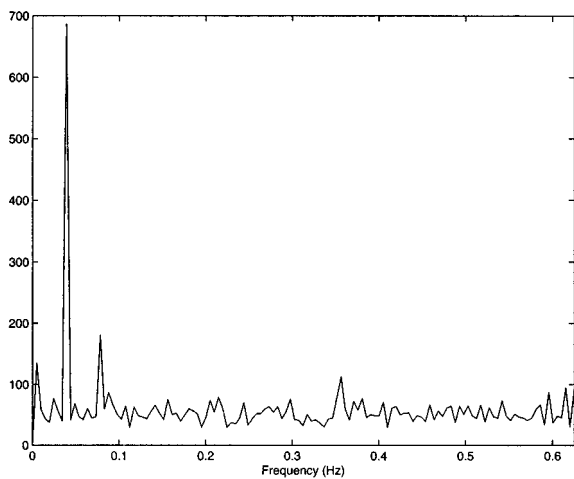


Fig. 8. Motion corrected signal.

we show some analysis of innovations. For LRST, the innovations signal is the error signal produced after fitting the AR(1) + white noise model to the residuals. For the GLM method, the innovations are the result of applying the GLM “whitening” filter to the residuals. We see that the GLM filter does a poor job,

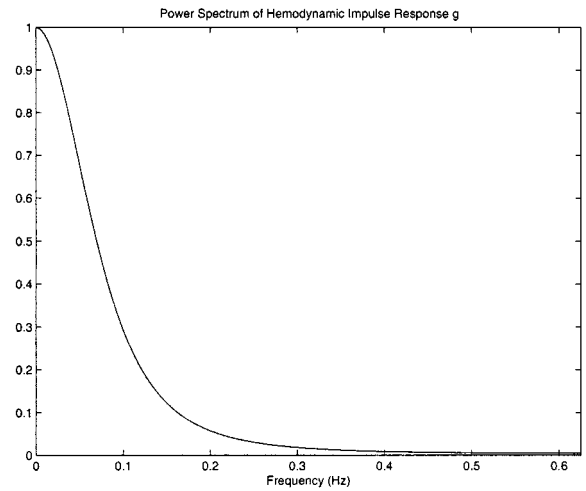


Fig. 9. Gamma frequency response.

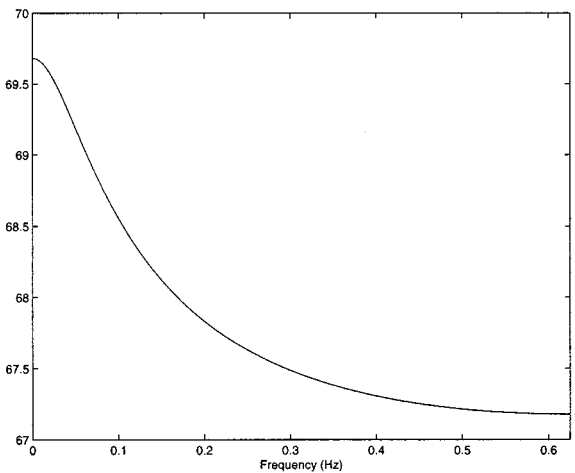


Fig. 10. AR(1) spectrum.

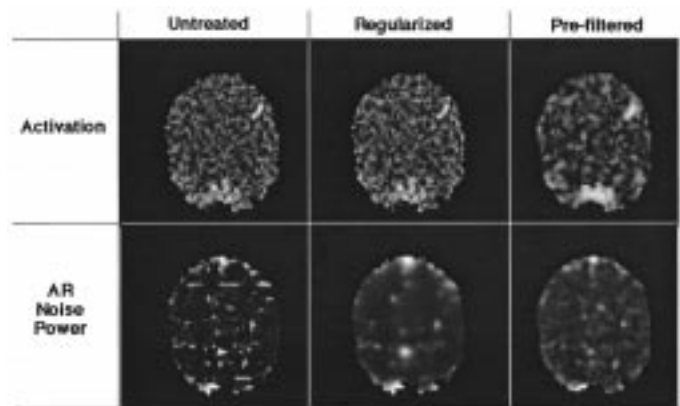


Fig. 11. Activation for each of three methods.

leaving much color in the innovations. The LRST innovations are by contrast essentially white as expected.

VI. CONCLUSION

In this paper, we have discussed the use of regularization to improve the modeling of fMRI time series. To avoid computational problems we have eschewed Tikhonov regulariza-

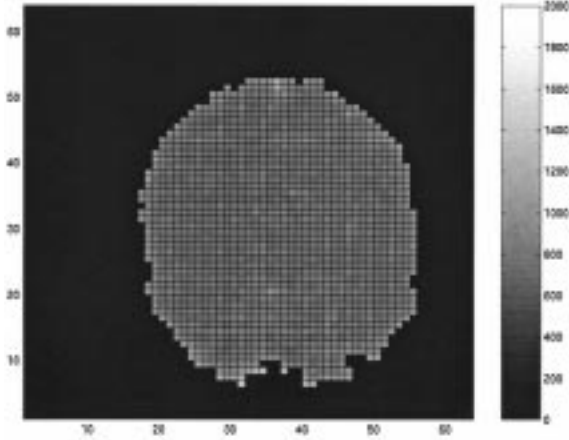


Fig. 12. Nure for LRST method.

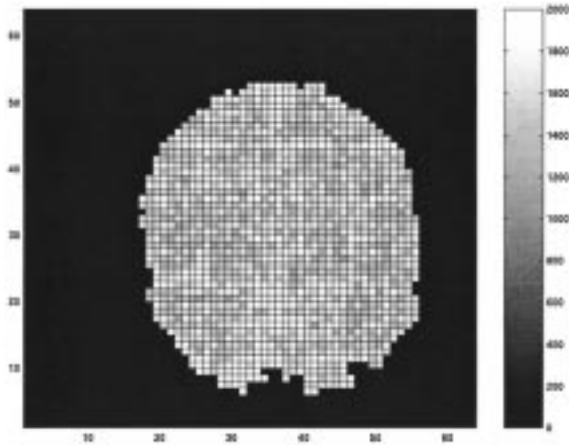


Fig. 13. Nure for GLM method.

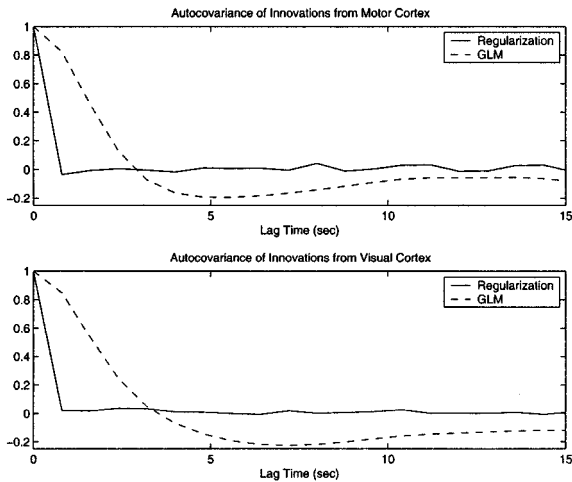


Fig. 14. ACF of innovations.

tion rather using local modeling. We found significant improvement in SNR due to regularization of the noise parameters. We also found significant improvements over prefiltering which has been a popular method in the fMRI literature. Effects on motion parameters (not reported here) were of smaller order. Effects on delay parameters will be reported elsewhere. We have also developed a discrepancy measure (nure) for comparing dif-

ferent estimation methods and used that to compare our LRST technique to the currently popular GLM method. Our method shows superior performance. Other topics for further study include; provision of standard errors for parameter and activation estimates; extensions to deal with single event data.

APPENDIX A COMPUTATIONAL DETAILS

The estimation of m_P decouples because it enters only the zero frequency term in (3.2). Firstly minimizing $J(\theta_P)$ with respect to m_P gives

$$\hat{m}_P = \Sigma_Q K_Q^h \text{Re} \left(\tilde{x}_{0,P-Q} - \tilde{\xi}_{0,P-Q}^T f_P \right) / n. \quad (7.1)$$

This nulls the first term in (3.5) and then \hat{m}_P can be calculated at the end, once f_P is found.

We now denote $\theta_P = [\beta_P^T, \alpha_P^T]^T$ to be the remaining parameters. $\beta_P = [\beta_P, f_P^T, D_P]^T$ and $\alpha_P = [\sigma_{\eta,P}^2, a_P]^T$. And we minimize $J_P(\theta_P)$ with respect to θ_P using cyclic descent, thus, the following:

given α_P get $\beta_P = [\beta_P, f_P^T, D_P]^T$;
given β_P get α_P .

Getting $\beta_P = [\beta_P, f_P^T, D_P]^T$: We proceed in two stages:

given b_P get D_P, f_P ;
given D_P, f_P get b_P .

Getting D_P, f_P : Set $\tilde{y}_{k,P}^o = e^{j\omega_k D_P} \tilde{y}_{k,P} = e^{j\omega_k D_P} (\tilde{x}_{k,P} - b_P u_k)$ so then

$$\hat{D}_P = \arg \cdot \max \sigma_{y_o \xi P}^T (\sigma_{\xi \xi P})^{-1} \sigma_{y_o \xi P} \quad (7.2)$$

i.e., plot the cross covariance for a lag corresponding to a maximum. Note that $\sigma_{\xi \xi P}$ only gets calculated once; only $\sigma_{y_o \xi P}$ is recalculated as we vary D_P . Once D_P is found then

$$f_P = (\sigma_{\xi \xi P})^{-1} \sigma_{y_o \xi P}. \quad (7.3)$$

But this need only be found at the last iteration.

Getting b_P : We introduce the unregularized estimator

$$b_P^o = \frac{\sigma_{uxP} - \sigma_{x_o \xi P}^T (\sigma_{\xi \xi P})^{-1} (\sigma_{u_o \xi P})}{\sigma_{uuP} - \sigma_{u_o \xi P} (\sigma_{\xi \xi P})^{-1} (\sigma_{u_o \xi P})^T} \quad (7.4)$$

and call Δ_P the denominator in that expression i.e.,

$$\Delta_P = \sigma_{uuP} - \sigma_{u_o \xi P} (\sigma_{\xi \xi P})^{-1} (\sigma_{u_o \xi P})^T.$$

The local constant estimator is simply

$$b_P = (\Sigma_Q \Delta_{P-Q} K_Q^h)^{-1} \Sigma_Q (\Delta_{P-Q} b_{P-Q}^o K_Q^h). \quad (7.5)$$

Getting α_P : Here an EM algorithm [25] proves most convenient and requires zero padding. Below a subscript zero denotes the current iteration while subscript one denotes next iteration. We first list the unregularized algorithm

$$\text{var}_P = \Sigma (|\tilde{w}_{0,k,P}|^2 / n + \Gamma_{0,k}) / n \quad (7.6)$$

$$a_{1,P} = \text{cov}_P / \text{var}_P$$

$$\text{cov}_P = \Sigma e^{j\omega_k} (|\tilde{w}_{0,k,P}|^2 / n + \Gamma_{0,k}) / n \quad (7.7)$$

$$\begin{aligned}
\sigma_{\eta,1,P}^2 &= \text{var}_P(1 - a_{1,P}^2) \\
\tilde{w}_{0,k,P} &= h_{k,P} \tilde{c}_{k,P} \\
\Gamma_{0,k} &= h_{k,P} \sigma_\nu^2 \\
h_{k,P} &= 1 - \frac{\sigma_\nu^2}{F_{0,k,P}} \\
F_{0,k,P} &= \sigma_\nu^2 + \frac{\sigma_{\eta,0,P}^2}{|1 - a_{0,P} e^{-j\omega_k}|^2}.
\end{aligned}$$

To estimate σ_ν^2 also the update equation is

$$\sigma_{\nu,1,P}^2 = \Sigma(|\tilde{c}_{k,P} - \tilde{w}_{0,k,P}|^2/n + \Gamma_{0,k})/n.$$

Turning now to spatial regularization, the algorithm is exactly as before except that now

$$\begin{aligned}
\text{var}_P &= \Sigma_Q \text{var}_{P-Q}^0 K_Q^h \\
\text{cov}_P &= \Sigma_Q \text{cov}_{P-Q}^0 K_Q^h
\end{aligned}$$

where var^0 , cov^0 denote the quantities in (7.6), (7.7).

APPENDIX B DERIVATIONS

It is clearer if we first develop the no-regularization case and then modify the results for regularization.

A. No Regularization

Getting D_P , f_P : Setting

$$\tilde{y}_{k,P} = (\tilde{x}_{k,P} - b_P u_k) = e^{-j\omega_k D_P} (\tilde{\xi}_{k,P}^T f_P) + v_{k,P}$$

the criterion (3.2) becomes (dropping momentarily the log term and the $1/2Mn^2$)

$$\begin{aligned}
J_P^0(\theta_P) &= \Sigma_1^{n-1} |\tilde{y}_{k,P} - e^{-j\omega_k D_P} \tilde{\xi}_{k,P}^T f_P|^2 / F_{k,P} \\
&= \Sigma_1^{n-1} |\tilde{y}_{k,P}^o - \tilde{\xi}_{k,P}^T f_P|^2 / F_{k,P} \\
\tilde{y}_{k,P}^o &= e^{j\omega_k D_P} \tilde{y}_{k,P}.
\end{aligned}$$

Optimizing this with respect to f_P gives (7.6). Using that, reduces $J_P^0(\theta_P)$ to

$$J_P^0(\theta_P) = \sigma_{yyP} - \sigma_{y\phi P}^T (\sigma_{\xi\xi P})^{-1} \sigma_{y\phi P} \quad (8.1)$$

and minimizing this with respect to D_P is the same as (7.5).

Getting b_P : We rewrite $J_P^0(\theta_P)$ as [with $(\tilde{x}_{k,P}^o, \tilde{u}_k^o) = e^{j\omega_k D_P} (\tilde{x}_{k,P}, \tilde{u}_k)$]

$$J_P^0(\theta_P) = \Sigma_1^{n-1} |\tilde{x}_{k,P}^o - b_P \tilde{u}_k^o - \tilde{\xi}_{k,P}^T f_P|^2 / F_{k,P}.$$

Given D_P , this is linear in b_P , f_P and so using a conditional expectation notation is given by

$$b_P = \text{cov}(u^o, x^o | \tilde{\xi}_P) / \text{var}(u^o | \tilde{\xi}_P)$$

and this is (7.4).

Getting $F_{k,P}$: For clarity, we drop the subscript P in this subsection. For the EM derivation, the data is $v = w + \nu$. The KL (Kullback Liebler) function (also called the Q function) for the EM algorithm is the conditional expectation given v of the log of the joint density of w and ν .

$$\text{KL}(\theta_1, \theta_0) = \text{KL}_\nu + \text{KL}_w.$$

Note, for later use, for a stationary signal y_t

$$\Sigma_1^n y_t^2 = \frac{1}{n} \Sigma_1^n |\tilde{y}_k|^2, \quad E|\tilde{y}_k|^2 \simeq n F_{y_k}$$

where F_{y_k} is a spectrum. And F_{wvk} denotes the cross spectrum between w , v . Also in this subsection successive iterates are denoted with indices respectively 0,1. Continuing we find

$$\begin{aligned}
\text{KL}_\nu &= \frac{1}{n\sigma_{\eta 1}^2} E_0 (\Sigma_1^n |\tilde{v}_k - \tilde{w}_k|^2 |\tilde{v}_k|) + n \ln \sigma_{\eta 1}^2 \\
\text{KL}_w &= \frac{1}{n\sigma_{\eta 1}^2} E_0 (\Sigma_1^n |\tilde{w}_k|^2 |1 - a_1 e^{j\omega_k}|^2 |\tilde{v}_k|) \\
&\quad + \Sigma_1^n \ln \left(\sigma_{\eta 1}^2 \frac{1}{|1 - a_1 e^{j\omega_k}|^2} \right).
\end{aligned}$$

Now KL_ν is free of unknown parameters. While

$$\begin{aligned}
\text{KL}_w &= \frac{1}{n\sigma_{\eta 1}^2} \Sigma_1^n (|\tilde{w}_{ok}|^2 + n\Gamma_{ok}) |1 - a_1 e^{j\omega_k}|^2 \\
&\quad + \Sigma_1^n \ln \left(\sigma_{\eta 1}^2 \frac{1}{|1 - a_1 e^{j\omega_k}|^2} \right)
\end{aligned}$$

where (with $F_{vk} = F_{0k}$)

$$\begin{aligned}
\tilde{w}_{ok} &= E_0(\tilde{w}_k | \tilde{v}_k) = F_{wvk} F_{ok}^{-1} \tilde{v}_k = F_{wk} F_{ok}^{-1} \tilde{v}_k \\
&= (F_{ok} - \sigma_\nu^2) F_{ok}^{-1} \tilde{v}_k = \left(1 - \frac{\sigma_\nu^2}{F_{ok}} \right) \tilde{v}_k \\
n\Gamma_{ok} &= \text{var}_0(\tilde{w}_k | \tilde{v}_k) = n(F_{wk} - F_{wvk} F_{ok}^{-1} F_{wvk}^*) \\
&= n(F_{wk} - F_{wk}^2 F_{ok}^{-1}) = nF_{wk} F_{ok}^{-1} (\sigma_\nu^2) \\
&= n \left(1 - \frac{\sigma_\nu^2}{F_{ok}} \right) \sigma_\nu^2.
\end{aligned}$$

Continuing we can collapse KL_w to

$$\text{KL}_w = (\text{var}_P(1 + a_1^2) - 2a_1 \text{cov}_P) \frac{n}{\sigma_{\eta 1}^2} + n \ln \sigma_{\eta 1}^2 \quad (8.2)$$

and minimizing this with respect to a_1 , $\sigma_{\eta 1}^2$ gives the quoted algorithm.

Turning to σ_ν^2 , on optimizing KL_ν we get

$$\begin{aligned}
\sigma_\nu^2 &= \Sigma \frac{|\tilde{v}_k - \tilde{w}_{ok}|^2}{n^2} + \Sigma \frac{\text{var}_o(\tilde{w}_{ok} | \tilde{v}_k)}{n^2} \\
&= \Sigma \frac{|\tilde{v}_k - \tilde{w}_{ok}|^2}{n^2} + \Sigma \frac{\Gamma_{ok}}{n}
\end{aligned}$$

which is the quoted expression.

B. Regularized Parameters

Given b_P optimize with respect to D_P , f_P and get the same results as in (7.2), (7.3). Then $J_P^0(\theta_P)$ collapses to (8.1) and then

$$\begin{aligned} J_P^0 &= \sigma_{yyP} - \sigma_{y\alpha P}^T (\sigma_{\xi\xi P})^{-1} \sigma_{y\alpha P} \\ &= \frac{1}{n} \sum_{k=1}^{n-1} |\tilde{x}_{k,P} - b_P \tilde{u}_k|^2 \\ &\quad - \frac{1}{n} \sum_{k=1}^{n-1} (\tilde{x}_{k,P}^o - b_P \tilde{u}_k^o) \tilde{\xi}_{k,P}^{*T} (\sigma_{\xi\xi P})^{-1} (\cdot) \\ &= \sigma_{xxP} - 2b_P \sigma_{uxP} + b_P^2 \sigma_{uuP} \\ &\quad - (\sigma_{x\alpha P} - b_P \sigma_{u\alpha P})^T (\sigma_{\xi\xi P})^{-1} (\sigma_{x\alpha P} - b_P \sigma_{u\alpha P}). \end{aligned}$$

The locally weighted criterion, thus, becomes

$$\Sigma K_Q^h J_{P-Q}^0(b_P)$$

and optimizing this on b_P gives (7.5).

Regularizing $F_{k,P}$: We use local log likelihood, so KL_w is now

$$KL_w = \Sigma_Q K_Q^h K_L^0_{w,P-Q}$$

and in the mask or neighborhood a_1 , $\sigma_{\eta,1}^2$ are fixed. Follow through the previous derivation and then get same algorithm but with var_P , cov_P as quoted.

REFERENCES

- [1] K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner, H. Cheng, T. J. Brady, and B. R. Rosen, "Dynamic MRI of human brain activity during primary sensory stimulation," in *Proc. Nat. Acad. Sci.*, vol. 89, 1992, pp. 5675–5679.
- [2] S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, and K. Ugurbil, "Intrinsic signal changes accompanying sensory stimulation," in *Proc. Nat. Acad. Sci.*, vol. 89, 1992, pp. 5951–5955.
- [3] M. S. Cohen and S. Y. Bookheimer, "Localization of brain function using MRI," *Trends Neurosci.*, vol. 17, pp. 208–276, 1994.
- [4] A. M. Dale and R. L. Buckner, "Selective averaging of rapidly presented individual trials using fMRI," *Human Brain Mapping*, vol. 5, pp. 1–12, 1997.
- [5] B. Rosen, R. L. Buckner, and A. M. Dale, "Event-related fMRI: Past, present and future," in *PNAS*, vol. 95, 1998, pp. 773–780.
- [6] N. Lange and S. L. Zeger, "Non-linear Fourier timeseries analysis for human brain mapping by functional MRI," *Appl. Stat.*, vol. 46, pp. 1–29, 1997.
- [7] V. Solo, E. Brown, and R. Weisskoff, "A signal processing approach to functional MRI," in *Proc. IEEE ICIP97*, vol. II, 1997, pp. 121–124.
- [8] V. Solo, "Transfer function order estimation with a H_∞ criterion," presented at the IEEE CDC 98, 1998.
- [9] K. J. Worsley and K. J. Friston, "Analysis of fMRI time-series revisited—Again," *Neuroimage*, vol. 2, pp. 173–181, 1995.
- [10] P. P. Mitra and B. Pesaran, "Analysis of dynamic brain imaging data," Bell Labs, Tech. Rep., Murray Hill, NJ, 1997.
- [11] E. E. Ruttiman, M. Unser, R. R. Dawlings, D. Roi, N. F. Ramsey, V. S. Mattay, D. W. Hommer, J. A. Frank, and D. R. Weinberger, "Statistical analysis of functional MRI data in the wavelet domain," *IEEE Trans. Med. Imag.*, vol. 17, pp. 142–152, Apr. 1998.
- [12] X. Descombes, F. Kruggel, and D. von Cramon, "Spatio-temporal fMRI analysis using Markov random fields," *IEEE Trans. Med. Imag.*, vol. 17, pp. 1028–1039, Dec. 1998.
- [13] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time series," *Human Brain Map.*, vol. 1, pp. 153–171, 1994.
- [14] R. M. Weisskoff, J. Baker, J. Belliveau, T. L. Davis, K. K. Kwong, M. S. Cohen, and B. R. Rosen, "Power spectrum analysis of functionally-weighted MR data: What's in the noise," in *Proc. Soc. Magnetic Resonance in Medicine*, 1993, pp. 1–7.
- [15] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger, "Linear systems analysis of functional MRI in human V1," *J. Neurosci.*, vol. 16, pp. 4207–4221, 1996.
- [16] J. Mandeville, J. Marota, B. Kosofsky, J. Keltner, R. Weissleder, B. Rosen, and R. Weisskoff, "Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation," *Magn. Res. Med.*, 1998, to be published.
- [17] P. Purdon, V. Solo, E. Brown, and R. Weisskoff, "Locally regularized spatio-temporal modeling and model comparison for functional MRI," *Neuroimage*, 1999, submitted for publication.
- [18] D. R. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden-Day, 1981.
- [19] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. New York, NY: Wiley, 1977.
- [20] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Applications*. London, U.K.: Chapman and Hall, 1996.
- [21] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *DARPA Image Understanding Workshop: DARPA*, 1981, pp. 121–130.
- [22] L. Ng and V. Solo, "A data-driven method for choosing smoothing parameters in optical flow problems," in *Proc. IEEE ICIP97*, 1997, pp. III:360–363.
- [23] H. Linhart and W. Zucchini, *Model Selection*. New York: Wiley, 1986.
- [24] S. Konishi and G. Kitagawa, "Generalized information criteria in model selection," *Biometrika*, vol. 83, pp. 875–890, 1996.
- [25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.