

MULTISCALE STATISTICAL SIGNAL PROCESSING: STOCHASTIC PROCESSES INDEXED BY TREES

M. Basseville, A. Benveniste, K. C. Chou, A. S. Willsky

Abstract

Motivated by the recently-developed theory of multiscale signal models and wavelet transforms, we introduce stochastic dynamic models evolving on homogeneous trees. In particular we introduce and investigate both AR and state models on trees. Our analysis yields generalizations of Levinson and Schur recursions and of Kalman filters, Riccati equations, and Rauch-Tung-Striebel smoothing.

1. MULTISCALE REPRESENTATIONS AND HOMOGENEOUS TREES

The recently-introduced theory of multiscale representations and wavelet transforms [4] provides a sequence of approximations of signals at finer and finer scales. In 1-D a signal $f(x)$ is represented at the m th scale by a sequence $f(m, n)$ which provides the amplitudes of time-scaled pulses located at the points $n2^{-m}$. The progression from one scale to the next thus introduces twice as many points and indeed provides a tree structure with the pair $(2^{-m}, n)$ at one scale associated with $(2^{-(m+1)}, 2n)$ and $(2^{-(m+1)}, 2n + 1)$ at the next. This provides the motivation for the development of a system and stochastic process theory when the index set is taken to be a homogeneous dyadic tree. In this paper we outline some of the basic ideas behind our work.

Let \mathcal{T} denote the index set of the tree and we use the single symbol t for nodes on the tree. The scale associated with t is denoted by $m(t)$, and we write $s \preceq t$ ($s \prec t$) if $m(s) \leq m(t)$ ($m(s) < m(t)$). We also let $d(s, t)$ denote the distance between s and t , and $s \wedge t$ the common "parent" node of s and t (e.g. $(2^{-m}, n)$ is the parent of $(2^{-(m+1)}, 2n)$ and $(2^{-(m+1)}, 2n + 1)$). In analogy with the shift operator z^{-1} used as the basis for describing discrete-time dynamics we also define several shift operators on the tree: 0 , the identity operator (no move); γ^{-1} , the fine-to-coarse shift (e.g. from $(2^{-(m+1)}, 2n$ or $2n + 1)$ to $(2^{-m}, n)$); α , the left coarse-to-fine shift ($(2^{-m}, n)$ to

$(2^{-(m+1)}, 2n)$); β , the right coarse-to-fine shift $((2^{-m}, n)$ to $(2^{-(m+1)}, 2n + 1)$); and δ , the exchange operator $((2^{-(m+1)}, 2n) \longleftrightarrow (2^{-(m+1)}, 2n + 1))$. Note that 0 and δ are **isometries** in that they are one-to-one, onto maps of \mathcal{T} that preserve distances. Also we have the relations

$$\delta^2 = \gamma^{-1}\alpha = \gamma^{-1}\beta = 0, \gamma^{-1}\delta = \gamma^{-1}, \delta\beta = \alpha \quad (1.1)$$

It is possible to code all points on the tree via shifts from an arbitrary origin node, i.e. as wt_0 , $w \in \mathcal{L}$, where

$$\mathcal{L} = (\gamma^{-1})^* \cup \{\alpha, \beta\}^* \delta (\gamma^{-1})^* \cup \{\alpha, \beta\}^* \quad (1.2)$$

The **length** of a word w is denoted $|w|$ and equals $d(wt, t)$ (e.g. $|\gamma^{-1}| = 1$, $|\delta| = 2$). Also, since we will be interested in coarse-to-fine dynamic models, we define some notation for **causal** moves:

$$w \preceq 0 \ (w \prec 0) \ \text{if} \ wt \preceq t \ (wt \prec t) \quad (1.3)$$

We also define for convenience the move $\delta^{(n)}$ which exchanges the n th bit:

$$\begin{aligned} \text{If } t = \alpha\gamma^{-1}t, \text{ then } \delta^{(n)}t &= \alpha\delta^{(n-1)}\gamma^{-1}t \\ \text{If } t = \beta\gamma^{-1}t, \text{ then } \delta^{(n)}t &= \beta\delta^{(n-1)}\gamma^{-1}t \end{aligned} \quad (1.4)$$

Finally, we introduce formal power series notation and operations.

$$S = \sum_{w \in \mathcal{L}} s_w \cdot w \quad (1.5)$$

$$\gamma S = \sum_{w \in \mathcal{L}} s_{\gamma^{-1}w} \cdot w \quad (1.6)$$

$$\delta^{(k)} S = \sum_{w \in \mathcal{L}} s_{\gamma^k w} \cdot w \quad (1.7)$$

2. MODELING OF ISOTROPIC PROCESSES ON TREES

A zero-mean process Y_t , $t \in \mathcal{T}$ is **isotropic** if

$$E[Y_t Y_s] = r_{d(t,s)} \quad (2.1)$$

i.e. if its second-order statistics are invariant under any isometry of \mathcal{T} . These processes have been the subject of some study, and a Bochner-like spectral theorem has been developed [1,2]. However, many questions remain including an explicit criterion

for a sequence r_n to be the covariance of such a process and the representations of isotropic processes as outputs of systems driven by white noise. Note first that the sequence $\{Y_{\gamma^{-n}t}\}$ is an ordinary time series so that r_n must be positive semidefinite; however, the constraints of isotropy require even more. To uncover this structure we seek here the characterization of the class of autoregressive (AR) models where an AR model of order p has the form

$$Y_t = \sum_{\substack{w \leq 0 \\ |w| \leq p}} a_w Y_{wt} + \sigma W_t \quad (2.2)$$

where W_t is a white noise with unit variance. Note that this model is “causal” - i.e. it has a coarse-to-fine direction of propagation - since $w \leq 0$. Also, a first thought might be to examine models with strict past dependence, i.e. Y_t a function of $W_{\gamma^{-n}t}$; however as shown in [2], the constraints of isotropy allow us to show that only AR(1) has such dependence. Thus we have that AR(p) involves a full set of 2^{p-1} a_w 's and one σ so that the number of parameters doubles as p increases by one. In addition as shown in [2], isotropy places numerous polynomial constraints on these parameters. As we now describe a better representation is provided by the generalization of lattice structures which involves only one new parameter as p increases by one.

Let $\mathcal{H}\{\dots\}$ denote the Gaussian linear space spanned by the variables in braces and define the (n th order) past of the node t :

$$\mathcal{Y}_{t,n} \triangleq \mathcal{H}\{Y_{wt} : w \leq 0, |w| \leq n\} \quad (2.3)$$

As for time series, the development of models of increasing order involves recursions for the forward and backward prediction errors. Specifically, define the **backward residual space**:

$$\mathcal{Y}_{t,n} = \mathcal{Y}_{t,n-1} \oplus \mathcal{F}_{t,n} \quad (2.4)$$

where $\mathcal{F}_{t,n}$ is spanned by the backward prediction errors

$$F_{t,n}(w) \triangleq Y_{wt} - E(Y_{wt} | \mathcal{Y}_{t,n-1}) \quad (2.5)$$

where $w \leq 0$, $|w| = n$. These variables are collected into a $2^{\lfloor \frac{n}{2} \rfloor}$ -dimensional vector (see [2] for the order), $F_{t,n}$. For $|w| < n$ and $w \asymp 0$ (i.e. $m(wt) = m(t)$) define the **forward prediction errors**:

$$E_{t,n}(w) \triangleq Y_{wt} - E(Y_{wt} | \mathcal{Y}_{\gamma^{-1}t, n-1}) \quad (2.6)$$

and let $\mathcal{E}_{t,n}$ denote the span of these residuals and $E_{t,n}$ the $2^{\lfloor \frac{n-1}{2} \rfloor}$ -dimensional vector of these variables (see [2]).

The key to the development of our models is the recursive computation of $F_{t,n}$ and $E_{t,n}$ as n increases. The general idea is the same as for time series but we must deal with the more complex geometry of the tree and the changing dimensions of $F_{t,n}$ and $E_{t,n}$. In particular, as shown in [2], it is necessary to distinguish between n even and odd and between different groups of the components of $F_{t,n}$ and $E_{t,n}$. For example, $F_{t,n}$ consists of $F_{t,n}(w)$ in eq.(2.5) with $|w| = n$, $w \preceq 0$. Suppose that n is even and consider elements of $F_{t,n}$ for which $|w| = n$, $w \prec 0$. In this case $w = \tilde{w}\gamma^{-1}$ for some $\tilde{w} \preceq 0$, with $|\tilde{w}| = n - 1$, and by an argument exactly analogous to the time series case we obtain the recursion:

$$F_{t,n}(w) = F_{\gamma^{-1}t,n-1}(\tilde{w}) - E[F_{\gamma^{-1}t,n-1}(\tilde{w})|E_{t,n-1}] \quad (2.7)$$

This procedure identifies several projections, as in eq.(2.7), to be calculated. A key result is that these projection operators can in fact be reduced to **scalar** projections involving a single new reflection coefficient and the local averages or **barycenters** of the residuals:

$$e_{t,n} = 2^{-[\frac{n-1}{2}]} \sum_{|w| < n, w \succ 0} E_{t,n}(w) \quad (2.8)$$

$$f_{t,n} = 2^{-[\frac{n}{2}]} \sum_{|w|=n, w \preceq 0} F_{t,n}(w) \quad (2.9)$$

For example, the projection in eq.(2.7) is the same for **all** such \tilde{w} and in fact equals $E[F_{\gamma^{-1}t,n-1}(\tilde{w})|e_{t,n-1}]$. This and related expressions follow from the properties of isotropy and from a very important fact: any local isometry, i.e. a map f from one subset of A onto another that preserves distances, can be extended to a full isometry on \mathcal{T} .

As a consequence of this result, we can obtain scalar Levinson recursions for the barycenters themselves. Note first that $e_{t,0} = E_{t,0} = Y_t = F_{t,0} = f_{t,0}$; that $e_{t,1} = E_{t,1}$, $f_{t,1} = F_{t,1}$; and that a straightforward calculation yields

$$f_{t,1} = f_{\gamma^{-1}t,0} - k_1 e_{t,0} \quad (2.10)$$

$$e_{t,1} = e_{t,0} - k_1 f_{\gamma^{-1}t,0} \quad (2.11)$$

$$-1 \leq k_1 = \frac{r_1}{r_0} \leq 1 \quad (2.12)$$

For n even we find [2] that

$$e_{t,n} = e_{t,n-1} - k_n f_{\gamma^{-1}t,n-1} \quad (2.13)$$

$$f_{t,n} = \frac{1}{2} (f_{\gamma^{-1}t,n-1} + e_{\delta(\frac{n}{2})t,n-1}) - k_n e_{t,n-1} \quad (2.14)$$

where the reflection coefficients k_n and the variances of the residuals satisfy

$$\begin{aligned} k_n &= \text{cor}(e_{t,n-1}, f_{\gamma^{-1}t,n-1}) \\ &= \text{cor}\left(e_{\delta(\frac{n}{2})_{t,n-1}}, e_{t,n-1}\right) \\ &= \text{cor}\left(e_{\delta(\frac{n}{2})_{t,n-1}}, f_{\gamma^{-1}t,n-1}\right) \end{aligned} \quad (2.15)$$

$$\text{cor}(x, y) = E(xy) / [E(x^2)E(y^2)]^{1/2} \quad (2.16)$$

$$\sigma_{e,n}^2 = E(e_{t,n}^2) = (1 - k_n^2) \sigma_{n-1}^2 \quad (2.17)$$

$$\sigma_{f,n}^2 = E(f_{t,n}^2) = \left(\frac{1 + k_n}{2} - k_n^2\right) \sigma_{n-1}^2 \quad (2.18)$$

where k_n must satisfy

$$-\frac{1}{2} < k_n < 1 \quad (2.19)$$

For n odd we have

$$e_{t,n} = \frac{1}{2} \left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})_{t,n-1}} \right) - k_n f_{\gamma^{-1}t,n-1} \quad (2.20)$$

$$f_{t,n} = f_{\gamma^{-1}t,n-1} - \frac{1}{2} k_n \left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})_{t,n-1}} \right) \quad (2.21)$$

$$k_n = \text{cor}\left(\frac{1}{2} \left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})_{t,n-1}} \right), f_{\gamma^{-1}t,n-1}\right) \quad (2.22)$$

$$\sigma_{e,n}^2 = \sigma_{f,n}^2 = \sigma_n^2 = (1 - k_n^2) \sigma_{f,n-1}^2 \quad (2.23)$$

where

$$-1 < k_n < 1 \quad (2.24)$$

Note that the constraints on the reflection coefficients are slightly different than for time series, and as shown in [2] these conditions are precisely those for a sequence r_n to be the covariance of an isotropic process. In addition, the following generalization of the Schur recursions allow us to calculate the k_n efficiently. Define the formal power series

$$P_n \triangleq \text{cov}(Y_t, e_{t,n}) \triangleq \sum_{w \leq 0} E(Y_t e_{wt,n}) \cdot w \quad (2.25)$$

$$Q_n \triangleq \text{cov}(Y_t, f_{t,n}) \triangleq \sum_{w \leq 0} E(Y_t f_{wt,n}) \cdot w \quad (2.26)$$

Then

$$P_0 = Q_0 = \sum_{w \leq 0} r_{|w|} \cdot w \quad (2.27)$$

while for n even

$$P_n = P_{n-1} - k_n \gamma Q_{n-1} \quad (2.28)$$

$$Q_n = \frac{1}{2} \left(\gamma Q_{n-1} + \delta^{(\frac{n}{2})} P_{n-1} \right) - k_n P_{n-1} \quad (2.29)$$

where

$$k_n = \frac{\gamma Q_{n-1}(0) + \delta^{(\frac{n}{2})} P_{n-1}(0)}{2P_{n-1}(0)} \quad (2.30)$$

and for n odd

$$P_n = \frac{1}{2} \left(P_{n-1} + \delta^{(\frac{n-1}{2})} P_{n-1} \right) - k_n \gamma Q_{n-1} \quad (2.31)$$

$$Q_n = \gamma Q_{n-1} - k_n \frac{1}{2} \left(P_{n-1} + \delta^{(\frac{n-1}{2})} P_{n-1} \right) \quad (2.32)$$

where

$$k_n = \frac{2\gamma Q_{n-1}(0)}{P_{n-1}(0) + \delta^{(\frac{n-1}{2})} P_{n-1}(0)} \quad (2.33)$$

In [2] we also show how these same k_n can be used to construct whitening and modeling filters for Y_t and we present a stability result analogous to the time series case: k_n must not achieve any of the extreme values in eq.(2.19), eq.(2.24). In addition we demonstrate that the class of AR(p) processes are completely equivalent to reflection coefficient sequences with $k_n = 0$, $n > p$ and we show that these processes are exactly the isotropic processes with impulse responses with support on a cylinder of radius $[\frac{p}{2}]$ about the strict past γ^{-n} .

3. STATE MODELS AND MULTIGRID ESTIMATION

A second class of models displaying coarse-to-fine structure is specified by state models of the form

$$x(t) = A(m(t))x(\gamma^{-1}t) + B(m(t))w(t) \quad (3.1)$$

where $w(t)$ is a vector white noise process with covariance I . The model eq.(3.1) describes a process that is Markov scale-to-scale and, because of this we can readily calculate its second order statistics. For example in the case in which A and B are constant and A is stable, eq.(3.1) can describe **stationary** processes, where the covariance of x satisfies the Lyapunov equation

$$P_x = AP_x A^T + BB^T \quad (3.2)$$

and the correlation function is

$$K_{xx}(t, s) = A^{d(t, s \wedge t)} P_x (A^T)^{d(t, s \wedge t)} \quad (3.3)$$

In the scalar case, or if $AP_x = P_x A^T$, eq.(3.1) describes an isotropic process, but in general eq.(3.1) describes a somewhat larger set of processes.

Consider now the estimation of $x(t)$ based on measurements

$$y(t) = C(m(t))x(t) + v(t) \quad (3.4)$$

where $v(t)$ is white noise of covariance $R(m(t))$, independent of x . In many problems we may only have data at the finest level; however in some applications such as geophysical signal processing or the fusion of multispectral data, data at multiple scales is collected and must be combined. In [3] we describe three different algorithmic structures for estimating $x(t)$ based on the measurements in eq.(3.4). One of these involves processing from one scale to the next. This structure resembles the Laplacian pyramid processing structure [4] and can be performed extremely quickly using discrete Haar transforms.

A second structure is based on the following equality which can be derived from the Markovian structure of eq.(3.1):

$$\hat{x}(t) = L_1 \hat{x}(\gamma^{-1}t) + L_2(\hat{x}(\alpha t) + \hat{x}(\beta t)) + L_3 y(t) \quad (3.5)$$

where L_1 , L_2 , and L_3 are gains (depending upon scale in general). Eq.(3.5) describes a set of coupled equations from scale to scale which can be solved by Gauss-Seidel relaxation that can be structured exactly as in multigrid algorithms for the solution of partial differential equations.

A third algorithm involves a single fine-to-coarse sweep followed by a coarse-to-fine correction. In the first step we recursively calculate the best estimate of $x(t)$ based on observations in its descendent subtree. This recursion involves three steps, which together define a new Riccati equation: a **backward prediction step** to predict from αt and βt to t ; a **merge step**, merging these two estimates; and an **update step** incorporating the measurement at t . The merge step is the new feature that has no counterpart for standard temporal models. Once we have reached the top node of the tree, the downward sweep has the same form as the Rauch-Tung-Striebel form of the optimal smoother for temporal models (allowing of course for the proliferation of parallel calculations as the algorithm passes from coarser to finer scales): the best smoothed estimate at t is calculated in terms of the best smoothed estimate at $\gamma^{-1}t$ and the filtered estimate at that node calculated during the upward sweep.

The work of M.B. and A.B. was supported in part by Grant CNRS GO134. The work of K.C.C. and A.S.W. was supported in part by AFOSR under Grant AFOSR-88-0032, in part by NSF under Grant ECS-8700903, and in part by ARO under Contract DAAL03-86-K-0171.

REFERENCES

- [1] J.P. Arnaus, G. Letac, "La formule de representation spectrale d'un processus gaussien stationnaire sur un arbre homogene," Publ. Lab. Stat. and Prob. UA 745, Toulouse.
- [2] M. Basseville, A. Benveniste, and A.S. Willsky, "Multi-Scale Autoregressive Processes," in preparation.
- [3] K.C. Chou, A.S. Willsky, A. Benveniste, and M. Basseville, "Recursive and Iterative Estimation Algorithms for Multi-Resolution Stochastic Processes," submitted to 1989 IEEE Conf. on Decision and Control.
- [4] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," Comm. on Pure and Appl. Math., Vol. 91, 1988, pp. 909-996.