Multiscale Autoregressive Processes, Part I: Schur–Levinson Parametrizations

Michèle Basseville, Albert Benveniste, *Fellow, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

Abstract—**In many applications (e.g., recognition of geophysical and biomedical signals and multiscale analysis of images), it is of interest to analyze and recognize phenomena occurring at different scales. The recently introduced wavelet transforms provide a time-and-scale decomposition of signals that offers the possibility of such analysis. At present, however, there is no corresponding statistical framework to support the development of optimal, multiscale statistical signal processing algorithms. In this paper we describe such a framework. The theory of multiscale signal representations leads naturally to models of signals on trees, and this provides the framework for our investigation. In particular, in this paper we describe the class of isotropic processes on homogeneous trees and develop a theory of autoregressive models in this context. This leads to generalizations of Schur and Levinson recursions, associated properties of the resulting reflection coefficients, and the initial pieces in a system theory for multiscale modeling.**

I. Introduction

THE investigation of multiscale representations of signals and the development of multiscale algorithms has been and remains a topic of much interest in many contexts. In some cases, such as in the use of fractal models for signals and images [4], [33] the motivation has directly been the fact that the phenomenon of interest exhibits patterns of importance at multiple scales. A second motivation has been the possibility of developing highly parallel and iterative algorithms based on such representations. Multigrid methods for solving partial differential equations [12], [27], [34], [36] are a good example. A third motivation stems from so-called "sensor fusion" problems in which one is interested in combining together measurements with very different spatial resolutions. Geophysical problems, for example, often have this character.

Manuscript received March 20, 1990; revised May 30, 1991. The work of M. Basseville and A. Benveniste was supported in part by Grant CNRS G0134 and by INRIA-NSF agreement. The work of A. S. Willsky was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002, in part by the National Science Foundation under Grants MIP-901528 and INT-9002393, and in part by the U.S. Army Research Office under Contract DAAL03-86-K-0171, and INRIA.

M. Basseville in with IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France, and with the Centre Nationale de la Recherche Scientifique (CNRS), France.

A. Benveniste is with IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France, and with the Institut National de Recherche en Informatique et en Automatique (INRIA), France.

A. S. Willsky is with the Laboratory for Information and Decision Systems, and the Department of Electrical Engineering and Computer Science, Massachusettsd Institute of Technology, Cambridge, MA 02139.

IEEE Log Number 9201074.

One of the more recent areas of investigation in multiscale analysis has been the development of a theory of multiscale representations of signals [30], [32] and the closely related topic of wavelet transforms [20]–[22], [24]–[26], [29]. These methods have drawn considerable attention in several disciplines including signal processing because they provide a natural way to perform a time-scale decomposition of signals. In addition examples that have been given of such transforms seem to indicate that it should be possible to develop efficient optimal processing algorithms based on these representations. The development of such optimal algorithms, e.g., for the reconstruction of noise-degraded signals, for the optimal fusion of multiresolution data, or for the detection and localization of transient signals of different duration, requires the development of a statistical framework for analyzing, fusing, and modeling signals and phenomena at multiple scales. The research presented in this paper and its companion [1] as well as in several other papers and reports [7], [8], [5], [16]–[18] has the development of such a framework as its objective.

Essentially all methods, including wavelet transforms [21], [31] and multirate filtering [19], [39], for representing and processing signals at multiple resolution involve pyramidal data structures. Each level in such a structure corresponds to signal representation at a particular scale, with successively decimated representations at coarser scales. Such a multiscale representation has a natural interpretation as a tree in which each level in the tree corresponds to a particular scale and each node at a given scale is connected both to a "parent" node at the neighboring coarser scale and to several "descendant" nodes at the neighboring finer scale. The simplest, and most common, structure of this type is the dyadic tree in which there is a factor of two decimation from scale to scale, and thus there are two descendant nodes for each parent. It is this structure that provides the starting point for our investigation. That is, our objective is to develop a theory of statistical modeling and signal processing for multiscale representations of signals defined on dyadic trees.

There are a number of criteria that must be satisfied for such a formalism to be of value. First, it must be rich enough to be useful for modeling significant classes of phenomena and signals. Second, it should provide the basis for the development of efficient algorithms for modeling, estimation, segmentation, and other signal analysis tasks. Finally, in our opinion for such a formalism to pro-

1053-587X/92$03.00 © 1992 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

vide a true foundation for multiscale statistical signal processing, it should be based on the extensions to this setting of the basic principles that have provided the framework for the powerful and successful classical methods of time series analysis.

One of the most powerful concepts in time series analysis is that of stationarity, and the extension and use of this concept to the multiscale context represents one of the goals of our work. In [8] we introduce and study stationarity and the related systems concepts of shift invariance beginning from first principles and developing a state-space and systems theory for multiscale systems and processes on the dyadic tree. Since shifts on the tree are shown to correspond to changes in scale [8], it is not surprising that stationarity in this sense corresponds to a notion of statistical scale invariance or self-similarity. Indeed, as developed in [7], [8], [16]–[18], [40], [41], this framework provides an excellent setting for the modeling of self-similar or fractal processes; moreover, as shown in [7], [17], the natural extension of these models to scale-varying systems and processes allows us to develop useful multiscale models for a surprisingly large class of processes.

In this paper we investigate an alternate framework to that in [8] for an important subclass of stationary processes on trees. Specifically in classical time series analysis we have not only the fundamental internal modeling framework of state-space methods but also the "black-box" or input-output framework of regression. This framework leads not only to the powerful methods of AR and ARMA modeling but also to efficient algorithms for building models directly from data. The principal objective of this paper and its companion [1] is to develop a similar framework for multiscale signal representations. In particular, we develop a theory of multiscale autoregressive modeling for a class of self-similar processes which we refer to as isotropic.

An obvious and critical aspect of both state-space and autoregressive modeling is that we require a notion of recursion or causality, i.e., in time series analysis we deal with the time axis asymmetrically (in terms of past and future). In the case of multiscale representations we also have a one-dimensional "axis" in which we may consider recursion, namely the axis of scale. In particular, the models considered here and in [8], [16]–[18] involve a notion of causality or recursion proceeding from coarse scales to finer scales, exactly as in signal synthesis (rather than analysis) using wavelet transforms or Laplacian pyramids. By adopting this asymmetric view of scale we gain exactly the same advantage as in time series analysis. That is, without any sacrifice in the generality of the representation, we gain all of the advantages of recursion. These can be used as in [16]–[18] to develop efficient scale-recursive optimal estimation algorithms or as we do here in finding and parametrizing the class of causal models on trees that produce isotropic processes. In particular in this paper and in [1] we extend the powerful and computationally efficient Schur and Levinson algorithms to the multiscale framework and use there as the basis for the AR parametrization of isotropic processes and the construction of lattice filters for multiscale modeling and whitening. As in standard time series analysis, these whitening filter structures provide us with a key ingredient for the efficient calculation of likelihoods for a variety of signal processing problems including parameter estimation, segmentation, and hypothesis testing.

In the next section we introduce the multiscale representation of signals on dyadic trees. We also introduce the class of multiscale isotropic processes and set the stage for introducing dynamic modes on trees by describing their structure and introducing a rudimentary transform theory. In addition, we describe the class of autoregressive (AR) models on trees. As we will see, the geometry and structure of a dyadic tree is such that the dimension of an AR model increases with the order of the model. Thus an nth order AR model is characterized by more than n coefficients whose interdependence is specified by a complex relation and the passage from order n to order $n + 1$ is far from simple. In contrast, as we will show, we obtain a far simpler picture if we consider the generalization of lattice structures.

In particular, in Section III we introduce forward and backward prediction error processes on dyadic trees, and, as for time series, we develop Levinson-like recursions for these processes as the order of prediction increases. While the basic structure of our analysis is very much the same as for time series, there are significant new issues that must be addressed for processes on dyadic trees. In particular, we will see that the dimension of the prediction error processes increases with the order of prediction (due essentially to the fact that the number of nodes at a given distance from a specified node increases geometrically with distance). This requires significant care in computing the projections, required in the Levinson's recursions, of error vectors of one order on those of the preceding order. The result is an apparently complicated set of expressions.

However, as we explore in Section IV, the constraints of isotropy lead to significant simplifications in the required projections. Indeed, we will show that all of these apparently different vector projections collapse into projections onto scalar averages (or barycenters) of the prediction error vectors. Thus, as for time series, only one reflection coefficient is needed to specify the Levinson recursion at each stage, and in Section IV we also develop Schur-like recursions for the computation of these reflection coefficients. As we will see, the constraints that the reflection coefficients must satisfy are somewhat different than for the case of time series.

The actual construction of whitening and shaping (or modeling) filters for the original process requires, of course, the full error processes and not just their barycenters. In part II of this paper [1] we build on the results of Sections III and IV to construct and analyze whitening and modeling lattice filters for AR processes. In particular, we will see that the constraints on the reflection coef-

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

ficients are again necessary and sufficient for stability, and we also use our analysis to study in detail the Wold decomposition for isotropic processes. One important result here is that all isotropic AR processes can be modeled using our Schur/Levinson/lattice construction.

II. Multiscale Representations and Stochastic Processes on Homogeneous Trees

The starting point for our development is the pyramidal structure of multiscale representations for signals in continuous or discrete time. For example in discrete time such a structure arises naturally in multirate digital filtering [19] and the problem of decimation with perfect reconstruction [39]. Similarly in continuous time [21], [31] a multiscale representation of a signal $f(x)$ consists of a sequence of approximations of that signal at finer and finer scales where the approximation of $f(x)$ at the mth scale is given by

$$f_m(x) = \sum_{n=-\infty}^{+\infty} f(m, n)\phi(2^m x - n). \tag{2.1}$$

The set of coefficients $\{f(m, n): n = 0, \pm 1, \pm 2, \cdots\}$ corresponds to the representation of a signal at the mth scale. Thanks to the scaling factor of 2^m in (2.1), there is a factor of 2 decimation between any scale and the previous, coarser one. This yields the dyadic tree structure of the full set of representations over all scales, in which any node (m, n) has two descendent nodes at the next, finer scale and one parent node at the preceding coarser scale. It is this data structure that we have in mind to study here. However, in order to introduce carefully the particular notion of stationarity we use here, namely that of isotropy, we must first step back and take a more fundamental, abstract view of the dyadic tree index set.

A. Homogenous Trees

Homogenous trees, and their structure, have been the subject of some work [2], [3], [15], [23], [14] in the past on which we build and which we now briefly review. A homogenous tree $\mathfrak{I}$ of order q is an infinite acyclic, undirected, connected graph such that every node of $\mathfrak{I}$ has exactly $(q + 1)$ branches to other nodes. Note that $q = 1$ corresponds to the usual integers with the obvious branches from one integer to its two neighbors. The case of $q = 2$, illustrated in Fig. 1, corresponds, as we will see, to the dyadic tree on which we focus throughout the paper. In 2-D signal processing it would be natural to consider the case of $q = 4$ leading to a pyramidal structure for our indexing of processes.

The tree $\mathfrak{I}$ has a natural notion of distance: $d(s, t)$ is the number of branches along the shortest path between the nodes of $s, t \in \mathfrak{I}$ (by abuse of notation we use $\mathfrak{I}$ to denote both the tree and its collection of nodes). One can then define the notion of an isometry on $\mathfrak{I}$ which is simply a one-to-one map of $\mathfrak{I}$ onto itself that preserves distances. For the case of $q = 1$, the group of all possible isometries corresponds to translations of the integers $(t \mapsto t + k)$,

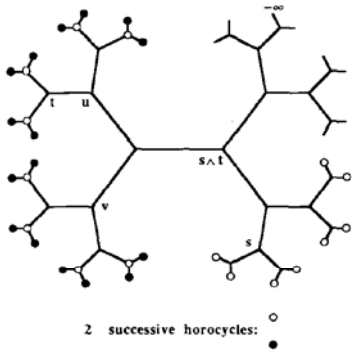

Fig. 1. The dyadic homogenous tree.

the reflection operation $(t \mapsto -t)$ and concatenations of these. For $q \geq 2$ the group of isometries of $\mathfrak{I}$ is significantly larger and more complex. One extremely important result is the following [23]:

Lemma 2.1 (Extension of Isometries): Let $\mathfrak{I}$ be a homogenous tree of order q, let A and A' be two subsets of nodes, and let f be a local isometry from A to A', i.e., f is bijection from A onto A' such that

$$d(f(s), f(t)) = d(s, t) \qquad \text{for all } s, t \in A. \tag{2.2}$$

Then there exists an isometry $\tilde{f}$ of $\mathfrak{I}$ which equals f when restricted to A. Furthermore, if $\tilde{f}_1$ and $\tilde{f}_2$ are two such extensions of f, their restrictions to the shortest path joining any two points of A are identical.

Another important concept is the notion of a boundary point [3], [14] of a tree. Consider the set of infinite sequences of $\mathfrak{I}$ where any such sequence consists of a sequence of distinct nodes $t_1, t_2, \cdots$ where $d(t_i, t_{i+1}) = 1$. A boundary point is an equivalence class of such sequences where two sequences are equivalent if they differ by a finite number of nodes. For $q = 1$, there are only two such boundary points corresponding to sequences increasing towards $+\infty$ or decreasing toward $-\infty$. For $q = 2$ the set of boundary points is uncountable. In this case let us choose one boundary point which we will denote by $-\infty$.

Once we have distinguished this boundary point we can identify a partial order on $\mathfrak{I}$. In particular, note that from any node t there is a unique path in the equivalence class defined by $-\infty$ (i.e., a unique path from t "toward" $-\infty$). Then if we take any two nodes s and t, their paths to $-\infty$ must differ by only a finite number of points and thus must meet at some node which we denote by $s \wedge t$ (see Fig. 1). We then can define a notion of the relative distance of two nodes to $-\infty$

$$\delta(s, t) = d(s, s \wedge t) - d(t, s \wedge t) \tag{2.3}$$

so that

$$s \preceq t \text{ ("}s \text{ is at least as close to } -\infty \text{ as } t\text{")} \quad \text{if } \delta(s, t) \leq 0 \tag{2.4}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

$$s \lessdot t \text{ (``}s \text{ is closer to } -\infty \text{ than } t\text{''})$$
$$\text{if } \delta(s, t) < 0. \tag{2.5}$$

This also yields an equivalence relation on nodes of $\mathfrak{I}$:

$$s \Leftrightarrows t \leftrightarrow \delta(s, t) = 0. \tag{2.6}$$

For example, the points s, v, and u in Fig. 1 are all equivalent. The equivalence classes of such nodes are referred to as horocycles.

These equivalence classes can best be visualized as in Fig. 2 by redrawing the tree, in essence by picking the tree up at $-\infty$ and letting the tree ''hang'' from this boundary point. In this case the horocycles appear as points on the same horizontal level and $s \leq t$ means that s lies on a horizontal level above or at the level of t. Note that in this way we make explicit the dyadic structure of the tree. With regard to multiscale signal representations, a shift on the tree toward $-\infty$ corresponds to a shift from a finer to a coarser scale and points on the same horocycle correspond to the points at different translational shifts in the signal representation at a single scale. Note also that we now have a simple interpretation for the nondenumerability of the set of boundary points: they correspond to dyadic representations of all real numbers.

B. Shifts and Transforms on $\mathfrak{I}$

The structure of Fig. 2 provides the basis for our development of dynamical models on trees since it identifies a ''time-like'' direction corresponding to shifts toward or away from $-\infty$. In order to define such dynamics we will need the counterpart of the shift operators z and z^{-1} in order to define shifts or moves in the tree. Because of the structure of the tree the description of these operators is a bit more complex, and in fact we introduce notation for five operators representing the following elementary moves on the tree, which are also illustrated in Fig. 3:

- 0 identity operator (no move),
- γ^{-1} backward shift (move one step toward $-\infty$),
- α left forward shift (move one step away from $-\infty$ toward the left),
- β right forward shift (move one step away from $-\infty$ toward the right),
- δ interchange operator (move to the nearest point in the same horocycle).

Note that 0 is an isometry; δ is invertible; α and β are one to one but not onto; γ^{-1} is onto but not one to one; and these operators satisfy the following relations (where the convention is that the leftmost operator is applied first):

$$\alpha\gamma^{-1} = \beta\gamma^{-1} = 0 \tag{2.7}$$
$$\delta\gamma^{-1} = \gamma^{-1} \tag{2.8}$$
$$\delta^2 = 0 \tag{2.9}$$
$$\beta\delta = \alpha. \tag{2.10}$$

Arbitrary moves on the tree can be encoded via finite strings or words using these symbols as the alphabet and the formulas (2.7)–(2.10). For example, referring to Fig. 3

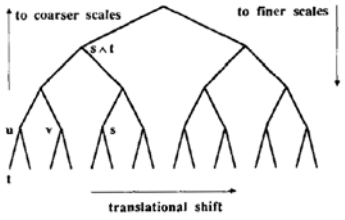

Fig. 2. Showing the scale structure.

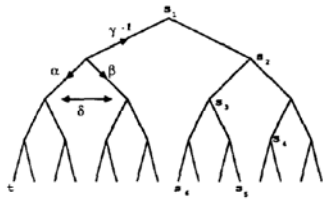

Fig. 3. Encoding the moves on the tree.

$$s_1 = t\gamma^{-4}, \quad s_2 = t\gamma^{-3}\delta, \quad s_3 = t\gamma^{-3}\delta\alpha$$
$$s_4 = t\gamma^{-3}\delta\beta\alpha, \quad s_5 = t\gamma^{-3}\delta\alpha\beta^2. \tag{2.11}$$

It is also possible to code all points on the tree via their shifts from a specified, arbitrary point t_0 taken as origin. Specifically, define the language

$$\mathfrak{L} = (\gamma^{-1})^* \cup (\gamma^{-1})^*\delta\{\alpha, \beta\}^* \cup \{\alpha, \beta\}^* \tag{2.12}$$

where K^* denotes arbitrary sequences of symbols in K including the empty sequence which we identify with the operator 0. Then any point $t \in \mathfrak{I}$ can be written as $t_0 w$, where $w \in \mathfrak{L}$. Note that the moves in $\mathfrak{L}$ are of three types: a pure shift back toward $-\infty$ $((\gamma^{-1})^*)$; a pure descent away from $-\infty$ $(\{\alpha, \beta\}^*)$; and a shift up followed by a descent down another branch of the tree $((\gamma^{-1})^*\delta\{\alpha, \beta\}^*)$. Our use of δ in the last category of moves ensures that the subsequent downward shift is on a different branch than the preceding ascent. This emphasizes an issue that arises in defining dynamics on trees. Specifically, we will avoid writing strings of the form $\gamma^{-1}\alpha$ or $\gamma^{-1}\beta$. For example $t\gamma^{-1}\alpha$ either equals t or $t\delta$ depending upon whether t is the left or right immediate descendant of another node. By using δ in our language we avoid this issue. One price we pay is that $\mathfrak{L}$ is not a semigroup since vw need not be in $\mathfrak{L}$ for $v, w \in \mathfrak{L}$. However, for future reference we note that, using (2.7)–(2.10), $w\delta$ and $w\gamma^{-1}$ are both in $\mathfrak{L}$ for any $w \in \mathfrak{L}$.

It is straightforward to define a length $|w|$ for each word in $\mathfrak{L}$, corresponding to the number of shifts required in the move specified by w. Note that

$$|\gamma^{-1}| = |\alpha| = |\beta| = 1$$
$$|0| = 0, \quad |\delta| = 2. \tag{2.13}$$

Thus $|\gamma^{-n}| = n$, $|w_{\alpha\beta}| =$ the number of α's and β's in

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

$w_{\alpha\beta} \in \{\alpha, \beta\}^*$, and $|\gamma^{-n}\delta w_{\alpha\beta}| = n + 2 + |w_{\alpha\beta}|$.[1] This notion of length will be useful in defining the order of dynamic models on $\mathfrak{I}$. We will also be interested exclusively in causal models, i.e., in models in which the output at some scale (horocycle) does not depend on finer scales. For this reason we are most interested in moves that either involve pure ascents on the tree, i.e., all elements of $\{\gamma^{-1}\}^*$, or elements $\gamma^{-n}\,\delta w_{\alpha\beta}$ of $\{\gamma^{-1}\}^*\,\delta\{\alpha, \beta\}^*$ in which the descent is no longer than the ascent, i.e., $|w_{\alpha\beta}| \leq n$. We use the notation $w \preceq 0$ to indicate that w is such a causal move. Note that we include moves in this causal set that are not strictly causal in that they shift a node to another on the same horocycle. We use the notation $w \precsim 0$ for such a move. The reasons for this will become clear when we examine autoregressive models.

Also, on occasion we find it useful to use a simplified notation for particular moves. Specifically, we define $\delta^{(n)}$ recursively, starting with $\delta^{(1)} = \delta$ and

$$\begin{aligned} &\text{If } t = t\gamma^{-1}\alpha, \quad \text{then } t\delta^{(n)} = t\gamma^{-1}\,\delta^{(n-1)}\alpha \\ &\text{If } t = t\gamma^{-1}\beta, \quad \text{then } t\delta^{(n)} = t\gamma^{-1}\,\delta^{(n-1)}\beta. \end{aligned} \tag{2.14}$$

What $\delta^{(n)}$ does is to map t to another point on the same horocycle in the following manner: we move up the tree n steps and then descend n steps; the first step in the descent is the opposite of the one taken on the ascent, while the remaining steps are the same. That is, if $t = t\gamma^{-n+1}w_{\alpha,\beta}$ then $t\delta^{(n)} = t\gamma^{-n+1}\,\delta w_{\alpha\beta}$. For example, referring to Fig. 3, $s_6 = t\delta^{(4)}$.

With the notation we have defined we can now define transforms as a way in which to encode convolutions much as z transforms do for temporal systems. In particular, we consider systems that are specified via noncommutative formal power series [11] of the form:

$$S = \sum_{w \in \mathfrak{L}} s_w \cdot w, \qquad s_w \in \boldsymbol{R}. \tag{2.15}$$

If the input to this system is u_t, $t \in \mathfrak{I}$, then the output is given by the generalized convolution:

$$(Su)_t = \sum_{w \in \mathfrak{L}} s_w u_{tw}. \tag{2.16}$$

For future reference we use the notation $S(0)$ to denote the coefficient of the empty word in S. Also it will be necessary for us to consider particular shifted versions of S:

$$\gamma[S] = \sum_{w \in \mathfrak{L}} s_{w\gamma^{-1}} \cdot w \tag{2.17}$$

$$\delta^{(k)}[S] = \sum_{w \in \mathfrak{L}} s_{w\delta^{(k)}} \cdot w \tag{2.18}$$

where we use (2.7)–(2.10) and (2.14) to write $w\gamma^{-1}$ and $w\delta^{(k)}$ as elements of $\mathfrak{L}$. Notice that, because of the relations (2.7)–(2.10),the operators $S \to \gamma[S]$ and $S \to \delta[S]$ cannot be thought of as multiplication operators on formal power series.

[1]Note another consequence of the ambiguity in $\gamma^{-1}\alpha$: its "length" should either be 0 or 2.

C. Isotropic Processes on Homogenous Trees

Consider a zero-mean stochastic process Y_t, $t \in \mathfrak{I}$ indexed by nodes on the tree. We say that such a process is isotropic if the covariance between Y at any two points depends only on the distance between the points, i.e., if there exists a sequence r_n, $n = 0, 1, 2, \cdots$ so that

$$E[Y_t Y_s] = r_{d(t,s)}. \tag{2.19}$$

An alternate way to think of an isotropic process is that its statistics are invariant under tree isometries. That is, if $f: \mathfrak{I} \to \mathfrak{I}$ is an isometry and if Y_t is an isotropic process, then $Z_t = Y_{f(t)}$ has the same statistics as Y_t. For time series this simply states that Y_{-t} and Y_{t+k} have the same statistics as Y_t. For dyadic trees the richness of the group of isometries makes isotropy a much stronger property.

Isotropic processes have been the subject of some study [2], [3], [23] in the past, and in particular a spectral theorem has been developed that is the counterpart of Bochner's theorem for stationary time series. In particular, Bochner's theorem states that a sequence r_n, $n = 0, 1, \cdots$ is the covariance function of a stationary time series if and only if there exists a nonnegative, symmetric spectral measure $S(d\omega)$ so that

$$\begin{aligned} r_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega n} S(d\omega) \\ &= \frac{1}{\pi}\int_0^{\pi} \cos(\omega n)\, S(d\omega). \end{aligned}$$

If we perform the change of variables $x = \cos\omega$ and note that $\cos(n\omega) = C_n(\cos\omega)$, where $C_n(x)$ is the nth Chebychev polynomial, we have

$$r_n = \int_{-1}^{1} C_n(x)\,\mu(dx) \tag{2.20}$$

where $\mu(dx)$ is a nonnegative measure on $[-1, 1]$ (also referred to as the spectral measure) given by

$$\mu(dx) = \frac{1}{\pi}(1 - x^2)^{-(1/2)} S(d\omega). \tag{2.21}$$

For example, for the white noise sequence with $r_n = \delta_{n0}$,

$$\mu(dx) = \frac{1}{\pi}(1 - x^2)^{-(1/2)}. \tag{2.22}$$

The analogous theorem for isotropic processes on dyadic trees requires the introduction of the Dunau polynomials [2], [23]:

$$P_0(x) = 1, \qquad P_1(x) = x \tag{2.23}$$

$$xP_n(x) = \tfrac{2}{3}P_{n+1}(x) + \tfrac{1}{3}P_{n-1}(x). \tag{2.24}$$

Theorem 2.1 [2], [3]: A sequence r_n, $n = 0, 1, 2, \cdots$ is the covariance function of an isotropic process on a dyadic tree if and only if there exists a nonnegative measure μ on $[-1, 1]$ so that

$$r_n = \int_{-1}^{1} P_n(x)\,\mu(dx). \tag{2.25}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

The simplest isotropic process on the tree is again white noise, i.e., a collection of uncorrelated random variables indexed by $\mathfrak{I}$, with $r_n = \delta_{n0}$, and the spectral measure μ in (2.25) in this case is [23]

$$\mu(dx) = \frac{1}{2\pi} \chi_{[-(2\sqrt{2}/3),(2\sqrt{2}/3)]}(x) \frac{(8 - 9x^2)^{1/2}}{1 - x^2} dx \quad (2.26)$$

where $\chi_A(x)$ is the characteristic function of the set A. A key point here is that the support of this spectral measure is smaller than the interval $[-1, 1]$. This appears to be a direct consequence of the large size of the boundary of the tree, which also leads to the existence of a far larger class of singular processes than one finds for time series. While Theorem 2.1 does provide a necessary and sufficient condition for a sequence r_n to be the covariance of an isotropic process, it does not provide an explicit and direct criterion in terms of the sequence values. For time series we have such a criterion based on the fact that r_n must be a positive semidefinite sequence. It is not difficult to see that r_n must also be positive semidefinite for processes on dyadic trees: form a time series by taking any sequence $Y_{t_1}, Y_{t_2}, \cdots$ where the t_i are all distinct and $d(t_i, t_{i+1}) = 1$; the covariance function of this series is r_n. However, thanks to the geometry of the tree and the richness of the group of isometries of $\mathfrak{I}$, there are many additional constraints on r_n. For example, consider the three nodes v, u, and $s \wedge t$ in Fig. 1, and let

$$X^T = [Y_v, Y_u, Y_{s\wedge t}]. \quad (2.27)$$

Then

$$E[XX^T] = \begin{bmatrix} r_0 & r_2 & r_2 \\ r_2 & r_0 & r_2 \\ r_2 & r_2 & r_0 \end{bmatrix} \geq 0 \quad (2.28)$$

which is a constraint that is not imposed on covariance functions of time series. Collecting all of the constraints on r_n into a useful form is not an easy task. However, as we develop in this paper, in analogy with the situation for time series, there is an alternative method for characterizing valid covariance sequences based on the generation of a sequence of reflection coefficients which must satisfy a far simpler set of constraints which once again differ somewhat from those in the time series setting.

D. Models for Stochastic Processes on Trees

As for time series it is of considerable interest to develop white-noise-driven models for processes on trees. The most general input-output form for such a model is simply

$$Y_t = \sum_{s\in\mathfrak{I}} c_{t,s} W_s \quad (2.29)$$

where W_t is a white noise process with unit variance and the $c_{t,s}$ are (real) weights defining the generalized convolution. In general, the output of this system is not isotropic and it is of interest to find models that do produce isotropic processes. One class introduced in [3] has the form

$$Y_t = \sum_{s\in\mathfrak{I}} c_{d(s,t)} W_s. \quad (2.30)$$

To show that this is isotropic, let (s, t) and (s', t') be two pairs of points such that $d(s, t) = d(s', t')$. By Lemma 2.1 there exists an isometry f so that $f(s) = s', f(t) = t'$. Then

$$\begin{aligned} E[Y_{s'} Y_{t'}] &= \sum_u c_{d(s',u)} c_{d(t',u)} \\ &= \sum_{u'} c_{d(s',f(u'))} c_{d(t',f(u'))} \\ &= \sum_{u'} c_{d(f(s),f(u'))} c_{d(f(t),f(u'))} \\ &= \sum_{u'} c_{d(s,u')} c_{d(t,u')} = E[Y_s Y_t]. \end{aligned} \quad (2.31)$$

The class of systems of the form of (2.30) is the generalization of the class of zero-phase LTI systems (i.e., systems with impulse responses of the form $h(t, s) = h(|t - s|)$). On the other hand, we know that for time series any LTI stable system, and in particular any causal, stable system, yields a stationary output when driven by white noise. As indicated previously, a major objective of this paper is to find the class of causal models on trees that produce isotropic processes when driven by white noise. Such a class of models will then also provide us with the counterpart of the Wold decomposition of a time series as a weighted sum of "past" values of a white noise process.

A logical starting point for such an investigation is the class of models introduced in Section II-B

$$Y_t = (SW)_t, \qquad S = \sum_{w\in\mathfrak{L}} s_w \cdot w. \quad (2.32)$$

However, it is not true that Y_t is isotropic for an arbitrary choice of S. For example, if $S = 1 + a\gamma^{-1}$, it is straightforward to check that Y_t is not isotropic. Thus we must look for a subset of this class of models. As we will see the correct model set is the class of autoregressive (AR) processes, where an AR process of order p has the form

$$Y_t = \sum_{\substack{w<0 \\ |w|\leq p}} a_w Y_{tw} + \sigma W_t \quad (2.33)$$

where W_t is a white noise with unit variance.

The form of (2.33) deserves some comment. First note that the constraints placed on w in the summation of (2.33) state that Y_t is a linear combination of the white noise W_t and the values, Y_{tw}, of Y at nodes that are both at distances at most p from t (i.e., $|w| \leq p$) and also on the same or previous horocycles ($w \leq 0$). Thus the model (2.33) is not strictly "causal" and is indeed an implicit specification since values of Y on the same horocycle depend on each other through (2.33) (see the second-order example to follow).

A question that then arises is: why not look instead at models in which Y_t depends only on its "strict" past, i.e.,

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

on points of the form $t\gamma^{-n}$. As shown in Appendix A, the additional constraints required of isotropic processes makes this class quite small. Specifically consider an isotropic process Y_t that does have this strict dependence:

$$Y_t = \sum_{n=0}^{\infty} a_n W_{t\gamma^{-n}}. \tag{2.34}$$

In Appendix A we show that the coefficients a_n must be of the form

$$a_n = \sigma a^n \tag{2.35}$$

so that the only process with strict past dependence as in (2.34) is the AR (1) process

$$Y_t = aY_{t\gamma^{-1}} + \sigma W_t. \tag{2.36}$$

Consider next the AR (2) processes, which specializing (2.33), has the form

$$Y_t = a_1 Y_{t\gamma^{-1}} + a_2 Y_{t\gamma^{-2}} + a_3 Y_{t\delta} + \sigma W_t. \tag{2.37}$$

Note first that this is indeed an implicit specification, since if we evaluate (2.37) at $t\delta$ rather than t we see that

$$Y_{t\delta} = a_1 Y_{t\gamma^{-1}} + a_2 Y_{t\gamma^{-2}} + a_3 Y_t + \sigma W_{t\delta}. \tag{2.38}$$

We can, of course, solve the pair (2.37), (2.38) to obtain the explicit formulas:

$$Y_t = \left(\frac{a_1}{1 - a_3}\right) Y_{t\gamma^{-1}} + \left(\frac{a_2}{1 - a_3}\right) Y_{t\gamma^{-2}} + \sigma V_t \tag{2.39}$$

$$Y_{t\delta} = \left(\frac{a_1}{1 - a_3}\right) Y_{t\gamma^{-1}} + \left(\frac{a_2}{1 - a_3}\right) Y_{t\gamma^{-2}} + \sigma V_{t\delta} \tag{2.40}$$

where

$$V_t = \frac{1}{1 - a_3^2}\{W_t + a_3 W_{t\delta}\}. \tag{2.41}$$

The structure of the representation (2.39)–(2.40) reveals that AR processes may be produced by propagating downward "wavefronts" of computations for those values of Y at one horocycle that all depend on the same values of Y at the preceding horocycle and the same set of values of W. Such vector representations may be of interest in some contexts such as in [16] in which we use similar but nonisotropic models to analyze some estimation problems. On the other hand, note that V_t is correlated with $V_{\delta t}$ and is uncorrelated with other values of V and thus is not an isotropic process (since $E[V_t V_{\gamma^{-2}t}] \neq E[V_t V_{\delta t}]$). In what follows in this paper and in part II we develop an alternate explicit representation for AR processes in which we again will encounter vector processes capturing the wavefront character of the computations but in which the driving noise will in fact be isotropic and white.

Another important point to note is that the second-order AR (2) model has four coefficients—three a's and σ, while for time series there would only be two a's. Indeed a simple calculation shows that the number of coefficients in our AR (p) model grows geometrically with the order, p, of the regression rather than linearly as for time series. On the other hand, these coefficients in our AR model are not independent and indeed there exist nonlinear relationships among the coefficients. For example for the second-order model (2.37) $a_3 \neq 0$ if $a_2 \neq 0$ since we know that the only isotropic process with strict past dependence is AR (1). Indeed, as shown in [6], the coefficients a_1, a_2, and a_3 must satisfy a fourth-order polynomial relation.

Because of the complex relationship among the a_w's in (2.33), the representation is not a completely satisfactory parameterization of this class of models. As we will see in subsequent sections, an alternate parametrization, provided by a generalization of Schur and Levinson recursions, provides us with a much better parametrization. In particular, this parametrization involves a sequence of reflection coefficients for AR processes on trees where exactly one new reflection coefficient is added as the AR order is increased by one.

III. Forward and Backward Prediction Errors and Levinson Recursions for Isotropic Processes on Trees

As outlined in the preceding section the direct parametrization of isotropic AR models in terms of their coefficients $\{a_w\}$ is not completely satisfactory since the number of coefficients grows exponentially with the order p, and at the same time there is a growing number of nonlinear constraints among the coefficients. In this and the following section we develop an alternate characterization involving one new coefficient when the order is increased by one. This development is based on the construction of "prediction" filters of increasing order, in analogy with the procedures developed for time series [9], [10] that lead to lattice filter models and whitening filters for AR processes. As is the case for time series, the single new parameter introduced at each stage, which we will also refer to as a reflection coefficient, is not subject to complex constraints involving reflection coefficients of other orders. Therefore, in contrast to the case of time series for which either the reflection coefficient representation or the direct parametrization in terms of AR coefficients are "canonic" (i.e., there are as many degrees of freedom as there are coefficients), the reflection coefficient representation for processes on trees appears to be the only natural canonic representation. Also, as for time series, we will see that each reflection coefficient is subject to bounds on its value which capture the constraint that r_n must be a valid covariance function of an isotropic process. Since this is a more severe and complex constraint on r_n than arises for time series, one would expect that the resulting bounds on the reflection coefficients would be somewhat different. This is the case, although somewhat surprisingly the constraints involve only a very simple modification to those for time series.

As for time series the recursion relations that yield the reflection coefficients arise from the development of forward and backward prediction error filters for Y_t. One crucial difference with time series is that the dimension of

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

the output of these prediction error filters increases with increasing filter order. This is a direct consequence of the structure of the AR model (2.33) and the fact that unlike the real line, the number of points at distance p from a node on a tree increases geometrically with p. For example, from (2.37)–(2.41) we see that Y_t and $Y_{t\delta}$ are closely coupled in the AR (2) model, and thus their prediction might best be considered simultaneously. For higher orders the coupling involves (a linearly growing number of) additional Y's. In this section we set up the proper definitions of these vectors of forward and backward prediction variables, and develop Levinson-like recursions for these as the order of prediction increases. Thanks to the constraints of isotropy we will see in the next section that the required projections in the Levinson recursions involve only one new coefficient as the filter order is increased by one.

A. Forward and Backward Prediction Errors

Let Y_t be an isotropic process on a tree, and let $\mathcal{H}\{\cdots\}$ denote the linear span of the random variables indicated between the braces. As developed in [10], the basic idea behind the construction of prediction models of increasing orders for time series is the construction of the past of a point t: $\mathcal{Y}_{t,n} = \mathcal{H}(\{Y_{t-k} | 0 \le k \le n\}$ and the consideration of the sequences of spaces as n increases. In analogy with this, we define the past of the node t on our tree:

$$\mathcal{Y}_{t,n} \triangleq \mathcal{H}\{Y_{tw}: w \preceq 0, |w| \le n\}. \tag{3.1}$$

One way to think of the past for time series is to take the set of all points within a distance n of t and then to discard the future points. This is exactly what (3.1) is: $\mathcal{Y}_{t,n}$ contains all points $\mathcal{Y}_s$ on previous horocycles ($s \prec t$) and on the same horocycle ($s \precsim t$) as long as $d(s, t) \le n$. A critical point to note is that in going from $\mathcal{Y}_{t,n-1}$ to $\mathcal{Y}_{t,n}$ we add new points on the same horocycle as t if n is even but not if n is odd (see the example to follow and Figs. 4–7).

In analogy with the time series case, the backward innovations or prediction error space, which we denote by $\mathcal{F}_{t,n}$, is defined as the space spanning the new information in $\mathcal{Y}_{t,n}$ which is orthogonal to $\mathcal{Y}_{t,n-1}$:

$$\mathcal{Y}_{t,n} = \mathcal{Y}_{t,n-1} \oplus \mathcal{F}_{t,n} \tag{3.2}$$

so that $\mathcal{F}_{t,n}$ is the orthogonal complement of $\mathcal{Y}_{t,n-1}$ in $\mathcal{Y}_{t,n}$ which we also denote by $\mathcal{F}_{t,n} = \mathcal{Y}_{t,n} \ominus \mathcal{Y}_{t,n-1}$. Define the backward prediction errors for the "new" elements of the "past" introduced at the nth step, i.e., for $w \preceq 0$ and $|w| = n$, we define

$$F_{t,n}(w) \triangleq Y_{tw} - E(Y_{tw} | \mathcal{Y}_{t,n-1}) \tag{3.3}$$

where $E(x | \mathcal{Y})$ denotes the linear least squares estimate of x based on data spanning $\mathcal{Y}$. Then

$$\mathcal{F}_{t,n} = \mathcal{H}\{F_{t,n}(w): |w| = n, w \preceq 0\}. \tag{3.4}$$

For time series the forward innovations process is the difference between Y_t and its estimate based on the past of Y_{t-1}. In a similar fashion, define the forward innovations

$$E_{t,n}(w) \triangleq Y_{tw} - E(Y_{tw} | \mathcal{Y}_{t\gamma^{-1},n-1}) \tag{3.5}$$

where w ranges over a set of words such that tw is on the same horocycle as t and at a distance at most $n-1$ from t (so that $\mathcal{Y}_{t\gamma^{-1},n-1}$ is the past of that point as well), i.e., $|w| < n$ and $w \precsim 0$. Define

$$\mathcal{E}_{t,n} \triangleq \mathcal{H}\{E_{t,n}(w): |w| < n \text{ and } w \precsim 0\}. \tag{3.6}$$

Let $E_{t,n}$ denote the column vector of the $E_{t,n}(w)$. A simple calculation shows that

$$\dim E_{t,n} = 2^{[(n-1)/2]} \tag{3.7}$$

where $[x]$ denotes the largest integer $\le x$. The elements

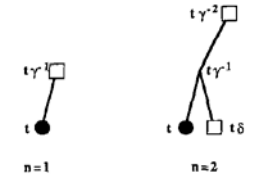

Fig. 4. Illustrating $E_{t,n}$ (dots) and $F_{t,n}$ (squares) for $n = 1, 2$.

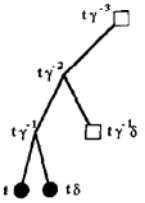

Fig. 5. Illustrating $E_{t,3}$ (dots) and $F_{t,3}$ (squares).

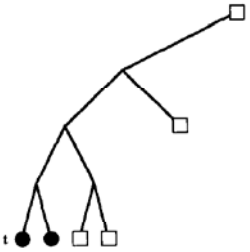

Fig. 6. Illustrating $E_{t,4}$ (dots) and $F_{t,4}$ (squares).

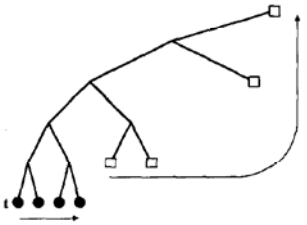

Fig. 7. Illustrating $E_{t,5}$ (dots) and $F_{t,5}$ (squares).

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

of $E_{t,n}$ are ordered according to a dyadic representation of the words w for which $|w| < n$, $w \neq 0$. Specifically, any such w other than 0 must have the form

$$w = \delta^{(i_k)} \cdots \delta^{(i_2)} \delta^{(i_1)} \tag{3.8}$$

with

$$1 \leq i_1 < i_2 < \cdots < i_k < \left[\frac{n}{2}\right] \tag{3.9}$$

and with $|w| = 2i_k$. For example, the points tw for $w = 0$, δ, $\delta^{(2)}$, and $\delta^{(2)}\delta$ are indicated by black circles in Fig. 7 and are ordered from left to right as indicated by the arrow.[2] Thus the words w of interest are in one-to-one correspondence with the numbers 0 and $\Sigma_{j=1}^{k} 2^{i_j}$, which provides us with our ordering.

In a similar fashion, let $F_{t,n}$ denote the column vector of the $F_{t,n}(w)$. In this case

$$\dim F_{t,n} = 2^{\lfloor n/2 \rfloor}. \tag{3.10}$$

The elements of $F_{t,n}$ are ordered as follows. Note that any word w for which $|w| = n$ and $w \not\leq 0$ can be written as $w = \gamma^{-k}\bar{w}$ for some $k \geq 0$ and $\bar{w} \neq 0$. For example, as illustrated in Fig. 7, for $n = 5$ the set of such w's is ($\gamma^{-1}\delta^{(2)}$, $\gamma^{-1}\delta^{(2)}\delta$, $\gamma^{-3}\delta$, and γ^{-5}). We order the w's as follows: first we group them in order of increasing k and then for fixed k we use the same ordering as for $E_{t,n}$ on the $\bar{w}$.

Example 3.1: In order to illustrate the geometry of the problem, consider the cases $n = 1, 2, 3, 4, 5$. The first two are illustrated in Fig. 4 and the last three are in Figs. 5–7, respectively. In each figure the points comprising $E_{t,n}$ are marked with dots, while those forming $F_{t,n}$ are indicated by squares.

$n = 1$ (see Fig. 4): To begin we have

$$\mathcal{Y}_{t,0} = \mathcal{H}\{Y_t\}.$$

The only word w for which $|w| = 1$ and $w \not\leq 0$ is $w = \gamma^{-1}$. Therefore

$$\begin{aligned} F_{t,1} &= F_{t,1}(\gamma^{-1}) \\ &= Y_{t\gamma^{-1}} - E(Y_{t\gamma^{-1}}|Y_t). \end{aligned}$$

Also

$$\mathcal{Y}_{t\gamma^{-1},0} = \mathcal{H}\{Y_{t\gamma^{-1}}\}$$

and the only word w for which $|w| < 1$ and $w \neq 0$ is $w = 0$. Thus

$$\begin{aligned} E_{t,1} &= E_{t,1}(0) \\ &= Y_t - E(Y_t|Y_{t\gamma^{-1}}). \end{aligned}$$

$n = 2$ (see Fig. 4): Here

$$\mathcal{Y}_{t,1} = \mathcal{H}\{Y_t, Y_{t\gamma^{-1}}\}.$$

[2]The left-to-right ordering of these points is due only to the fact that t was taken at the left of the horocycle.

In this case $|w| = 2$ and $w \not\leq 0$ implies that $w = \delta$ or γ^{-2}. Thus

$$\begin{aligned} F_{t,2} &= \begin{pmatrix} F_{t,2}(\delta) \\ F_{t,2}(\gamma^{-2}) \end{pmatrix} \\ &= \begin{pmatrix} Y_{t\delta} - E(Y_{t\delta}|Y_t, Y_{t\gamma^{-1}} \\ Y_{t\gamma^{-2}} - E(Y_{t\gamma^{-2}}|Y_t, Y_{t\gamma^{-1}}) \end{pmatrix}. \end{aligned}$$

Similarly,

$$\mathcal{Y}_{t\gamma^{-1},1} = \mathcal{H}\{Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}\}$$

and 0 is the only word satisfying $|w| < 2$ and $w \neq 0$. Hence

$$\begin{aligned} E_{t,2} &= E_{t,2}(0) \\ &= Y_t - E(Y_t|Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}). \end{aligned}$$

$n = 3$ (see Fig. 5): In this case

$$\mathcal{Y}_{t2} = \mathcal{H}\{Y_t, Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\delta}\}$$

$$\begin{aligned} F_{t,3} &= \begin{pmatrix} F_{t,3}(\gamma^{-1}\delta) \\ F_{t,3}(\gamma^{-3}) \end{pmatrix} \\ &= \begin{pmatrix} Y_{t\gamma^{-1}\delta} - E(Y_{t\gamma^{-1}\delta}|Y_t, Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\delta}) \\ Y_{t\gamma^{-3}} - E(Y_{t\gamma^{-3}}|Y_t, Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\delta}) \end{pmatrix}. \end{aligned}$$

Also

$$\mathcal{Y}_{t\gamma^{-1},2} = \mathcal{H}\{Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\gamma^{-3}}, Y_{t\gamma^{-1}\delta}\}$$

and there are two words, namely, 0 and δ, satisfying $|w| < 3$ and $w \neq 0$.

$$\begin{aligned} E_{t,3} &= \begin{pmatrix} E_{t,3}(0) \\ E_{t,3}(\delta) \end{pmatrix} \\ &= \begin{pmatrix} Y_t - E(Y_t|Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\gamma^{-3}}, Y_{t\gamma^{-1}\delta}) \\ Y_{t\delta} - E(Y_{t\delta}|Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\gamma^{-3}}, Y_{t\gamma^{-1}\delta}) \end{pmatrix}. \end{aligned}$$

$n = 4$ (see Fig. 6):

$$\mathcal{Y}_{t,3} = \mathcal{H}\{Y_t, Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\delta}, Y_{t\gamma^{-3}}, Y_{t\gamma^{-1}\delta}\}$$

$$F_{t,4} = \begin{pmatrix} F_{t,4}(\delta^{(2)}) \\ F_{t,4}(\delta^{(2)}\delta) \\ F_{t,4}(\gamma^{-2}\delta) \\ F_{t,4}(\gamma^{-4}) \end{pmatrix}$$

$$\mathcal{Y}_{t\gamma^{-1},3} = \mathcal{H}\{Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\gamma^{-3}}, Y_{t\gamma^{-4}}, Y_{t\gamma^{-1}\delta}, Y_{t\gamma^{-2}\delta}\}$$

$$E_{t,4} = \begin{pmatrix} E_{t,4}(0) \\ E_{t,4}(\delta) \end{pmatrix}.$$

$n = 5$ (see Fig. 7):

$$\begin{aligned} \mathcal{Y}_{t,4} = \mathcal{H}\{&Y_t, Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{\delta t}, Y_{t\gamma^{-3}}, \\ &\cdot Y_{t\gamma^{-1}\delta}, Y_{t\gamma^{-4}}, Y_{t\gamma^{-2}\delta}, Y_{t\gamma^{-1}\delta\alpha}, Y_{t\gamma^{-1}\delta\beta}\} \end{aligned}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

$$F_{t,5} = \begin{pmatrix} F_{t5}(\gamma^{-1}\,\delta^{(2)}) \\ F_{t,5}(\gamma^{-1}\,\delta^{(2)}\,\delta) \\ F_{t,5}(\gamma^{-3}\,\delta) \\ F_{t,5}(\gamma^{-5}) \end{pmatrix}$$

$$\mathcal{Y}_{t\gamma^{-1},4} = \mathcal{H}\{Y_{t\gamma^{-1}},\ Y_{t\gamma^{-2}},\ Y_{t\gamma^{-3}},\ Y_{t\gamma^{-1}\delta},\ Y_{t\gamma^{-4}},\ \cdot\ Y_{t\gamma^{-2}\delta},\ Y_{t\gamma^{-5}},\ Y_{t\gamma^{-3}\delta},\ Y_{t\gamma^{-2}\delta\alpha},\ Y_{t\gamma^{-2}\delta\beta}\}$$

$$E_{t,5} = \begin{pmatrix} E_{t,5}(0) \\ E_{t,5}(\delta) \\ E_{t,5}(\delta^{(2)}) \\ E_{t,5}(\delta^{(2)}\,\delta) \end{pmatrix}.$$

Let us make a few comments about the structure of these prediction error vectors. Note first that for n odd, dim $F_{t,n}$ = dim $E_{t,n}$, while for n even dim $F_{t,n}$ = 2 dim $E_{t,n}$. Indeed, for n even $F_{t,n}$ includes some points on the same horocycle as t (namely, tw for $|w| = n$, $w \not\simeq 0$), e.g., for $n = 2$, $F_{t,2}(\delta)$ is an element of $F_{t,2}$. These are the points that are on the backward-expanding boundary of the "past." At the next stage, however, these points become part of $E_{t,n}$, e.g., for $n = 3$, $E_{t,3}(\delta)$ is an element of $E_{t,3}$. This captures the fact mentioned previously that as the order of an AR model increases, an increasing number of points on the same horocycle is coupled.

As a second point, note that we have already provided a simple interpretation (3.2) of $\mathfrak{F}_{t,n}$ as an orthogonal complement. As for time series, this will be crucial in the development of our recursions. We will also need similar representations for $\mathcal{E}_{t,n}$. It is straightforward to check that for n odd

$$\mathcal{Y}_{t,n} \ominus \mathcal{Y}_{t\gamma^{-1},n-1} = \mathcal{E}_{t,n} \tag{3.11}$$

(this can be checked for $n = 1$ and 3 from example 3.1), while for n even

$$\mathcal{Y}_{t,n} \ominus \mathcal{Y}_{t\gamma^{-1},n-1} = \mathcal{E}_{t,n} + \mathcal{E}_{t\delta^{(n/2)},n} \tag{3.12}$$

note that the right-hand side of (3.12) is not, however, an orthogonal sum, see part II. For example for $n = 2$ this can be checked from the calculations in example 3.1 plus the fact that

$$E_{t\delta,2} = Y_{t\delta} - E[Y_{t\delta}|Y_{t\gamma^{-1}},\ Y_{t\gamma^{-2}}].$$

Finally, it is important to note that the process $E_{t,n}$ (for n fixed) is not in general an isotropic process (we will provide a counterexample shortly). However, if Y_t is AR (p) and $n \geq p$, then, after an appropriate normalization $E_{t,n}$ is white noise. This is in contrast to the case of time series in which case the prediction errors for all order models are stationary (and become white if $n \geq p$). In the case of processes on trees $E_{t,n}$ has statistics that are in general invariant with respect to some of the isometries of $\mathfrak{Z}$ but not all of them.

B. Calculation of Prediction Errors by Levinson Recursions on the Order

We are now in a position to develop recursions in n for the $F_{t,n}(w)$ and $E_{t,n}(w)$. Our approach follows that for time series except that we must deal with the more complex geometry of the tree. In particular, because of this geometry and the changing dimensions of $F_{t,n}$ and $E_{t,n}$, it is necessary to distinguish the cases of n even and n odd.

1) n Even: Consider first $F_{t,n}(w)$ for $|w| = n$, $w \not\leq 0$. There are two natural subclasses for these words w. In particular, either $w < 0$ or $w \not\simeq 0$.

Case 1: Suppose that $w < 0$. Then $w = \gamma^{-1}\tilde{w}$ for some $\tilde{w} \not\leq 0$ with $|\tilde{w}| = n - 1$. We then can perform the following computation, using (3.3) and properties of orthogonal projections:

$$\begin{aligned} F_{t,n}(w) &= F_{t,n}(t\gamma^{-1}\tilde{w}) = Y_{t\gamma^{-1}\tilde{w}} - E(Y_{t\gamma^{-1}\tilde{w}}|\,\mathcal{Y}_{t,n-1}) \\ &= Y_{t\gamma^{-1}\tilde{w}} - E(Y_{t\gamma^{-1}\tilde{w}}|\,\mathcal{Y}_{t\gamma^{-1},n-2}) \\ &\quad - E(Y_{t\gamma^{-1}\tilde{w}}|\,\mathcal{Y}_{t,n-1} \ominus \mathcal{Y}_{t\gamma^{-1},n-2}). \end{aligned}$$

Using (3.3) (applied at $t\gamma^{-1}$, $n - 1$) and (3.11) (applied at the odd integer $n - 1$), we then can compute

$$\begin{aligned} F_{t,n}(w) &= F_{t\gamma^{-1},n-1}(\tilde{w}) - E(Y_{t\gamma^{-1}\tilde{w}}|E_{t,n-1}) \\ &= F_{t\gamma^{-1},n-1}(\tilde{w}) - E(F_{t\gamma^{-1},n-1}(\tilde{w})|E_{t,n-1}) \end{aligned} \tag{3.13}$$

where the last equality follows from the orthogonality of $E_{t,n-1}$ and $\mathcal{Y}_{t\gamma^{-1},n-2}$ (from (3.11)). Equation (3.13) then provides us with a recursion for $F_{t,n}(w)$ in terms of variables evaluated at words $\tilde{w}$ of shorter length.

Case 2: Suppose that $w \not\simeq 0$. Then, since $|w| = n$, it is not difficult to see that $w = \delta^{(n/2)}\tilde{w}$ for some $\tilde{w}$ satisfying $|\tilde{w}| < n$, $\tilde{w} \not\simeq 0$ (for example, for $n = 4$, the only w satisfying $|w| = n$ and $w \not\simeq 0$ are $\delta^{(2)}$ and $\delta^{(2)}\,\delta$, see example 3.1). As in case 1 we have that

$$\begin{aligned} F_{t,n}(w) &= Y_{t\delta^{(n/2)}\tilde{w}} - E(Y_{t\delta^{(n/2)}\tilde{w}}|\,\mathcal{Y}_{t\gamma^{-1},n-2}) \\ &\quad - E(Y_{t\delta^{(n/2)}\tilde{w}}|\,\mathcal{Y}_{t,n-1} \ominus \mathcal{Y}_{t\gamma^{-1},n-2}). \end{aligned} \tag{3.14}$$

Now for n even we can show that

$$\mathcal{Y}_{t\gamma^{-1},n-2} = \mathcal{Y}_{t\delta^{(n/2)}\gamma^{-1},n-2}.$$

For example, for $n = 4$ these both equal $\{Y_{t\gamma^{-1}},\ Y_{t\gamma^{-2}},\ Y_{t\gamma^{-3}},\ Y_{t\gamma^{-1}\delta}\}$. Using this together with (3.5) and the orthogonality of $E_{t,n-1}$ and $\mathcal{Y}_{t\gamma^{-1},n-2}$ we can reduce (3.14) to

$$F_{t,n}(w) = E_{t\delta^{(n/2)},n-1}(\tilde{w}) - E(E_{t\delta^{(n/2)},n-1}(\tilde{w})|E_{t,n-1}) \tag{3.15}$$

which again expresses each $F_{t,n}(w)$ in terms of prediction errors evaluated at shorter words. As an additional comment, note that the number of words satisfying case 1 is the same as the number for case 2 (i.e., one-half dim $F_{t,n}$). Consider next $E_{t,n}(w)$ for $|w| < n$ and $w \not\simeq 0$. In this

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

case, we compute

$$E_{t,n}(w) = Y_{tw} - E(Y_{tw}|\,\mathcal{Y}_{t\gamma^{-1},n-2}) - E(Y_{tw}|\,\mathcal{Y}_{t\gamma^{-1},n-1} \ominus \mathcal{Y}_{t\gamma^{-1},n-2}) = E_{t,n-1}(w) - E(E_{t,n-1}(w)|F_{t\gamma^{-1},n-1}) \quad (3.16)$$

where the last equality follows from (3.2).

2) n Odd: Let us first consider the special case of $n = 1$ which will provide the starting point for our recursions. From example 3.1

$$F_{t,1} = Y_{t\gamma^{-1}} - E(Y_{t\gamma^{-1}}|Y_t) = Y_{t\gamma^{-1}} - k_1 Y_t = F_{t\gamma^{-1},0} - k_1 E_{t,0} \quad (3.17)$$

where k_1 is the first reflection coefficient, exactly as for time series

$$k_1 = \frac{E[Y_{t\gamma^{-1}} Y_t]}{E[Y^2_{t\gamma^{-1}}]} = \frac{r_1}{r_0}. \quad (3.18)$$

Similarly,

$$E_{t,1} = Y_t - E(Y_t|Y_{t\gamma^{-1}}) = Y_t - k_1 Y_{t\gamma^{-1}} = E_{t,0} - k_1 F_{t\gamma^{-1},0}. \quad (3.19)$$

Consider next the computation of $F_{t,n}(w)$ for $n \geq 3$ and odd. Note that for n odd it is impossible for w to satisfy $|w| = n$ and $w \not\simeq 0$. Therefore the condition

$$|w| = n \quad \text{and } w \not\preceq 0$$

is equivalent to

$$w = \gamma^{-1}\tilde{w}, \quad |\tilde{w}| = n - 1, \quad \tilde{w} \not\preceq 0.$$

Therefore, proceeding as before,

$$F_{t,n}(w) = Y_{t\gamma^{-1}\tilde{w}} - E(Y_{t\gamma^{-1}\tilde{w}}|\,\mathcal{Y}_{t\gamma^{-1},n-2}) - E(Y_{t\gamma^{-1},\tilde{w}}|\,\mathcal{Y}_{t,n-1} \ominus \mathcal{Y}_{t\gamma^{-1},n-2}) = F_{t\gamma^{-1},n-1}(\tilde{w}) - E(F_{t\gamma^{-1},n-1}(\tilde{w})|E_{t,n-1}, E_{t\delta^{((n-1)/2)},n-1}) \quad (3.20)$$

where the last equality follows from (3.12) applied at the even integer $n - 1$.

Consider next the computation of $E_{t,n}(w)$ for $n \geq 3$ and odd, and for $|w| < n$, $w \not\simeq 0$. There are two cases (each corresponding to one-half the components of $E_{t,n}$) depending upon whether $|w|$ is $n - 1$ or smaller.

Case 1: Suppose that $|w| < n - 1$. In this case, exactly the same type of argument yields

$$E_{t,n}(w) = E_{t,n-1}(w) - E(E_{t,n-1}(w) - E(E_{t,n-1}(w)|F_{t\gamma^{-1},n-1}). \quad (3.21)$$

Case 2: Suppose that $|w| = n - 1$. In this case $w = \delta^{((n-1)/2)}\tilde{w}$ where $\tilde{w} \not\simeq 0$ and computations analogous to those performed previously yield

$$E_{t,n}(w) = E_{t\delta^{((n-1)/2)},n-1}(\tilde{w}) - E(E_{t\delta^{((n-1)/2)},n-1}(\tilde{w})|F_{t\gamma^{-1},n-1}) \quad (3.22)$$

where in this case we use the fact that

$$\mathcal{Y}_{t\gamma^{-1},n-2} = \mathcal{Y}_{t\delta^{((n-1)/2)}\gamma^{-1},n-2}.$$

For example, for $n = 5$ these both equal

$$\{Y_{t\gamma^{-1}}, Y_{t\gamma^{-2}}, Y_{t\gamma^{-3}}, Y_{t\gamma^{-4}}, Y_{t\gamma^{-1}\delta}, Y_{t\gamma^{-2}\delta}\}.$$

We have now identified six formulas—(3.13), (3.15), (3.16), (3.20), (3.21), and (3.22)—for the order-by-order recursive computation of the forward and backward prediction errors. Of course, we must still address the issue of computing the projections defined in these formulas. As we make explicit in the next section the richness of the group of isometries and the constraints of isotropy provide the basis for a significant simplification of these projections by showing that we need only to compute projections onto the local averages or barycenters of the prediction errors. Moreover, scalar recursions for these barycenters provide us both with a straightforward method for calculating the sequence of reflection coefficients and with a generalization of the Schur recursions.

Finally, as mentioned previously $E_{t,n}$ is not, in general, an isotropic process unless Y_t is AR (p) and $n \geq p$, in which case it is white noise. To illustrate this, consider the computations of $E[E_{t,1}E_{t\delta,1}]$ and $E[E_{t,1}E_{t\gamma^{-2},1}]$ which should be equal if $E_{t,1}$ is isotropic. From (3.18), (3.19) we find that

$$E[E_{t,1}E_{t\delta,1}] = r_2 - \frac{r_1^2}{r_0}$$

while

$$E[E_{t,1}E_{t\gamma^{-2},1}] = r_2 - \frac{r_1^2}{r_0} + \frac{r_1(r_1 r_2 - r_0 r_3)}{r_0^2}.$$

In general these expressions are not equal to that $E_{t,1}$ is not isotropic. However, from the calculations in Appendix A we see that these expressions are equal and indeed $E_{t,1}$ is white noise if Y_t is AR (1). A stronger result that will be proved in the part II of this paper is that $E_{t,n}$, suitably normalized, is isotropic for all $n \geq p$ if and only if Y_t is AR (p).

IV. Prediction Error Barycenters, Reflection Coefficients, and Schur Recursions for Isotropic Processes on Trees

As previewed in the introduction, the various projections (3.13), (3.15), (3.16), (3.20)–(3.22) required in the Levinson recursions are more complex than their time series counterparts. Fortunately, thanks to the constraints of isotropy, these projections can be simplified considerably and indeed can all be computed in terms of projections onto scalar processes representing the barycenters of the vector prediction errors. In this section we prove this re-

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

sult and use it as the basis for a set of scalar Levinson recursions for the barycenter processes. Each stage of this recursion involves a single reflection coefficient, and we present a generalization of the Schur recursions which provide a procedure for computing the reflection coefficient sequence from the given isotropic covariance sequence.

A. Projections onto $\mathcal{E}$ and $\mathcal{F}$ and their Barycenters

Let us define the average values of the components of the prediction errors:

$$e_{t,n} = 2^{-\lfloor (n-1)/2 \rfloor} \sum_{|w|<n, w \precsim 0} E_{t,n}(w) \tag{4.1}$$

$$f_{t,n} = 2^{-\lfloor n/2 \rfloor} \sum_{|w|=n, w \preceq 0} F_{t,n}(w). \tag{4.2}$$

The following result is critical.

Lemma 4.1: The six collections of projections necessary for the order recursive computation of the prediction errors for all required words w and $\tilde{w}$ can be reduced to a total of four projections onto the barycenters of the prediction error vectors. In particular, as follows.

For n even: For any word w' such that $|w'| = n - 1$ and for any word w'' such that $|w''| < n$ and $w'' \precsim 0$, we have that

$$E(F_{t\gamma^{-1},n-1}(w')|E_{t,n-1}) = E(E_{t\delta^{(n/2)},n-1}(w'')|E_{t,n-1}) \tag{4.3}$$

$$= E(F_{t\gamma^{-1},n-1}(w_0)|e_{t,n-1}) \tag{4.4}$$

$$= E(E_{t\delta^{(n/2)},n-1}(0)|e_{t,n-1}) \tag{4.5}$$

(refer to (3.13), (3.15)) where w_0 is any of the w'. Also for any w such that $|w| < n$ and $w \precsim 0$, we have that

$$E(E_{t,n-1}(w)|F_{t\gamma^{-1},n-1}) = E(E_{t,n-1}(0)|f_{t\gamma^{-1},n-1}) \tag{4.6}$$

(refer to (3.16)).

For n odd: For any w' and w'' satisfying the constraints $|\cdot| < n$ and $\cdot \precsim 0$ we have that

$$E(E_{t,n-1}(w')|F_{t\gamma^{-1},n-1}) = E(E_{t\delta^{((n-1)/2)},n-1}(w'')|F_{t\gamma^{-1},n-1}) \tag{4.7}$$

$$= E(E_{t,n-1}(0)|f_{t\gamma^{-1},n-1}) \tag{4.8}$$

(refer to (3.21), (3.22)). In addition, for any $\tilde{w} \preceq 0$ such that $|\tilde{w}| = n - 1$

$$E(F_{t\gamma^{-1},n-1}(\tilde{w})|E_{t,n-1}, E_{t\delta^{((n-1)/2)},n-1}) = E(F_{t\gamma^{-1},n-1}(w_0)|\tfrac{1}{2}(e_{t,n-1} + e_{t\delta((n-1)/2),n-1})) \tag{4.9}$$

(refer to (3.20)) where w_0 is any of the $\tilde{w}$.

These results rely heavily on the structure of the dyadic tree, the isometry extension lemma, and the isotropy of Y. As an illustration consider the cases $n = 4$ and 5 illustrated in Figs. 8 and 9. Consider $n = 4$ first. Note that the distance relationships of each of the elements of $F_{t\gamma^{-1},3}$ and of $E_{t\delta^{(2)},3}$ to $E_{t,3}$ are the same. Furthermore, all

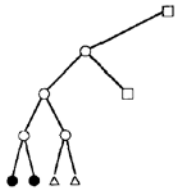

Fig. 8. Illustrating $E_{t,3}$ (dots), $F_{t\gamma^{-1},3}$ (squares), $E_{t\delta^{(2)},3}$ (triangles), and $\mathcal{Y}_{t\gamma^{-1},2}$ (circles).

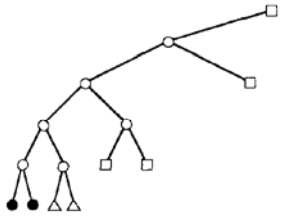

Fig. 9. Illustrating $E_{t,4}$ (dots), $E_{t\delta^{(2)},4}$ (triangles), $F_{t\gamma^{-1},4}$ (squares), and $\mathcal{Y}_{t\gamma^{-1},3}$ (circles).

three of these vectors contain errors in estimates based on $\mathcal{Y}_{t\gamma^{-1},2}$. Hence because of this symmetry and the isotropy of Y, the projections of any of the elements of $F_{t\gamma^{-1},3}$ or $E_{t\delta^{(2)},3}$ onto $E_{t,3}$ must be the same, as stated in (4.3). Furthermore, the two elements of $E_{t,3}$ have identical geometric relationship with respect to the elements of the other two error vectors. Hence the projections onto $E_{t,3}$ must weight its two elements equally, i.e., the projection must depend only on the average of the two, $e_{t,3}$, as stated in (4.4), (4.5). Similarly, the two elements of $F_{t\gamma^{-1},3}$ have identical geometric relations to each of the elements of $E_{t,3}$ so that (4.6) must hold. Similar geometric arguments apply to Fig. 9 and (4.7)–(4.9) evaluated at $n = 5$. Perhaps the only one deserving comment is (4.9). Note, however, in this case that each of the elements of $F_{t\gamma^{-1},4}$ has the same geometric relationship to all of the elements of $E_{t,4}$ and $E_{t\delta^{(2)},4}$ and therefore the projection onto the combined span of these elements must weight the elements of $E_{t,4}$ and $E_{t\delta^{(2)},4}$ equally and thus is a function of $(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1})/2$.

Proof of Lemma 4.1: As we have just illustrated the ideas behind each of the statements in the lemma are the same and thus we will focus explicitly only on the demonstration of (4.4). The other formulas are then obtained by analogous arguments.

The demonstration of (4.4) depends on the following three lemmas which are proved in Appendix B by exploiting symmetry and the isometry extension lemma.

Lemma 4.2: The best linear estimate

$$G_{t,n} = E(F_{t\gamma^{-1},n-1}(w)|E_{t,n-1}) \tag{4.10}$$

for n even is the same for all $|w| = n - 1$, $w \preceq 0$.

Lemma 4.3: The cross covariance

$$H_{t,n} = E(F_{t\gamma^{-1},n-1}(w)E_{t,n-1}(w')) \tag{4.11}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

is the same for all $|w| = n - 1$, $w \leq 0$ *and all* $|w'| < n$ *and* $w' \not\geq 0$.

Lemma 4.4: The covariance $\Sigma_{E,n}$ *of* $E_{t,n}$ *has the following structure. Let* $\Sigma(\alpha_0, \cdots, \alpha_d)$ *denote a* $2^d \times 2^d$ *covariance matrix, depending upon scalars* $\alpha_0, \cdots, \alpha_d$ *and with the following recursively defined structure:*

$$\Sigma(\alpha_0) = \alpha_0 \tag{4.12}$$

$$\Sigma(\alpha_0, \cdots, \alpha_d) = \begin{bmatrix} \Sigma(\alpha_0, \cdots, \alpha_{d-1}) & \alpha_d U_{d-1} \\ \alpha_d U_{d-1} & \Sigma(\alpha_0, \cdots, \alpha_{d-1}) \end{bmatrix} \tag{4.13}$$

where U_d *is a* $2^d \times 2^d$ *matrix all of whose values are* 1 *(i.e.,* $U_d = 1_d 1_d^T$ *where* 1_d *is a* 2^d*-dimensional vector of* 1*'s). Then there exist numbers* $\alpha_0, \alpha_1, \cdots, \alpha_{[(n-1)/2]}$ *so that*

$$\Sigma_{E,n} = \Sigma(\alpha_0, \cdots, \alpha_{[(n-1)/2]}). \tag{4.14}$$

From Lemma 4.2 we see that we need only show that $G_{t,n}$ depends only on $e_{t,n-1}$. However, from Lemma 4.4 it is a simple calculation to verify that $1_{[(n-1)/2]}$ is an eigenvector of $\Sigma_{E,n}$. Then, consider any $X \in \mathcal{E}_{t,n-1}$ of the form

$$X = \sum_{\substack{|w'| < n \\ w' \not\geq 0}} \lambda_{w'} E_{t,n-1}(w') \tag{4.15}$$

where

$$\sum_{\substack{|w'| < n \\ w' \not\geq 0}} \lambda_{w'} = 0. \tag{4.16}$$

Then, since $e_{t,n-1}$ is also as in (4.15) but with all λ_w equal, we have that

$$2^{[(n-1)/2]} E(X e_{t,n-1}) = (\lambda_{w'_1}, \cdots, \lambda_{w'_{2^{[(n-1)/2]}}}) \Sigma_{E,n} 1_{2^{[(n-1)/2]}} = 0. \tag{4.17}$$

Thus we have an orthogonal decomposition of $\mathcal{E}_{t,n-1}$ into the space spanned by X as in (4.15), (4.16) and the one-dimensional subspace spanned by $e_{t,n-1}$. However, thanks to Lemma 4.3, for any X satisfying (4.15), (4.16)

$$E[F_{t\gamma^{-1},n-1}(w) X] = \left(\sum_{w'} \lambda_{w'} \right) H_{t,n} = 0. \tag{4.18}$$

Thus the projection (4.10) is equal to the projection onto $e_{t,n-1}$, proving our result.

Remark: Lemma 4.4 allows us to say a great deal about the structure of $\Sigma_{E,n}$. In particular, it is straightforward to verify that the eigenvectors of $\Sigma_{E,n}$ are the discrete Haar basis [21]. For example, in dimension 8 the eigenvectors are the columns of the matrix

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \end{bmatrix} \tag{4.19}$$

Also, as shown in Section IV-A and in Appendix C, the structure of $\Sigma_{E,n}$ allows us to develop an extremely efficient procedure for calculating $\Sigma_{E,n}^{-1/2}$. Indeed this procedure involves a set of scalar computations and a recursive construction similar to the iterative construction of $\Sigma(\alpha_0, \alpha_1, \cdots, \alpha_d)$, with a total complexity of $O(l \log l)$, where $l = [(n-1)/2]$.

Finally, let us note an extremely important consequence of Lemma 4.1. Recall that the Levinson recursions developed in Section III involved projections of each

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

of the components of the vector prediction errors onto entire vectors of prediction error vectors. What Lemma 4.1 says most obviously is that we need only project onto the barycenters of prediction error vectors. For example, from (4.3)–(4.5) we see that for any $|w'| = n - 1$,

$$E(F_{t\gamma^{-1},n-1}(w')|E_{t,n-1}) = E(F_{t\gamma^{-1},n-1}(w')|e_{t,n-1}).$$

However, what this lemma also states is that this projection does not in fact depend on the specific choice of ω', so that it equals its barycenter. That is,

$$E(F_{t\gamma^{-1},n-1}(w')|E_{t,n-1}) = E(f_{t\gamma^{-1},n-1}|e_{t,n-1}).$$

Thus the required projections can be reduced completely to projections of scalars onto scalars. In the next two subsections we develop these purely scalar recursions, introducing the associated reflection coefficient sequence. In part II, these results allow us to develop lattice structures for the full prediction error processes.

B. Scalar Recursions for the Barycenters

As just indicated, an immediate consequence of Lemma 4.1, the definitions of the barycenters, and the computations in Section III-B is the following set of recursions for the barycenters themselves:

For n even:

$$e_{t,n} = e_{t,n-1} - E(e_{t,n-1}|f_{t\gamma^{-1},n-1}) \tag{4.20}$$

$$f_{t,n} = \tfrac{1}{2}(f_{t\gamma^{-1},n-1} + e_{t\delta^{(n/2)},n-1}) - \tfrac{1}{2}E(f_{t\gamma^{-1},n-1} + e_{t\delta^{(n/2)},n-1}|e_{t,n-1}). \tag{4.21}$$

For n odd, $n > 1$:

$$e_{t,n} = \tfrac{1}{2}(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}) - \tfrac{1}{2}E(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}|f_{t\gamma^{-1},n-1}) \tag{4.22}$$

$$f_{t,n} = f_{t\gamma^{-1},n-1} - E(f_{t\gamma^{-1},n-1}|\tfrac{1}{2}(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1})) \tag{4.23}$$

while for $n = 1$,

$$f_{t,1} \doteq F_{t,1}, \qquad e_{t,1} = E_{t,1} \tag{4.24}$$

and thus, (3.17)–(3.19) provide the necessary formulas.

It remains now to compute explicitly the projections indicated in (4.20)–(4.23). As the following result states, we only need compute one number k_n at each stage of the recursion, where k_n is the correlation coefficient between a variable being estimated and the variable on which the estimate is based. We've already seen this for $n = 1$ in (3.17)–(3.19), which yields also the first element of the sequence k_n which we refer to as the reflection coefficient sequence.

Theorem 4.1: For n even:

$$e_{t,n} = e_{t,n-1} - k_n f_{t\gamma^{-1},n-1} \tag{4.25}$$

$$f_{t,n} = \tfrac{1}{2}(f_{t\gamma^{-1},n-1} + e_{t\delta^{(n/2)},n-1}) - k_n e_{t,n-1} \tag{4.26}$$

where

$$\begin{aligned} k_n &= \operatorname{cor}(e_{t,n-1}, f_{t\gamma^{-1},n-1}) \\ &= \operatorname{cor}(e_{t\delta^{(n/2)},n-1}, e_{t,n-1}) \\ &= \operatorname{cor}(e_{t\delta^{(n/2)},n-1}, f_{t\gamma^{-1},n-1}) \end{aligned} \tag{4.27}$$

and $\operatorname{cor}(x, y) = E(xy)/[E(x^2)E(y^2)]^{1/2}$.

For n odd:

$$e_{t,n} = \tfrac{1}{2}(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}) - k_n f_{t\gamma^{-1},n-1} \tag{4.28}$$

$$f_{t,n} = f_{t\gamma^{-1},n-1} - \tfrac{1}{2}k_n(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}) \tag{4.29}$$

where

$$k_n = \operatorname{cor}(\tfrac{1}{2}(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}), f_{t\gamma^{-1},n-1}). \tag{4.30}$$

Keys to providing this result are the following two lemmas, the first of which is proven in Appendix B and the second of which can be proven in an analogous manner:

Lemma 4.5: For n odd:

$$E(e^2_{t\delta^{((n+1)/2)},n}) = E(e^2_{t,n}) = E(f^2_{t\gamma^{-1},n}) \triangleq \sigma^2_n. \tag{4.31}$$

Lemma 4.6: For n even $\frac{1}{2}(e_{t,n} + e_{t\delta^{(n/2)},n})$ *and* $f_{t\gamma^{-1},n}$ *have the same variance.*

Proof of Theorem 4.1: We begin with the case of n even. Since $n - 1$ is odd, Lemma 4.5 yields

$$E(e^2_{t,n-1}) = E(e^2_{t\delta^{(n/2)},n-1}) = E(f^2_{t\gamma^{-1},n-1}) \triangleq \sigma^2_{n-1}. \tag{4.32}$$

From (4.20)–(4.21) we than see that (4.25)–(4.27) are correct if

$$\begin{aligned} E[e_{t,n-1}f_{t\gamma^{-1},n-1}] &= E[e_{t\delta^{(n/2)},n-1}e_{t,n-1}] \\ &= E[e_{t\delta^{(n/2)},n-1}f_{t\gamma^{-1},n-1}] \triangleq g_{n-1} \end{aligned} \tag{4.33}$$

so that

$$k_n = \frac{g_{n-1}}{\sigma^2_{n-1}}. \tag{4.34}$$

However, the first equality in (4.33) follows directly from Lemma 4.1 while the second equality results from the first with t replaced by $t\delta^{(n/2)}$ and the fact that

$$\mathcal{F}_{t\gamma^{-1},n-1} = \mathcal{F}_{t\delta^{(n/2)}\gamma^{-1},n-1}. \tag{4.35}$$

For n odd the result directly follows from Lemma 4.6 and (4.22), (4.23).

Corollary: The variances of the barycenters satisfy the following recursions. For n even

$$\sigma^2_{e,n} = E(e^2_{t,n}) = (1 - k_n^2)\sigma^2_{n-1} \tag{4.36}$$

$$\sigma^2_{f,n} = E(f^2_{t,n}) = \left(\frac{1 + k_n}{2} - k_n^2\right)\sigma^2_{n-1} \tag{4.37}$$

where k_n *must satisfy*

$$-\tfrac{1}{2} \le k_n \le 1. \tag{4.38}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

For n odd

$$\sigma^2_{e,n} = \sigma^2_{f,n} = \sigma^2_n = (1 - k_n^2)\sigma^2_{f,n-1} \quad (4.39)$$

where

$$-1 \le k_n \le 1. \quad (4.40)$$

Proof: Equation (4.36) follows directly from (4.25) and (4.27) and the standard formulas for the estimation variance. Equation (4.37) follows in a similar way from (4.26) and (4.27) where the only slightly more complex feature is the use of (4.27) to evaluate the mean-squared value of the term in parentheses in (4.26). Equation (4.39) follows in a similar way from (4.28)–(4.30) and Lemma 4.6. The constraints (4.38) and (4.40) are immediate consequences of the nonnegativity of the various variances. Equality in one of these constraints yields the case of singular processes, i.e., processes for which some of the barycenter error processes are identically zero, corresponding to perfect prediction.

As we had indicated previously, the constraint of isotropy represents a significantly more severe constraint on the covariance sequence r_n. It is interesting to note that these additional constraints manifest themselves in the simple modification (4.38) of the constraint on k_n for n even over the form (4.40) that one also finds in the corresponding theory for time series. Also, as in the case of time series the satisfaction of (4.38) or (4.40) with equality corresponds to the class of deterministic or singular processes for which perfect prediction is possible. We will have more to say about these and related observations in the part II.

C. Schur Recursions and Computation of the Reflection Coefficients

As with the usual Levinson recursions for time series, we can use the recursions (4.25)–(4.26) and (4.28)–(4.29) for the barycenter error processes together with the definitions (4.27) and (4.30) of the reflection coefficients and the recursions (4.36), (4.37) and (4.39) to obtain explicit recursions for the computation of the k_n sequence directly for the given isotropic covariance sequence. We leave this straightforward computation to the reader and focus here on an alternative computational procedure generalizing the so-called Schur recursions [28], [37] for the cross-spectral densities between a given time series and its forward and backward prediction errors. In considering the generalization of these recursions to isotropic processes on trees, we must replace the z-transform power series for cross-spectral densities by corresponding formal power series of the type introduced in Section II. Specifically, for $n \ge 0$ define P_n and Q_n as

$$P_n \triangleq \text{cov}\,(Y_t, e_{t,n}) \triangleq \sum_{w \le 0} E(Y_t e_{tw,n}) \cdot w \quad (4.41)$$

$$Q_n \triangleq \text{cov}\,(Y_t, f_{t,n}) \triangleq \sum_{w \le 0} E(Y_t f_{tw,n}) \cdot w \quad (4.42)$$

where we begin with P_0 and Q_0 specified in terms of the correlation function r_n of Y_t:

$$P_0 = Q_0 = \sum_{w \le 0} r_{|w|} \cdot w. \quad (4.43)$$

Recalling the definitions (2.17), (2.18) of $\gamma[S]$ and $\delta^{(k)}[S]$ for S a formal power series and letting $S(0)$ denote the coefficient of $w = 0$, we have the following generalization of the Schur recursions.

Theorem 4.2: The following Schur recursions on formal power series yield the sequence of reflection coefficients

For n even

$$P_n = P_{n-1} - k_n \gamma[Q_{n-1}] \quad (4.44)$$

$$Q_n = \tfrac{1}{2}(\gamma[Q_{n-1}] + \delta^{(n/2)}[P_{n-1}]) - k_n P_{n-1} \quad (4.45)$$

where

$$k_n = \frac{\gamma[Q_{n-1}](0) + \delta^{(n/2)}[P_{n-1}](0)}{2P_{n-1}(0)}. \quad (4.46)$$

For n odd

$$P_n = \tfrac{1}{2}(P_{n-1} + \delta^{((n-1)/2)}[P_{n-1}]) - k_n \gamma[Q_{n-1}] \quad (4.47)$$

$$Q_n = \gamma[Q_{n-1}] - k_n \tfrac{1}{2}(P_{n-1} + \delta^{((n-1)/2)}[P_{n-1}]) \quad (4.48)$$

where

$$k_n = \frac{2\gamma[Q_{n-1}](0)}{P_{n-1}(0) + \delta^{((n-1)/2)}[P_{n-1}](0)}. \quad (4.49)$$

Proof: Note first that for $n = 1$, (4.47), (4.48) agree with (3.17)–(3.19) since $P_0 = \delta^{(0)}[P_0]$, $\gamma[Q_0](0) = r_1$, and $P_0(0) = r_0$. Next, since the proofs for n even and odd are essentially the same, we describe only the case of n even. To begin, write the recursions (4.25), (4.26) for tw instead of t. Then, premultiplying these recursions by Y_t, taking expectations, and summing over w, we get the recursions (4.44), (4.45). To derive the new expression (4.46) for the reflection coefficient, we note simply that from the definition of Q_n, we must have $Q_n(0) = 0$. Evaluating (4.45) at 0 then directly yields (4.46).

V. Conclusion

In the first part of this paper we have described a new framework for modeling and analyzing signals at multiple scales. Motivated by the structure of the computations involved in the theory of multiscale signal representations and wavelet transforms, we have examined the class of isotropic processes on a homogenous tree of order 2. Thanks to the geometry of this tree, an isotropic process possesses many symmetries and constraints. These make the class of isotropic autoregressive processes somewhat difficult to describe if we look only at the usual AR coefficient representation. In this paper we have presented the first half of the development of the generalization of lattice structures which provides a much better parametrization of AR processes on dyadic trees. In particular we

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

have developed Levinson recursions for forward and backward prediction error processes, where the "forward" and "backward" directions refer to finer or coarser scales, respectively. Because of the geometry of the tree these prediction errors are vector of processes of dimension increasing with the order of prediction. However, thanks to the symmetries required of isotropic processes, the required computation in these vector recursions can be directly related to those in the scalar recursions for the barycenters of the prediction error vectors. In this paper we have developed these scalar Levinson recursions and the corresponding Schur recursions for the computation of the required reflection coefficient sequence. In part II [1] we develop lattice structures for both whitening and modeling filters for isotropic processes and use these results to obtain a detailed analysis of AR processes and of the Wold decomposition of isotropic processes on trees.

Appendix A
AR (1) and Isotropic Processes with Strict Past Dependence

We wish to show that AR (1) processes are the only isotropic processes with strict past dependence. To do this let us introduce the notation $]-\infty, t]$ to denote the path from t back towards $-\infty$, i.e., the set $\{t\gamma^{-n} | n \geq 0\}$, and consider a process of the form

$$Y_t = \sum_{s \in]-\infty, t]} a_{d(t,s)} W_s \tag{A.1}$$

where W_t is unit variance white noise.

We now consider the conditions under which (A.1) is stationary. Let t_1 and t_2 be any two nodes, let $t = t_1 \wedge t_2$, and define the distances $n_1 = d(t_1, t)$, $n_2 = d(t_2, t)$. Note that $d(t_1, t_2) = n_1 + n_2$. Also let $r(t_1, t_2) = E(Y_{t_1} Y_{t_2})$. Then from (A.1), the fact that W_t is white, and the definition of t, n_1, and n_2, we have

$$\begin{aligned} r(t_1, t_2) &= \sum_{s_1 \in]-\infty, t_1]} \sum_{s_2 \in]-\infty, t_2]} a_{d(t_1, s_1)} a_{d(t_2, s_2)} E(W_{s_1} W_{s_2}) \\ &= \sum_{s \in]-\infty, t]} a_{d(t_1, s)} a_{d(t_2, s)} \\ &= \sum_{m \geq 0} a_{n_1 + m} a_{n_2 + m}. \end{aligned}$$

For Y_t to be isotropic we must have that

$$\begin{aligned} r(t_1, t_2) &= r(d(t_1, t_2)) \\ &= r(n_1 + n_2). \end{aligned}$$

Therefore, for $n_1 \geq 0$, $n_2 \geq 0$ we must have that

$$r(n_1 + n_2) = \sum_{m \geq 0} a_{n_1 + m} a_{n_2 + m}. \tag{A.2}$$

In particular, for $n \geq 2$ we can deduce from (A.2) that we have the following two relationships:

$$\begin{aligned} r(2n) &= r(n + n) \\ &= \sum_{m \geq 0} a_{n+m}^2 \\ &= r(2n - 2) - a_{n-1}^2 \end{aligned} \tag{A.3}$$

$$\begin{aligned} r(2n) &= r((n + 1) + (n - 1)) \\ &= \sum_{m \geq 0} a_{m+n+1} a_{m+n-1} \\ &= r(2n - 2) - a_{n-2} a_n \end{aligned} \tag{A.4}$$

from which we deduce that

$$a_n a_{n-2} = a_{n-1}^2, \qquad n \geq 2$$

or equivalently

$$\frac{a_n}{a_{n-1}} = \text{constant}, \qquad n \geq 1.$$

Thus $a_n = \sigma a^n$, so that

$$Y_t = \sum_{s \in]-\infty, t]} \sigma a^{d(t,s)} W_s$$

from which we immediately see that Y_t satisfies

$$Y_t = a Y_{t\gamma^{-1}} + \sigma W_t.$$

Appendix B
Properties of the Statistics of the Forward and Backward Residuals

In this Appendix we prove some of the results on the structure of the statistics of the prediction errors $E_{t,n}$ and $F_{t,n}$ and their barycenters. The keys to the proofs of all of these results, and to the others stated in Section IV without proof, are the constraints of isotropy and the construction of specific isometries.

1. *Proof of Lemma 4.2*

Let

$$G_{t,n}(w) \triangleq E(F_{t\gamma^{-1},n-1}(w) | E_{t,n-1}) \tag{B.1}$$

where n is even and $|w| = n - 1$, $w \preceq 0$. We wish to show that $G_{t,n}(w)$ is identical for all such w. By definition

$$G_{t,n}(w) = E([Y_{t\gamma^{-1}w} - E(Y_{t\gamma^{-1}w} | \mathcal{Y}_{t\gamma^{-1},n-2})] | E_{t,n-1}). \tag{B.2}$$

Define the set of nodes

$$\mathfrak{I}_{t,n} = \{s = tv;\ |v| \leq n,\ v \preceq 0\}. \tag{B.3}$$

The points $t\gamma^{-1}w$ in (B.2) correspond to the points $s = tv$ in $\mathfrak{I}_{t,n}$ with $|v| = n$. Let w', w'' be any two words satisfying $|w| = n - 1$, $w \prec 0$. Suppose that we can find a local isometry ϕ: $\mathfrak{I}_{t,n} \to \mathfrak{I}_{t,n}$ such that

$$\begin{aligned} \phi(t\gamma^{-1}w') &= t\gamma^{-1}w'' \\ \phi(t\gamma^{-1}w'') &= t\gamma^{-1}w' \\ \phi(t) &= t \\ \phi(\mathfrak{I}_{t,n-1}) &= \mathfrak{I}_{t,n-1}. \end{aligned} \tag{B.4}$$

By the isometry extension lemma ϕ can be extended to an isometry on $\mathfrak{I}$.

Consider $G_{t,n}(w')$ and $G_{\phi(t),n}(w'')$ which are linear projections onto, respectively, $E_{t,n-1}$ and $E_{\phi(t),n-1}$. Since the processes Y_t and $Y_{\phi(t)}$ have the same statistics, these two

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

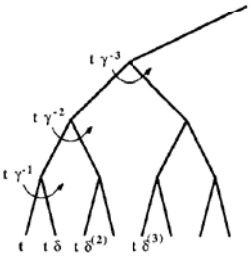

Fig. 10. Illustrating local isometries that interchange points on a given horocycle.

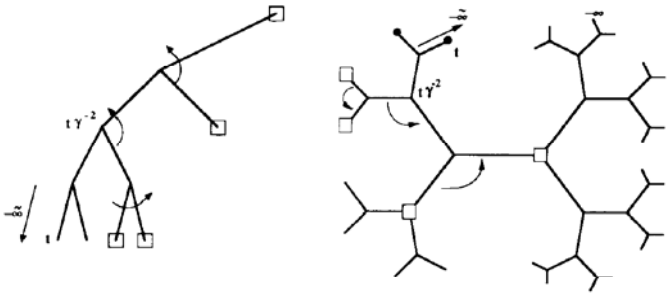

Fig. 11. (a) Illustrating i) the nodes involved in $\mathfrak{I}_{t,4}$; ii) the extreme points of $\mathfrak{I}_{t,4}$ (enclosed in boxes); iii) the alternate choice of $-\tilde{\infty}$ so that these extreme points are on the same horocycle and so that t is closer to $-\tilde{\infty}$ than the closest pivot point $t\gamma^{-2}$ (see the proof of Lemma 4.2 in Appendix B); (b) A redrawing of the tree of part (a) in a more symmetric fashion.

projection operators are identical. Furthermore, from (B.4) we see that $\phi(t) = t$ and $E_{\phi(t),n-1} = E_{t,n-1}$, so that we can conclude that $G_{t,n}(w') = G_{t,n}(w'')$.

Thus it remains to show that we can construct such local isometries for any such w' and w''. To do this, let us briefly reexamine the nature of local isometries that interchange points on a given horocycle. Consider the situation depicted in Fig. 10. One way to think of constructing an isometry interchanging t and $t\delta$ is to pivot the tree at $t\gamma^{-1}$ and "flip" or rotate everything below this point. Similarly by pivoting at $t\gamma^{-2}$ we interchange t and $t\delta^{(2)}$, while pivoting at $t\gamma^{-3}$ leads to an interchange of t and $t\delta^{(3)}$. Note several points about these isometries. First, by composing several of them we can interchange any two points on the same horocycle. Second, each of these isometries leaves fixed all nodes above and including the pivot node. Third, each of these isometries leaves globally invariant the set of nodes extending from the pivot node down to the portion of the horocycle in question that is descendent from the pivot. For example, each of the three pivot isometries described in the Fig. 10 maps the set of nodes drawn in this figure onto itself.

Next, let us recall that the notion of horocycle depends explicitly on the choice of the point $-\infty$. From the preceding discussion we will have completed the proof if we can show that there exists an alternate choice, say $-\tilde{\infty}$, such that 1) the extreme points tw in $\mathfrak{I}_{tn}$ with $|w| = n$ are on the same horocycle, and 2) the point t is "above," i.e., closer to $-\tilde{\infty}$ than, all of the pivot points involved in interchanging the above mentioned extreme points of $\mathfrak{I}_{t,n}$.

It is easily checked that these conditions are satisfied by any choice of $-\tilde{\infty}$ so that the path from $t\gamma^{-n/2}$ toward $-\tilde{\infty}$ passes through t. This is illustrated in Fig. 11 for $n = 4$. In Fig. 11(b) we have redrawn the tree of Fig. 11(a) in a more symmetric fashion so that the nature of the flips performed by these isometries is more apparent.

2. *Proof of Lemma 4.3*

Set

$$H_{t,m}(w, w') = E[F_{t\gamma^{-1},n-1}(w)E_{t,n-1}(w')] \quad \text{(B.5)}$$

where n is even $|w| = n - 1$, $w \underline{\leq} 0$ and $|w'| < n$, $w' \not\geq 0$. We wish to show that $H_{t,n}(w, w')$ is identical for all such w, w' pairs. An argument analogous to that in the preceding subsection shows that this will be true if we can construct two classes of isometries:

1) For any w_1, w_2 satisfying $|w| = n - 1$, $w \underline{\leq} 0$, $\phi(tw_1) = tw_2$, $\phi(tw_2) = tw_1$, ϕ leaves $\mathfrak{I}_{t,n-1}$ invariant and leaves fixed any point of the form tw', with $|w'| < n$, $w' \not\geq 0$.

2) For any w_1', w_2' satisfying $|w'| < n$, $w' \not\geq 0$, $\psi(t\gamma^{-1}w_1') = t\gamma^{-1}w_2'$, $\psi(t\gamma^{-1}w_2') = t\gamma^{-1}w_1'$, ψ leaves $\mathfrak{I}_{t\gamma^{-1},n-1}$ invariant and leaves fixed any point of the form $t\gamma^{-1}w$, with $|w| = n - 1$, $w \underline{\leq} 0$.

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

It is straightforward to check that the isometries constructed in the proof of Lemma 4.2 form a class satisfying 1. The construction of the second class of isometries is analogous to that used in the preceding section. However, in this case the points to be interchanged are already on the same horocycle (relative to the original choice of $-\infty$) and the points to be kept fixed are closer to $-\infty$ than any of the required pivot points (see Fig. 12 for the case $n = 4$). Thus the existence of the required isometries is immediate.

3. Proof of Lemma 4.4

As in Section II-C, let w_m denote the $2^{[(n-1)/2]}$ words such that $|w| < n$, $w \neq 0$, and for any two such words let

$$J_{t,n}(w_i, w_j) = E[E_{t,n}(w_i)E_{t,n}(w_j)]. \tag{B.6}$$

Let $n_1 = |w_1|$ and $n_2 = |w_2|$. Consider first the case when $n_1 \neq n_2$. What we must show in this case is that $J_{t,n}(w_i, w_j)$ is the same for all pairs w_i, w_j with these respective lengths. By an argument analogous to the ones used previously, this will be true if for any two pairs (w_i, w_j), (w_i', w_j') with $|w_i| = |w_i'| = n_1$, $|w_j| = |w_j'| = n_2$ we can find a local isometry ϕ of $\mathfrak{I}_{t,n}$ so that ϕ leaves $\mathfrak{I}_{t\gamma^{-1},n-1}$ invariant and performs the interchanges

$$tw_i \leftrightarrow tw_i', \quad tw_j \leftrightarrow tw_j'.$$

Direct calculations shows that the class of isometries ψ defined in the previous subsection contains the required isometry.

Suppose now that $|w_i| = |w_j| = n_1$, and let $s = d(tw_i, tw_j)$. An analogous argument shows that

$$J_{t,n}(w_i, w_j) = J_{t,n}(0, w_k), \qquad \text{where } |w_k| = s. \tag{B.7}$$

Again an appropriate element of the class of isometries ψ yields an isometry leaving $\mathfrak{I}_{t\gamma^{-1},n-1}$ invariant and performing the interchange

$$tw_i \leftrightarrow t, \; tw_j \leftrightarrow tw_k. \tag{B.8}$$

This finishes the proof of Lemma 4.4.

4. Proof of Lemmas 4.5 and 4.6

The proofs of the two equalities in (4.31) and the equality stated in Lemma 4.6 are all quite similar. We focus here explicitly only on showing that for n odd $e_{t,n}$ and $f_{t\gamma^{-1},n}$ have the same variance. This will be true if we can construct an isometry that interchanges the set of nodes associated with $E_{t,n}$ with the set of nodes associated with $F_{t\gamma^{-1},n}$ while leaving the set of nodes $\mathfrak{I}_{t\gamma^{-1},n-1}$ invariant. As in the proof of Lemma 4.2, the key here is to make a choice for an alternate boundary point $-\tilde{\infty}$ so that all of the relevant nodes are on the same horocycle with respect to this choice. The following construction does this. Consider the node $t\gamma^{-(n+1)/2}$. There are three directions defined from this node: one toward the original choice of $-\infty$, one toward t, and a third direction. Take $-\tilde{\infty}$ be any boundary point so that the path to it from $t\gamma^{-(n+1)/2}$ is along this third direction. Then the pivot about $t\gamma^{-(n+1)/2}$ with respect to $-\tilde{\infty}$ performs the required interchange and leaves the set of nodes $\mathfrak{I}_{t\gamma^{-1},n-1}$ invariant. This is illustrated in Fig. 13 for $n = 5$, where for simplicity of presentation we have displayed the tree in a symmetric fashion as in Fig. 11(b).

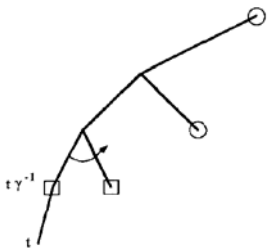

Fig. 12. Illustrating i) the node t together with the nodes comprising $\mathfrak{I}_{t\gamma^{-1},3}$; ii) the nodes (in squares) on the same horocycle to be interchanged by the isometry ψ needed in the proof of Lemma 4.3 (see Section III-B); and iii) the nodes (in circles) to be left fixed by ψ.

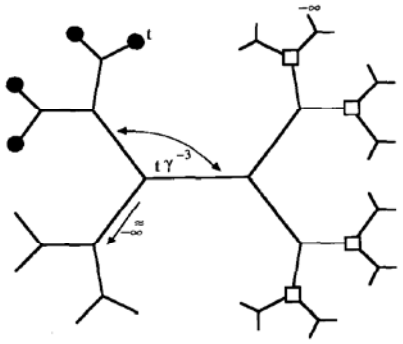

Fig. 13. Illustrating $E_{t,5}$ (dots), $F_{t\gamma^{-1},5}$ (squares), and the choice for $-\tilde{\infty}$ required in the proof of Lemma 4.5 (Appendix B). The required pivot, relative to $-\tilde{\infty}$, to interchange the dots and squares is at $t\gamma^{-3}$.

Appendix C
Calculation of $\Sigma^{-1/2}(\alpha_0 \cdots \alpha_k)$

We shall first make use of the following formula: for S and T symmetric matrices,

$$\begin{pmatrix} S & T \\ T & S \end{pmatrix}^{-1/2} = \frac{1}{2}\begin{pmatrix} X+Y & X-Y \\ X-Y & X+Y \end{pmatrix} \tag{C.1}$$

where

$$X = (S+T)^{-1/2}$$
$$Y = (S-T)^{-1/2}. \tag{C.2}$$

From (4.13) and (C.1), (C.2) we see that the computation of $\Sigma^{-1/2}(\alpha_0, \cdots, \alpha_k)$ can be performed by a simple construction from the inverse square roots of

$$\Sigma_+ = \Sigma(\alpha_0, \cdots, \alpha_{k-1}) + \alpha_k U_{k-1}$$
$$= \Sigma(\alpha_0 + \alpha_k, \cdots, \alpha_{k-1} + \alpha_k) \tag{C.3}$$
$$\Sigma_- = \Sigma(\alpha_0, \cdots, \alpha_{k-1}) - \alpha_k U_{k-1}$$
$$= \Sigma(\alpha_0 - \alpha_k, \cdots, \alpha_{k-1} - \alpha_k). \tag{C.4}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

If we introduce the following notation:

$$\text{Bloc}\,(X, Y) = \frac{1}{2}\begin{pmatrix} X+Y & X-Y \\ X-Y & X+Y \end{pmatrix} \qquad \text{(C.5)}$$

then $\Sigma^{-1/2}(\alpha_0, \cdots, \alpha_k)$ can be calculated via the following recursion:

$$\Sigma^{-1/2}(\alpha_0, \cdots, \alpha_k) = \begin{cases} \alpha_0^{-1/2} & \text{if } k = 0 \\ \text{Bloc}\,(\Sigma_+^{-1/2}, \Sigma_-^{-1/2}) & \text{if } k \geq 1 \end{cases} \qquad \text{(C.6)}$$

which involves a sequence of scalar calculations.

Acknowledgment

The authors would like to express their thanks to J.-P. Conze, Y. Guivarc'h, and A. Raugi from the Institut de Recherche en Mathématique de Rennes, for bringing to their attention the work of Arnaud, Cartier, Dunau, and LeTac on isotropic processes on the homogeneous tree, and also to K. Chou and B. Claus for fruitful discussions.

Some of the work of A. S. Willsky was performed while he was a visitor at INRIA.

References

[1] M. Basseville, A. Benveniste, and A. S. Willsky, "Multiscale autoregressive processes, Part II: Lattice structures for whitening and modeling," this issue, pp. 1935–1954.

[2] J. P. Arnaud, "Fonctions sphériques et fonctions définies positives sur l'arbre homogène," *C. R. Acad. Sci.*, t. 290, série A, pp. 99–101, Jan. 14, 1980.

[3] J. P. Arnaud and G. Letac, "La formule de représentation spectrale d'un processus Gaussien stationnaire sur un arbre homogène," Publ. Lab. Stat. and Prob., UA 745, Toulouse, France.

[4] M. Barnsley, *Fractals Everywhere*. San Diego, CA: Academic, 1988.

[5] M. Basseville and A. Benveniste, "Multiscale statistical signal processing," in *Proc. ICASSP-89*, (Glasgow, U.K.), 1989, pp. 2065–2068.

[6] M. Basseville, A. Benveniste, and A. S. Willsky, "Multiscale autoregressive processes," INRIA Res. Rep. 1206, 1990.

[7] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 766–784, 1992.

[8] A. Benveniste, R. Nikoukhah, and A. S. Willsky, "Multiscale system theory," in *Proc. 29th IEEE Conf. Decision Contr.* (Honolulu, HI), Dec. 1990.

[9] A. Benveniste, "Introduction: Estimation et factorisation spectrale, quelques points de vue féconds," in *Outils et Modèles Mathématiques pour l'Automatique, l'Analyse de Systèmes et le Traitement du Signal*, vol. 2. Editions du CNRS, 1982, pp. 231–266.

[10] A. Benveniste, "Méthodes d'orthogonalisation en treillis pour le problème de la réalisation stochastique," in *Outils et Modèles Mathematiques pour l'Automatique, l'Analyse de Systèmes et le Traitement du Signal*, vol. 2. Editions du CNRS, 1982 pp. 267–308.

[11] J. Berstel and C. Reutenauer, "Les séries rationnelles et leurs langages," (Collection "Etudes et Recherches en Informatique"). Masson, 1984.

[12] A. Brandt, "Multilevel adaptive solutions to boundary value problems," *Math. Comput.*, vol. 13, pp. 333–390, 1977.

[13] W. Briggs, *A Multigrid Tutorial*. Philadelphia, PA: SIAM, 1987.

[14] P. Cartier, "Harmonic analysis on trees," in *Proc. Symp. Pure Math.*, (Providence, RI), vol. 26, 1974, pp. 419–424.

[15] P. Cartier, "Géométrie et analyse sur les arbres," *Séminaire Bourbaki*, n407, Feb. 1972.

[16] K. C. Chou, A. S. Willsky, A. Benveniste, and M. Basseville, "Recursive and iterative estimation algorithms for multiresolution stochastic processes," in *Proc. IEEE Conf. Decision Contr.* (Tampa, FL), Dec. 1989.

[17] K. Chou, "A stochastic modeling approach to multiscale signal processing," Ph.D. dissertation, Dept. Elec. Eng. Comput. Sci., M.I.T., 1991.

[18] K. C. Chou and A. S. Willsky, "Multiscale Riccati equation and a two-sweep algorithm for the optimal fusion of multiresolution data," in *Proc. 29th IEEE Conf. Decision Contr.* (Honolulu, HI), Dec. 1990.

[19] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[20] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math. Phys.*, vol. 27, pp. 1271–1283, 1986.

[21] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 91, pp. 909–996, 1988.

[22] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, 1990.

[23] J. L. Dunau, "Etude d'une classe de marches aléatoires sur l'arbre homogène," in *Arbres homogènes et Couples de Gelfand*, J. P. Arnaud, J. L. Dunau, G. Letac. Publications du Laboratoire de Statistique et Probabilités, Univ. Paul Sabatier, Toulouse, France, no. 02–83, June 1983.

[24] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, pp. 85–102, 1984/1985.

[25] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integreable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, pp. 723–736, 1984.

[26] A. Grossmann, R. Kronland-Martinet, and J. Morlet, "Reading and understanding continuous wavelet transforms," in *Wavelets, Time-Frequency Methods, and Phase Space, 2-20, Proc. Int. Conf. Wavelets* (Marseilles, France), Dec. 14–18, 1987, 1989.

[27] W. Hackbusch and U. Trottenberg, Eds., *Multigrid Methods*. New York: Springer-Verlag, 1982.

[28] T. Kailath, "A theorem of I. Schur and its impact on modern signal processing," in *Schur Methods in Operator Theory and Signal Processing*, I. Gohberg Ed., Operator Theory: Advances and Applications, vol. 18. Basel, Boston, Stuttgart: Birkhäuser, 1986.

[29] R. Kronland-Martinet, J. Morlet, and A. Grossmann, "Analysis of sound patterns through wavelet transforms," *Int. J. Patt. Recog. Artificial Intell.*, vol. 1, no. 2, pp. 273–301, Aug. 1987.

[30] S. G. Mallat, "A compact multiresolution representation: The wavelet model," in *Proc. IEEE Workshop Comput. Vision* (Miami, FL), Dec. 1987.

[31] S. G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 7, pp. 674–693, 1989.

[32] S. G. Mallat, "Multiresolution approximation and wavelet orthonormal bases of L^2," *Trans. Amer. Math. Soc.*, June 1989.

[33] B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman, 1982.

[34] S. McCormick, *Multigrid Methods*, vol. 3 of the SIAM Frontiers Series. Philadelphia, PA: SIAM, 1987.

[35] Y. Meyer, "Wavelets and operators," in *Proc. Special Year Modern Anal. 1986/1987* (Urbana, IL), 1989.

[36] D. Paddon and H. Holstein, Eds., *Multigrid Methods for Integral and Differential Equations*. Oxford, England: Clarendon, 1985.

[37] E. A. Robinson and S. Treitel, "Maximum entropy and the relationship of the partial autocorrelation to the reflection coefficients of a layered system," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, no. 2, pp. 224–235, 1980.

[38] M. J. Smith and T. P. Barnwell, "Exact reconstruction techniques for tree-structured subband coders," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 434–441, 1986.

[39] P. P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions, and perfect reconstruction techniques," *IEEE ASSP Mag.*, vol. 4, no. 3, pp. 4–20, 1987.

[40] G. W. Wornell, "A Karhunen-Loeve-like expansion for $1/f$-processes via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, no. 9, pp. 859–861, July 1990.

[41] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 611–623, Mar. 1992.

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.

Michèle Basseville was born on May 2, 1952 in Paris, France. In 1976, she graduated from Ecole Normale Supérieure de Fontenay-aux-Roses, where she studied mathematics, and she received the Docteur-es-Sciences degree (thèse d'Etat) in 1982.

Since 1976, she has been with the Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), where she held a CNRS position, "Chargée de Recherche," from 1976 until 1990, and where she is now Directeur de Recherche at CNRS. Her main domain of interest has been the detection of abrupt changes in signals and dynamical systems, and its application to both recognition-oriented signal processing and industrial process monitoring for conditional maintenance. In 1982, she obtained the Docteur-es-Sciences degree (thèse d'Etat) for her contribution to several on-line signal segmentation algorithms, which have been proved useful since then, especially for geophysical signal processing and continuous speech recognition. Since 1980, she and A. Benveniste have been involved in a joint project on vibration monitoring for both large mechanical structures and rotating machines. Since 1988, they have been investigating other possible applications for monitoring, among them gas turbines. She has edited, together with A. Benveniste, the collective monograph *Detection of Abrupt Changes in Signals and Dynamical Systems* (Lecture Notes in Control and Information Sciences no. 77). She is currently writing a book, *Detection of Abrupt Changes—Theory and Applications*, with I. V. Nikiforov (IPOU, Moscow), which is to appear in the Prentice-Hall Information and System Sciences Series. Since 1987, she has been involved in the joint project INRIA/MIT on a statistical theory for multiresolution analysis of signals and images, with A. Benveniste and A. Willsky. Her main interest in this project is the development of statistical algorithms for the detection and recognition of events in multiscale signal processing.

Albert Benveniste (M'81–SM'89–F'91) was born May 8, 1949 in Paris, France. He graduated from Ecole des Mines de Paris in 1971 and received the Thèse d'Etat degree in mathematics, probability theory, in 1975.

From 1976 to 1979 he was Associate Professor of Mathematics at Université de Rennes I. Since 1979 he has been Directeur de Recherche at INRIA. For his thesis, he worked on probability theory, stochastic processes, and ergodic theory. In parallel, he pursued work on automatic control and signal processing in the area of adaptive systems for time-varying systems, with applications to data communications. In 1980 he began joint work with M. Basseville on the subject of change detection in signals and dynamical systems with application to vibration mechanics and fault detection in process control. Since 1987, he has been involved in developing, jointly with A. S. Willsky, a statistical theory of multiresolution signal and image processing, based on dynamical systems and Gaussian random fields on homogeneous trees. Since 1981, he has been interested in computer science, in the area of real-time languages and systems. Cooperating with P. Le Guernic, he participated in the definition of and theoretic studies on the Signal language.

Dr. Benveniste was corecipient in 1980 of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL Best Transaction Paper Award for his paper on blind deconvolution in data communications. In 1990 he received the CNRS Silver Medal. From 1986 to 1990 he was Vice-Chairman of the IFAC Committee on Theory and is Chairman of this committee for 1991–1993. From 1987 to 1990 he was Associate Editor for IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He is currently Associate Editor for the *International Journal of Adaptive Control and Signal Processing*, the *International Journal of Discrete Event Dynamical Systems*, and Associate Editor at Large for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He has coauthored with M. Métivier and P. Priouret the book *Adaptive Algorithms and Stochastic Approximations*, and has been an Editor, with M. Basseville, of the collective monograph *Detection of Abrupt Changes in Signals and Systems*.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received the S.B. and Ph.D. degrees from the Massachusetts Institute of Technology in 1969 and 1973, respectively.

From 1969 through 1973 he held a Fannie and John Hertz Foundation Fellowship. He joined the M.I.T. faculty in 1973 and his present position is Professor of Electrical Engineering. From 1974 to 1981 he served as Assistant Director of the M.I.T. Laboratory for Information and Decision Systems. He is also a founder and Member of the Board of Directors of Alphatech, Inc. He has held visiting positions at Imperial College, London; L'Université de Paris-Sud; and the Institut de Recherche en Informatique et Systèmes Aléatoires in Rennes, France. He is Editor of the M.I.T. Press series on signal processing, optimization, and control. His present research interests are in problems involving multidimensional and multiresolution estimation and imaging, discrete-event systems, and the asymptotic analysis of control and estimation systems. He is the author of the research monograph *Digital Signal Processing and Control and Estimation Theory* and is coauthor of the undergraduate text *Signals and Systems*.

Dr. Willsky was Program Chairman for the 17th IEEE Conference on Decision and Control, has been an Associate Editor of several journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, has served as a Member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society, and was Program Chairman for the 1981 Bilateral Seminar on Control Systems held in the People's Republic of China. He is also the Special Guest Editor of the 1992 special issue of the IEEE TRANSACTIONS ON INFORMATION THEORY on wavelet transforms and multiresolution signal analysis. In 1988 he was made a Distinguished Member of the IEEE Control Systems Society. In addition, he has given several planary lectures at major scientific meetings, including the 20th IEEE Conference on Decision and Control and the 1991 IEEE International Conference on Systems Engineering. In 1975 he received the Donald P. Eckman Award from the American Automatic Control Council. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph.

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 29, 2009 at 11:41 from IEEE Xplore. Restrictions apply.