March 1994

LIDS-P-2234

Multiscale Smoothing Error Models[1]

Mark R. Luettgen[2]

Alan S. Willsky[3]

Abstract

A class of multiscale stochastic models based on scale-recursive dynamics on trees has recently been introduced. These models are interesting because they can be used to represent a broad class of physical phenomena and because they lead to efficient algorithms for estimation and likelihood calculation. In this paper, we provide a complete statistical characterization of the error associated with smoothed estimates of the multiscale stochastic processes described by these models. In particular, we show that the smoothing error is itself a multiscale stochastic process with parameters which can be explicitly calculated.

[1]This work was supported by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002, by the Office of Naval Research under Grant N00014-91-J-1004 and by the Army Research Office under Grant DAAL03-92-G-0115.

[2]Please send all correspondence to this author at: Alphatech, Inc., 50 Mall Road, Burlington, MA 01803. Email: luettgen@alphatech.com. Tel: 617-273-3388 x224. Fax: 617-273-9345.

[3]Room 35-433, MIT Laboratory for Information and Decision Systems, Cambridge, MA 02139. Email: willsky@mit.edu. Tel: 617-253-2356. Fax: 617-258-8553.

1 Introduction

A class of multiscale models describing stochastic processes indexed by the nodes of a tree has recently been introduced in [1, 2]. These models can be used to capture a surprisingly rich class of physical phenomena. For instance, experimental results in [2] illustrate that they can be used to model the statistical self-similarity exhibited by stochastic processes with generalized power spectra of the form $1/f^{\beta}$, and in [3] we describe how they can be used to represent *any* 1-D Markov process or 2-D Markov random field. Moreover, this class of models leads to efficient algorithms for estimation and likelihood calculation and as a result provides a useful framework for a variety of signal and image processing problems [1, 2, 4, 5, 6].

Knowledge of the error statistics of smoothed estimates of such processes is essential for the development of a number of important new applications, including for instance so-called mapping problems [7], the multiscale counterpart to the model validation problem in [8], and certain oceanographic problems [9]. Several such applications have been developed in the context of 1-D Gauss-Markov models by exploiting relatively recent results which show that the smoothing error processes associated with Gauss-Markov models are themselves Gauss-Markov processes [7, 8, 10, 11][4]. In this paper, we derive a dynamic model for the smoothing error process associated with multiscale stochastic models. In particular, we show that the smoothing error is itself a multiscale stochastic process with parameters which can be explicitly computed. These results generalize previous results for Gauss-Markov processes, since these processes correspond to a degenerate form of the multiscale models, and provide the necessary framework for applications such as those mentioned above.

This paper is organized as follows. In Section 2 we briefly review the class of multiscale stochastic

[4]More generally, Levy et al. [12] have recently shown that the smoothing error processes associated with the class of Gaussian reciprocal processes, which contains the class of Gauss-Markov processes, are themselves Gaussian reciprocal. See also [13] for similar results corresponding to 2-D Gauss-Markov random fields.

models of interest here and the scale-recursive estimation algorithm associated with them. In Section 3 we derive a multiscale model for the smoothing error process.

2 Multiscale Stochastic Modeling and Optimal Estimation

The models presented in this section describe multiscale Gaussian stochastic processes indexed by nodes on a *tree*. A q^{th} order tree is a pyramidal structure of nodes connected such that each node of the tree has q offspring (see Figure 1). We denote nodes on the tree with an abstract index s, and define an upward (fine-to-coarse) shift operator $\bar{\gamma}$ such that $s\bar{\gamma}$ is the parent of node s. We also define a corresponding set of downward shift operators $\alpha_1, \cdots, \alpha_q$ such that $s\alpha_1, \cdots, s\alpha_q$ are the offspring of node s. In addition, we denote the set of nodes on the tree as $\mathcal{T}$ and the set of nodes which includes node s and all of its descendants as $\mathcal{T}_s$, i.e. $\mathcal{T}_s = \{\sigma | \sigma = s \text{ or } \sigma \text{ is a descendant of } s\}$. Also, the complement of $\mathcal{T}_s$ is denoted $\mathcal{T}_s^c$. The statistical characterization of model state $x(s) \in \mathcal{R}^n$ is then given by:

$$x(s) \quad = \quad A(s)x(s\bar{\gamma}) + B(s)w(s) \tag{1}$$

under the assumptions that $x(0) \sim \mathcal{N}(0, P(0))$, $w(s) \sim \mathcal{N}(0, I)$, $A(s)$ and $B(s)$ are matrices of appropriate size, and $s = 0$ is the root node at the top of the tree. The driving noise $w(s) \in \mathcal{R}^m$ is white, i.e. $w(s)$ and $w(\sigma)$ are independent if $s \neq \sigma$, and independent of the initial condition $x(0)$.

The class of models (1) has a statistical structure that can be exploited to develop efficient signal processing algorithms. In particular, note that any given node on the q^{th}-order tree can be viewed as a boundary between $q+1$ subsets of nodes (q corresponding to paths leading towards offspring and one corresponding to a path leading towards a parent). An important property of the model (1) is that, conditioned on the value of the state at any node, the values of the state corresponding to the $q+1$ corresponding subsets of nodes are independent. This fact is the basis for the development in [1, 2] of an algorithm for computing smoothed estimates of $x(s)$ based on noisy measurements $y(s) \in \mathcal{R}^p$ of the form:

$$y(s) \quad = \quad C(s)x(s) + v(s) \tag{2}$$

where $v(s) \sim \mathcal{N}(0, R(s))$, and is independent of both $w(s)$ and $x(0)$. The algorithm for computing the smoothed estimates of $x(s)$ is a generalization to q^{th}-order trees of the well-known Rauch-Tung-Striebel algorithm for smoothing 1-D Gauss-Markov processes. We briefly review this algorithm next, and then derive a general model for the error associated with the smoothed estimates.

We denote the set of states defined at nodes in $\mathcal{T}_s$ as X_s, i.e. $X_s = \{x(\sigma)\}_{\sigma \in \mathcal{T}_s}$, and similarly $Y_s = \{y(\sigma)\}_{\sigma \in \mathcal{T}_s}$. The set of measurements in the subtree strictly below s is denoted $Y_s^{\alpha_q}$, i.e. $Y_s^{\alpha_q} = \{y(\sigma) | \sigma \text{ is a descendant of } s\}$. We also define $\hat{x}(s|Y)$ as the expected value of $x(s)$ given measurements in the set Y and the corresponding error covariance as $P(s|Y)$.

The upward sweep of the smoothing algorithm begins with the initialization of $\hat{x}(s|Y_s^{\alpha_q})$ and $P(s|Y_s^{\alpha_q})$ at the finest level. In particular, for every s at this finest scale we set $\hat{x}(s|Y_s^{\alpha_q})$ to zero and $P(s|Y_s^{\alpha_q})$ to the solution at the finest level of the tree of the Lyapunov equation:

$$P(s) \quad = \quad A(s)P(s\bar{\gamma})A^T(s) + B(s)B^T(s) \tag{3}$$

where $P(s)$ denotes the covariance of the process $x(s)$ at node s. Suppose then that we have $\hat{x}(s|Y_s^{\alpha_q})$ and $P(s|Y_s^{\alpha_q})$ at a given node s. This estimate is *updated* to incorporate the measurement $y(s)$ according to the following:

$$\hat{x}(s|Y_s) \quad = \quad \hat{x}(s|Y_s^{\alpha_q}) + K(s)[y(s) - C(s)\hat{x}(s|Y_s^{\alpha_q})] \tag{4}$$

$$P(s|Y_s) \quad = \quad [I - K(s)C(s)]P(s|Y_s^{\alpha_q}) \tag{5}$$

where $K(s) = P(s|Y_s^{\alpha_q})C^T(s)[C(s)P(s|Y_s^{\alpha_q})C^T(s) + R(s)]^{-1}$.

Suppose next that we have the updated estimates $\hat{x}(s\alpha_i|Y_{s\alpha_i})$ at all of the immediate descendants of node s. The next step involves the use of these estimates to predict $x(s)$ at the next coarser scale, i.e. to compute $\hat{x}(s|Y_{s\alpha_i})$. Using the following *upward* model for the multiscale process [1, 2]:

$$x(s\bar{\gamma}) = F(s)x(s) + \bar{w}(s) \tag{6}$$

with the measurement equation again given by (2), and where $F(s) = P(s\bar{\gamma})A^T(s)P(s)^{-1}$ and $\mathrm{E}[\bar{w}(s)\bar{w}^T(s)] = P(s\bar{\gamma}) - P(s\bar{\gamma})A^T(s)P(s)^{-1}A(s)P(s\bar{\gamma}) \equiv Q(s)$, we compute the fine-to-coarse *predicted* estimates:

$$\hat{x}(s|Y_{s\alpha_i}) = F(s\alpha_i)\hat{x}(s\alpha_i|Y_{s\alpha_i}) \tag{7}$$

$$P(s|Y_{s\alpha_i}) = F(s\alpha_i)P(s\alpha_i|Y_{s\alpha_i})F^T(s\alpha_i) + Q(s\alpha_i) \tag{8}$$

The estimates $\hat{x}(s|Y_{s\alpha_i}), i = 1, \cdots, q$ are then *merged* to obtain

$$\hat{x}(s|Y_s^{\alpha_q}) = P(s|Y_s^{\alpha_q})\sum_{i=1}^{q} P^{-1}(s|Y_{s\alpha_i})\hat{x}(s|Y_{s\alpha_i}) \tag{9}$$

$$P(s|Y_s^{\alpha_q}) = [(1-q)P(s)^{-1} + \sum_{i=1}^{q} P^{-1}(s|Y_{s\alpha_i})]^{-1} \tag{10}$$

The recursion proceeds up the tree until one obtains the smoothed estimate of the root node, $\hat{x}(0|Y_0)$. This estimate initializes a *downward sweep* in which $\hat{x}(s|Y_0)$ is computed according to

$$\hat{x}(s|Y_0) = \hat{x}(s|Y_s) + J(s)[\hat{x}(s\bar{\gamma}|Y_0) - \hat{x}(s\bar{\gamma}|Y_s)] \tag{11}$$

$$P(s|Y_0) = P(s|Y_s) + J(s)[P(s\bar{\gamma}|Y_0) - P(s\bar{\gamma}|Y_s)]J^T(s) \tag{12}$$

$$J(s) = P(s|Y_s)F^T(s)P^{-1}(s\bar{\gamma}|Y_s) \tag{13}$$

Note that (12) characterizes the smoothing error covariance at any given lattice site s, but does not provide information about the correlation structure of the error process. The goal in the next section is to provide a multiscale model for the smoothing error process, i.e. to show that the error satisfies a recursion of the form (1), and to calculate the associated model parameters. This then provides the complete statistical characterization of the smoothing error that we seek.

3 Multiscale Smoothing Error Models

Given two nodes s and $\sigma \in \mathcal{T}_s^c$ on the tree, we can always represent $x(\sigma)$ in terms of $x(s\bar{\gamma})$ and an additive noise term $\varphi_{\sigma,s\bar{\gamma}}$:

$$x(\sigma) = \Phi_{\sigma,s\bar{\gamma}} x(s\bar{\gamma}) + \varphi_{\sigma,s\bar{\gamma}} \tag{14}$$

by tracing a path from σ to $s\bar{\gamma}$ and using the upward dynamics (6) and downward dynamics (1) to eliminate state variables along the way. The state transition matrix $\Phi_{\sigma,s\bar{\gamma}}$ is a function of the upward and downward prediction matrices A and F along the path, whereas $\varphi_{\sigma,s\bar{\gamma}}$ is a linear function of the upward and downward driving noises w and $\bar{w}$. For instance, the state $x(s\alpha_i)$ at the i^{th} offspring of s can be written in terms of the state $x(s\alpha_j)$ at the j^{th} offspring as:

$$x(s\alpha_i) = [A(s\alpha_i)F(s\alpha_j)]x(s\alpha_j) + [A(s\alpha_i)\bar{w}(s\alpha_j) + B(s\alpha_i)w(s\alpha_i)] \tag{15}$$

By construction, $\varphi_{\sigma,s\bar{\gamma}}$ is independent of the set of states $x(s\bar{\gamma}) \cup X_s$, as well as the corresponding set of measurements $y(s\bar{\gamma}) \cup Y_s$. This implies that $\hat{x}(\sigma|Y_s) = \Phi_{\sigma,s\bar{\gamma}}\hat{x}(s\bar{\gamma}|Y_s)$ which, using (14), implies that:

$$\tilde{x}(\sigma|Y_s) = \Phi_{\sigma,s\bar{\gamma}}\tilde{x}(s\bar{\gamma}|Y_s) + \varphi_{\sigma,s\bar{\gamma}} \tag{16}$$

where we have defined the error in $\hat{x}(s|Y)$ as $\tilde{x}(s|Y) \equiv x(s) - \hat{x}(s|Y)$. As a result, we see that $\tilde{x}(s|Y_s)$ has the following Markov property:

$$\begin{aligned}
\mathbf{E}\{\tilde{x}(s|Y_s)|\tilde{x}(\sigma|Y_s), \sigma \in \mathcal{T}_s^c\} &= \mathbf{E}\{\tilde{x}(s|Y_s)|\tilde{x}(s\bar{\gamma}|Y_s), \{\varphi_{\varsigma,s\bar{\gamma}}, \varsigma \in \mathcal{T}_s^c\}\} \\
&= \mathbf{E}\{\tilde{x}(s|Y_s)|\tilde{x}(s\bar{\gamma}|Y_s)\} + \mathbf{E}\{\tilde{x}(s|Y_s)|\{\varphi_{\varsigma,s\bar{\gamma}}, \varsigma \in \mathcal{T}_s^c\}\} \\
&= \mathbf{E}\{\tilde{x}(s|Y_s)|\tilde{x}(s\bar{\gamma}|Y_s)\}
\end{aligned} \tag{17}$$

The first equality in (17) follows from (16), the second from the orthogonality of $\varphi_{\sigma,s\bar{\gamma}}$ to $x(s\bar{\gamma})$ and Y_s, and the last from the orthogonality of $\varphi_{\sigma,s\bar{\gamma}}$ to $x(s)$ and Y_s. Now, using the upward dynamics (6), the upward sweep prediction equation (7) and standard linear least squares formulae we can write:

$$\tilde{x}(s|Y_s) = J(s)\tilde{x}(s\bar{\gamma}|Y_s) + \check{w}(s) \tag{18}$$

where $J(s)$ is given by (13) and where, from (17), $\check{w}(s)$ is *independent* of $\{\tilde{x}(\sigma|s)\}_{\sigma\in\mathcal{T}_s^c}$, and has covariance:

$$P(s|Y_s) - P(s|Y_s)F^T(s)P^{-1}(s\bar{\gamma}|Y_s)F(s)P(s|Y_s) \tag{19}$$

Next, note that the independence of $\check{w}(s)$ and $\{\tilde{x}(\sigma|Y_s)\}_{\sigma\in\mathcal{T}_s^c}$ implies that $\check{w}(s)$ is also independent of the *residual* information about $x(s)$ which is contained in the set of all available measurements Y_0, but *not* contained in Y_s. In particular, at each node in $\mathcal{T}_s^c$ a residual component $\nu_s(\sigma)$ which is orthogonal to the measurements in the set Y_s can be defined as:

$$\begin{aligned} \nu_s(\sigma) &= y(\sigma) - \mathrm{E}\{y(\sigma)|Y_s\} \\ &= C(\sigma)\tilde{x}(\sigma|Y_s) + v(\sigma) \end{aligned} \tag{20}$$

Denoting $\nu_s \equiv \{\nu_s(\sigma)\}_{\sigma\in\mathcal{T}_s^c}$, it is clear that span Y_0 = span $\{Y_s, \nu_s\}$, that $\nu_s \perp Y_s$ and that $\nu_s \perp \check{w}(s)$. Taking the expected value of both sides of (18) conditioned on ν_s, we obtain:

$$\mathrm{E}\{\tilde{x}(s|Y_s)|\nu_s\} = J(s)\mathrm{E}\{\tilde{x}(s\bar{\gamma}|Y_s)|\nu_s\} \tag{21}$$

Finally, noting that

$$\hat{x}(s|Y_0) = \hat{x}(s|Y_s) + \mathrm{E}\{\tilde{x}(s|Y_s)|\nu_s\} \tag{22}$$

and then subtracting (21) from (18) results in:

$$\tilde{x}(s|Y_0) = J(s)\tilde{x}(s\bar{\gamma}|Y_0) + \check{w}(s) \tag{23}$$

which is a multiscale model for the smoothing error of precisely the same form as (1).

This model is, of course, consistent with the error covariance computation in (12). In particular, using the Lyapunov equation for (23) we obtain:

$$P(s|Y_0) = J(s)P(s\bar{\gamma}|Y_0)J^T(s) + P(s|Y_s) - P(s|Y_s)F^T(s)P^{-1}(s\bar{\gamma}|Y_s)F(s)P(s|Y_s)$$

$$= P(s|Y_s) + J(s)[P(s\bar{\gamma}|Y_0) - P(s\bar{\gamma}|Y_s)]J^T(s) \tag{24}$$

In addition, on first-order trees, the model (1) reduces to a standard Gauss-Markov model, and hence (23) generalizes to q^{th}-order trees the corresponding 1-D time-series result. The derivation here is related to, but is in fact substantially simpler than, the derivation based on backwards prediction error models in [8].

References

[1] K. Chou, A. Willsky, A. Benveniste, and M. Basseville, "Recursive and iterative estimation algorithms for multiresolution stochastic processes," in *Proceedings of the 28th IEEE Conference on Decision and Control*, (Tampa, FL), pp. 1184–1189, December 1989.

[2] K. Chou, A. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion and regularization," Tech. Rep. LIDS-P-2085, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 1991. To appear in *IEEE Transactions on Automatic Control*, April 1994.

[3] M. Luettgen, W. Karl, A. Willsky, and R. Tenney, "Multiscale representations of Markov random fields," Tech. Rep. LIDS-P-2130, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 1992. To appear in *IEEE Transactions on Signal Processing*, December 1993.

[4] M. Luettgen, W. Karl, and A. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," Tech. Rep. LIDS-P-2115, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 1992. To appear in *IEEE Transactions on Image Processing*, January 1994.

[5] M. Luettgen and A. Willsky, "Likelihood calculation for a class of multiscale stochastic models, with applications to texture discrimination," Tech. Rep. LIDS-P-2186, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 1993. Submitted to *IEEE Transactions on Image Processing.*

[6] M. Luettgen and A. Willsky, "Likelihood calculation for a class of multiscale stochastic models," in *Proceedings of the 32nd IEEE Conference on Decision and Control*, (San Antonio,

TX), December 1993. To appear.

[7] M. Bello, A. Willsky, B. Levy, and D. Castanon, "Smoothing error dynamics and their use in the solution of smoothing and mapping problems," *IEEE Transactions on Information Theory*, vol. 32, pp. 483–495, 1986.

[8] M. Bello, A. Willsky, and B. Levy, "Construction and applications of discrete-time smoothing error models," *International Journal of Control*, vol. 50, pp. 203–223, 1989.

[9] P. Fieguth, W. Karl, and A. Willsky, "Multiresolution statistical estimation of North Pacific ocean height from Topax/Poseiden satellite altimetry," in *Proceedings of the 1994 SPIE Conference on Neural and Stochastic Methods in Image and Signal Processing*, (San Diego, CA), July 1994. To appear.

[10] F. Badawi, A. Lindquist, and M. Pavon, "A stochastic realization approach to the smoothing problem," *IEEE Transactions on Automatic Control*, vol. 24, pp. 878–887, 1979.

[11] R. Ackner and T. Kailath, "Discrete-time complementary models and smoothing," *International Journal of Control*, vol. 49, pp. 1665–1682, 1989.

[12] B. Levy, R. Frezza, and A. Krener, "Modeling and estimation of discrete time Gaussian reciprocal processes," *IEEE Transactions on Automatic Control*, vol. 35, pp. 1013–1023, 1990.

[13] B. Levy, "Non-causal estimation for Markov random fields," in *Proceedings of the International Symposium MTNS-89, Vol. 1: Realization and Modeling in System Theory*, (Basel), Birkhauser-Verlag, 1990.

List of Figures

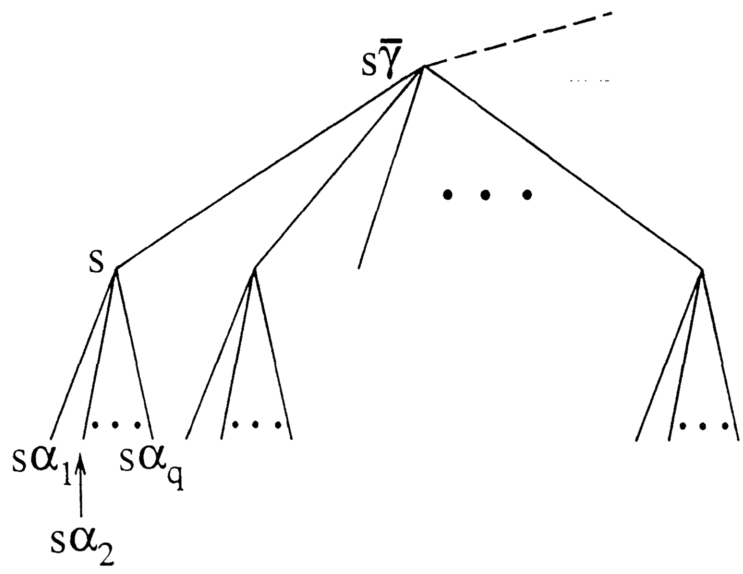

Figure 1: Multiscale stochastic processes are indexed by the q^{th}-order tree. The parent of a node s on the tree is denoted $s\bar{\gamma}$, and its q offspring are denoted $s\alpha_1, \cdots, s\alpha_q$.