Multi-scale autoregressive processes

Michèle Basseville, Albert Benveniste, Alan S. Willsky

136 p

UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1206

Programme 5
Automatique, Productique,
Traitement du Signal et des Données

MULTI-SCALE AUTOREGRESSIVE PROCESSES

Michèle BASSEVILLE
Albert BENVENISTE
Alan S. WILLSKY

Avril 1990

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CÉDEX
FRANCE
Téléphone : 99 36 20 00
Télex : UNIRISA 950 473 F
Télécopie : 99 38 38 32

Multi-Scale Autoregressive Processes

Publication Interne n° 525 - Mars 1990 - 136 Pages

Michèle Basseville * (IRISA/CNRS), Albert Benveniste * (IRISA/INRIA), Alan S. Willsky (MIT-LIDS)

ABSTRACT : In many applications, it is of interest to analyze and recognize phenomena occuring at different scales. The recently introduced wavelet transforms provide a time-and-scale decomposition of signals that offers the possibility of such analysis. However, there is no corresponding statistical framework. In this report, we describe such a framework. The theory of multiscale signal representations leads naturally to models of signals on trees. This report is made of **two parts**. *In the first part*, we describe the class of isotropic processes on homogenous trees and develop a theory of autoregressive models in this context. This leads to generalizations of Schur and Levinson recursions. *In the second part*, we investigate the properties of whitening and modeling filters for these processes. This report is a completely rewritten and extended version of the results announced in the research reports IRISA no 445/INRIA no 970 and IRISA no 446/INRIA no 971.

Processus Autorégressifs Multirésolution

RÉSUMÉ : Dans de nombreuses applications, il est intéressant d'analyser et de reconnaître des phénomènes se produisant à différentes échelles. Les transformées en ondelettes récemment introduites fournissent une décomposition de signaux en temps et en échelle. Mais à l'heure actuelle il n'existe pas de contexte statistique correspondant. Dans ce rapport, nous décrivons un tel contexte. La théorie des représentations multi-échelle conduit naturellement à des modèles de signaux indexés par des arbres. Ce rapport est constitué de **deux parties**. Dans la *première partie*, nous décrivons la classe des processus isotropes sur les arbres homogènes et développons une théorie des modèles autorégressifs dans ce contexte. Ceci conduit à des généralisations des récurrences de Schur et Levinson. Dans la *deuxième partie*, nous étudions les propriétés des filtres blanchisseurs et modélisateurs pour ces processsus. Ce rapport est une version complètement remaniée et étendue des résultats annoncés dans les rapports IRISA no 445/INRIA no 970 et IRISA no 446/INRIA no 971.

* Le travail de ces auteurs est également soutenu par le GDR CNRS G0134 "Traitement du Signal et Images".

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(L. A. 227)
UNIVERSITÉ DE RENNES 1 I. N. S. A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(LABORATOIRE DE RENNES)

Part I

Multi-Scale Autoregressive Processes Part I: Schur-Levinson parametrizations

Michèle Basseville[1], Albert Benveniste[1]
IRISA, Campus de Beaulieu
35042 RENNES CEDEX, FRANCE

Alan S. Willsky[2]
Laboratory for Information and Decision Systems, and
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, USA

Abstract

In[3] many applications (e.g. recognition of geophysical and biomedical signals and multiscale analysis of images), it is of interest to analyze and recognize phenomena occuring 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 homogenous 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.

EDICS: 5.2, 5.2.1, 6.1.4.

Keywords: Multiresolution signal analysis, statistical models, Schur-Levinson parametrizations.

---

[1]M.B. is also with the Centre National de la Recherche Scientifique (CNRS) and A.B. is also with Institut National de Recherche en Informatique et en Automatique (INRIA). The research of these authors was also supported in part by Grant CNRS GO134

[2]The work of this author was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-88-0032, in part by the National Science Foundation under Grant ECS-8700903, and in part by the US Army Research Office under Contract DAAL03-86-K-0171. In addition some of this research was performed while A.S.W. was a visitor at and received partial support from INRIA.

[3]Authorization is given to publish this abstract separately

1 Introduction

The investigation of multi-scale 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, 30] 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 [9, 24, 31, 33] or for performing Monte Carlo experiments [14] 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. Finally, renormalization group ideas, from statistical physics, now find application in methods for improving convergence in large-scale simulated annealing algorithms for Markov random field estimation [19].

One of the more recent areas of investigation in multi-scale analysis has been the development of a theory of multi-scale representations of signals [27, 29] and the closely related topic of wavelet transforms [15, 21, 26, 16, 23, 17, 22]. These methods have drawn considerable attention in several disciplines including signal processing because they appear to be a natural way to perform a time-scale decomposition of signals and because 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 or for the detection and localization of transient signals of different duration—requires, of course, the development of a corresponding theory of stochastic processes and their estimation. The research presented in this and several other papers and reports [13, 5, 14] has the development of this theory as its objective.

In the next section we introduce multi-scale representations of signals and wavelet

transforms and from these we motivate the investigation of stochastic processes on dyadic trees. In that section we also introduce the class of isotropic processes on dyadic trees and set the stage for introducing dynamic models on trees by describing their structure and introducing a rudimentary transform theory. In Section 2 we also introduce 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 3 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 4, 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 4 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 3 and 4 to construct and analyze whitening and modeling lattice filters for AR processes. In particular, we will see that the constraints on the reflection coefficients 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.

2 Multiscale Representations and Stochastic Processes on Homogenous Trees

2.1 Multiscale Representations and Wavelet Transforms

The multi-scale representation [28, 29] of a continuous 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

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

As $m \to \infty$ the approximation consists of a sum of many highly compressed, weighted, and shifted versions of the function $\phi(x)$ whose choice is far from arbitrary. In particular in order for the $(m+1)$st approximation to be a refinement of the mth, we require that $\phi(x)$ be exactly representable at the next scale:

$$\phi(x) = \sum_n h(n)\phi(2x - n) \tag{2.2}$$

Furthermore in order for (2.1) to be an orthogonal series, $\phi(t)$ and its integer translates must form an orthogonal set. As shown in [16], $h(n)$ must satisfy several conditions for this and several other properties of the representation to hold. In particular $h(n)$ must be the impulse response of a quadrature mirror filter [16, 35]. The simplest example of such a ϕ, h pair is the Haar approximation with

$$\phi(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & otherwise \end{cases} \tag{2.3}$$

and

$$h(n) = \begin{cases} 1 & n = 0 \\ 0 & otherwise \end{cases} \tag{2.4}$$

Multiscale representations are closely related to wavelet transforms. Such a transform is based on a single function $\psi(x)$ that has the property that the full set

of its scaled translates $\left\{2^{m/2}\psi(2^m x - n)\right\}$ form a complete orthonormal basis for L^2. In [16] it is shown that ϕ and ψ are related via an equation of the form

$$\psi(x) = \sum_n g(n)\phi(2x - n) \tag{2.5}$$

where $g(n)$ and $h(n)$ form a *conjugate mirror filter* pair [35], and that

$$f_{m+1}(x) = f_m(x) + \sum_n d(m,n)\psi(2^m x - n) \tag{2.6}$$

$f_m(x)$ is simply the partial orthonormal expansion of $f(x)$, up to scale m, with respect to the basis defined by ψ. For example if ϕ and h are as in eq. (2.3), eq. (2.4), then

$$\psi(x) = \begin{cases} 1 & 0 \leq x < 1/2 \\ -1 & 1/2 \leq x < 1 \\ 0 & otherwise \end{cases} \tag{2.7}$$

$$g(n) = \begin{cases} 1 & n = 0 \\ -1 & n = 1 \\ 0 & otherwise \end{cases} \tag{2.8}$$

and $\left\{2^{m/2}\psi(2^m x - n)\right\}$ is the *Haar basis.*

From the preceding remarks we see that we have a *dynamical* relationship between the coefficients $f(m,n)$ at one scale and those at the next. Indeed this relationship defines a lattice on the points (m,n), where $(m+1,k)$ is connected to (m,n) if $f(m,n)$ influences $f(m+1,k)$. In particular the Haar representation naturally defines a dyadic tree structure on the points (m,n) in which each point has two descendents corresponding to the two subdivisions of the support interval of $\phi(2^m x - n)$, namely those of $\phi(2^{(m+1)}x - 2n)$ and $\phi(2^{(m+1)}x - 2n - 1)$. This observation provides the motivation for the development of models for stochastic processes on dyadic trees as the basis for a statistical theory of multiresolution stochastic processes.

2.2 Homogenous Trees

Homogenous trees, and their structure, have been the subject of some work [2, 3, 12, 18, 11] in the past on which we build and which we now briefly review. A *homogenous tree $\mathcal{T}$ of order q* is an infinite acyclic, undirected, connected graph such that every node of $\mathcal{T}$ 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 Figure 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 $\mathcal{T}$ has a natural notion of distance: $d(s,t)$ is the number of branches along the shortest path between the nodes of $s, t \in \mathcal{T}$ (by abuse of notation we use $\mathcal{T}$ to denote both the tree and its collection of nodes). One can then define the notion of an isometry on $\mathcal{T}$ which is simply a one-to-one map of $\mathcal{T}$ onto itself that preserves distances. For the case of $q = 1$, the group of all possible isometries corresponds to translations of the integers ($t \mapsto t + k$) the reflection operation ($t \mapsto -t$) and concatenations of these. For $q \geq 2$ the group of isometries of $\mathcal{T}$ is significantly larger and more complex. One extremely important result is the following [18]:

Lemma 2.1 (Extension of Isometries) *Let $\mathcal{T}$ 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*

$$d(f(s), f(t)) = d(s,t) \text{ for all } s, t \in A \tag{2.9}$$

Then there exists an isometry $\tilde{f}$ of $\mathcal{T}$ 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 segments joining any two points of A are identical.

Another important concept is the notion of a *boundary point* [3, 11] of a tree. Consider the set of infinite sequences of $\mathcal{T}$ where any such sequence consists of a sequence of distinct nodes $t_1, t_2, \ldots$ 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 $\mathcal{T}$. 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 Figure 1). We then can define a notion of the *relative distance* of two nodes to $-\infty$

$$\delta(s,t) = d(s, s \wedge t) - d(t, s \wedge t) \tag{2.10}$$

so that

$$s \preceq t \text{ ("s is at least as close to } -\infty \text{ as t") if } \delta(s,t) \leq 0 \tag{2.11}$$

$$s \prec t \text{ ("s is closer to } -\infty \text{ than t") if } \delta(s,t) < 0 \tag{2.12}$$

This also yields an equivalence relation on nodes of $\mathcal{T}$:

$$s \asymp t \leftrightarrow \delta(s,t) = 0 \tag{2.13}$$

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

These equivalence classes can best be visualized as in Figure 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 $s \preceq t$ 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.

2.3 Shifts and Transforms on $\mathcal{T}$

The structure of Figure 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 Figure 3

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

Note that 0 and δ are isometries; 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 left-most operator is applied first):

$$\alpha\gamma^{-1} = \beta\gamma^{-1} = 0 \tag{2.14}$$

$$\delta\gamma^{-1} = \gamma^{-1} \tag{2.15}$$

$$\delta^2 = 0 \tag{2.16}$$

$$\beta\delta = \alpha \tag{2.17}$$

Arbitrary moves on the tree can then be encoded via finite strings or *words* using these symbols as the alphabet and the formulas (2.14)–(2.17). For example, referring to Figure 3

$$s_1 = t\gamma^{-4} \quad , \quad s_2 = t\gamma^{-3}\delta \quad , \quad s_3 = t\gamma^{-3}\delta\alpha$$
$$s_4 = t\gamma^{-3}\delta\beta\alpha \quad , \quad s_5 = t\gamma^{-3}\delta\alpha\beta^2 \tag{2.18}$$

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

$$\mathcal{L} = (\gamma^{-1})^* \cup (\gamma^{-1})^*\delta\{\alpha,\beta\}^* \cup \{\alpha,\beta\}^* \tag{2.19}$$

where K^* denotes arbitrary sequences of symbols in K including the empty sequence which we identify with the operator 0. Then any point $t \in \mathcal{T}$ can be written as $t_0 w$, where $w \in \mathcal{L}$. Note that the moves in $\mathcal{L}$ 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 $\mathcal{L}$ is not a semigroup since vw need not be in $\mathcal{L}$ for $v, w \in \mathcal{L}$. However, for future reference we note that, using (2.14)–(2.17), $w\delta$ and $w\gamma^{-1}$ are both in $\mathcal{L}$ for any $w \in \mathcal{L}$.

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

$$|\gamma^{-1}| = |\alpha\ | = |\beta| = 1$$
$$|0| = 0 \quad , \quad |\delta| = 2 \tag{2.20}$$

Thus $|\gamma^{-n}| = n$, $|w_{\alpha\beta}|$ = the number of α's and β's in $w_{\alpha\beta} \in \{\alpha,\beta\}^*$, and $|\gamma^{-n}\delta w_{\alpha\beta}| = n + 2 + |w_{\alpha\beta}|$.[4] This notion of length will be useful in defining the *order* of dynamic

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

models on $\mathcal{T}$. 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 $w \asymp 0$ for such a move. The reasons for this will become clear when we examine autoregressive models.

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

$$\begin{aligned} &\text{If } t = t\gamma^{-1}\alpha, \text{ then } t\delta^{(n)} = t\gamma^{-1}\delta^{(n-1)}\alpha \\ &\text{If } t = t\gamma^{-1}\beta, \text{ then } t\delta^{(n)} = t\gamma^{-1}\delta^{(n-1)}\beta \end{aligned} \tag{2.21}$$

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 Figure 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 [8] of the form:

$$S = \sum_{w\in\mathcal{L}} s_w \cdot w \tag{2.22}$$

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

$$(Su)_t = \sum_{w\in\mathcal{L}} s_w u_{tw} \tag{2.23}$$

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:

$$\gamma[S] = \sum_{w \in \mathcal{L}} s_{w\gamma^{-1}} \cdot w \tag{2.24}$$

$$\delta^{(k)}[S] = \sum_{w \in \mathcal{L}} s_{w\delta^{(k)}} \cdot w \tag{2.25}$$

where we use (2.14)–(2.17) and (2.21) to write $w\gamma^{-1}$ and $w\delta^{(k)}$ as elements of $\mathcal{L}$. Notice that, because of the relations (2.14)–(2.17), the operators $S \longrightarrow \gamma[S]$ and $S \longrightarrow \delta[S]$ can *not* be thought of as multiplication operators on formal power series.

2.4 Isotropic Processes on Homogenous Trees

Consider a zero-mean stochastic process $Y_t, t \in \mathcal{T}$ 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 $r_n, n = 0, 1, 2, \ldots$ so that

$$E[Y_t Y_s] = r_{d(t,s)} \tag{2.26}$$

An alternate way to think of an isotropic process is that its statistics are invariant under tree isometries. That is, if $f : \mathcal{T} \to \mathcal{T}$ 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 stranger property.

Isotropic processes have been the subject of some study [2, 3, 18] 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 $r_n, n = 0, 1, \ldots$ 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

$$r_n = \int_{-1}^{1} C_n(x)\mu(dx) \tag{2.27}$$

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

$$\mu(dx) = \frac{1}{\pi}(1-x^2)^{-\frac{1}{2}}S(d\omega) \tag{2.28}$$

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

$$\mu(dx) = \frac{1}{\pi}(1-x^2)^{-\frac{1}{2}} \tag{2.29}$$

The analogous theorem for isotropic processes on dyadic trees requires the introduction of the Dunau polynomials [2, 18]:

$$P_0(x) = 1 \quad , \quad P_1(x) = x \tag{2.30}$$

$$xP_n(x) = \frac{2}{3}P_{n+1}(x) + \frac{1}{3}P_{n-1}(x) \tag{2.31}$$

Theorem 2.1 [2, 3]: *A sequence $r_n, n = 0, 1, 2, \ldots$ 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*

$$r_n = \int_{-1}^{1} P_n(x)\mu(dx) \tag{2.32}$$

The simplest isotropic process on the tree is again white noise, i.e. a collection of uncorrelated random variables indexed by $\mathcal{T}$, with $r_n = \delta_{n0}$, and the spectral measure μ in (2.32) in this case is [18]

$$\mu(dx) = \frac{1}{2\pi}\chi_{\left[-\frac{2\sqrt{2}}{3},\frac{2\sqrt{2}}{3}\right]}(x)\frac{(8-9x^2)^{\frac{1}{2}}}{1-x^2}dx \tag{2.33}$$

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 doesn't 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 semi-definite 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 $Y_{t_1}, Y_{t_2}, \ldots$ where the t_i are all distinct and $d(t_i, t_{i+1}) = 1$; the covariance fuction of this series is r_n. However, thanks to the geometry of the tree and the richness of the group of isometries of $\mathcal{T}$, there are many additional constraints on r_n. For example, consider the three nodes v, u, and $s \wedge t$ in Figure 1, and let

$$X^T = [Y_v, Y_u, Y_{s \wedge t}] \tag{2.34}$$

Then

$$E[XX^T] = \begin{bmatrix} r_0 & r_2 & r_2 \\ r_2 & r_0 & r_2 \\ r_2 & r_2 & r_0 \end{bmatrix} \geq 0 \tag{2.35}$$

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.

2.5 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

$$Y_t = \sum_{s \in \mathcal{T}} c_{t,s} W_s \tag{2.36}$$

where W_t is a white noise process with unit variance. 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

$$Y_t = \sum_{s \in T} c_{d(s,t)} W_s \tag{2.37}$$

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

$$\begin{aligned} E[Y_{s'}Y_{t'}] &= \sum_u c_{d(s',u)} c_{d(t',u)} \\ &= \sum_{u'} c_{d(s',f(u'))} c_{d(t',f(u'))} \\ &= \sum_{u'} c_{d(f(s),f(u'))} c_{d(f(t),f(u'))} \\ &= \sum_{u'} c_{d(s,u')} c_{d(t,u')} = E[Y_s Y_t] \end{aligned} \tag{2.38}$$

The class of systems of the form of (2.37) are 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. 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 2.3

$$Y_t = (SW)_t \quad , \quad S = \textstyle\sum_{w \in \mathcal{L}} s_w \cdot w \tag{2.39}$$

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 straighforward 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

$$Y_t = \sum_{\substack{w \prec 0 \\ |w| \leq p}} a_w Y_{tw} + \sigma W_t \tag{2.40}$$

where W_t is a white noise with unit variance.

The form of (2.40) deserves some comment. First note that the constraints placed on w in the summation of (2.40) 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 ($w \preceq 0$). Thus the model (2.40) is not strictly "causal" and is indeed an implicit specification since values of Y on the same horocycle depend on each other through (2.40) (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. 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:

$$Y_t = \sum_{n=0}^{\infty} a_n W_{t\gamma^{-n}} \tag{2.41}$$

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

$$a_n = \sigma a^n \tag{2.42}$$

so that the *only* process with strict past dependence as in (2.41) is the AR(1) process

$$Y_t = aY_{t\gamma^{-1}} + \sigma W_t \tag{2.43}$$

Consider next the AR(2) process, which specializing (2.40), has the form

$$Y_t = a_1 Y_{t\gamma^{-1}} + a_2 Y_{t\gamma^{-2}} + a_3 Y_{t\delta} + \sigma W_t \tag{2.44}$$

Note first that this is indeed an implicit specification, since if we evaluate (2.44) at $t\delta$ rather than t we see that

$$Y_{t\delta} = a_1 Y_{t\gamma^{-1}} + a_2 Y_{t\gamma^{-2}} + a_3 Y_t + \sigma W_{t\delta} \tag{2.45}$$

We can, of course, solve the pair (2.44), (2.45) to obtain the explicit formulae

$$Y_t = \left(\frac{a_1}{1-a_3^2}\right) Y_{t\gamma^{-1}} + \left(\frac{a_2}{1-a_3^2}\right) Y_{t\gamma^{-2}} + \sigma V_t \tag{2.46}$$

$$Y_{t\delta} = \left(\frac{a_1}{1-a_3^2}\right) Y_{t\gamma^{-1}} + \left(\frac{a_2}{1-a_3^2}\right) Y_{t\gamma^{-2}} + \sigma V_{t\delta} \tag{2.47}$$

where

$$V_t = \frac{1}{1-a_3^2}\{W_t + a_3 W_{t\delta}\} \tag{2.48}$$

The structure of the representation (2.46)–(2.47) 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 [13] 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 our AR(p) model has $(2^p - 1)$ a's and one σ in contrast to the p a's and one σ for time series. On the other hand, the coefficients in our AR model are *not* independent and indeed there exist nonlinear relationships among the coefficients. For example for the second-order model (2.44) $a_3 \neq 0$ if $a_2 \neq 0$ since we know that the only isotropic process with strict past dependence is AR(1). In Appendix B we show that the coefficients a_1, a_2, and a_3 in (2.44) are related by a 4th-order polynomial relation.

Because of the complex relationship among the a_w's in (2.40), 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.

3 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 [6, 7] 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 the output of

these prediction error filters *increases with increasing filter order.* This is a direct consequence of the structure of the AR model (2.40) 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 (2.44)-(2.48) 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.

3.1 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 [7], 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 : $\mathcal{Y}_{t,n} = \mathcal{H}\{Y_{t-k} | 0 \leq k \leq n\}$ 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:

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

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 ($s \asymp t$) as long as $d(s,t) \leq n$. 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 Figures 4 – 7).

In analogy with the time series case, the backward innovations or prediction errors, which we denote by $\mathcal{F}_{t,n}$, are defined as the variables spanning the new

information in $\mathcal{Y}_{t,n}$ which are 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 $|w| = n$, define

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

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

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

For time series the forward innovations 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

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

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 $w \succ 0$. Define

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

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

$$\dim E_{t,n} = 2^{\left[\frac{n-1}{2}\right]} \tag{3.7}$$

where $[x]$ denotes the largest integer $\leq x$. The elements of $E_{t,n}$ are ordered according to a dyadic representation of the words w for which $|w| < n$, $w \succ 0$. Specifically any such w other than 0 must have the form

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

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 $w = 0, \delta, \delta^{(2)}$, and $\delta^{(2)}\delta$ are ordered in the Figure 7[5] according to the left-to-right arrow. Thus the words w of interest are in one-to-one correspondence with the numbers 0 and $\sum_{j=1}^{k} 2^{i_j}$, 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

$$\dim F_{t,n} = 2^{[\frac{n}{2}]} \tag{3.10}$$

The elements of $F_{t,n}$ are ordered as follows. Note that any word w for which $|w| = n$ and $w \preceq 0$ can be written as $w = \gamma^{-k}\tilde{w}$ for some $k \geq 0$ and $\tilde{w} \succ 0$. For example, as illustrated in Figure 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 $\tilde{w}$.

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 Figure 4 and the last three are in Figures 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 Figure 4): To begin we have

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

The only word w for which $|w| = 1$ and $w \preceq 0$ 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

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

[5]In the figure these points appear to be ordered left-to right along the horocycle. This, however, is due only to the fact that t was taken at the left of the horocycle.

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

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

$n = 2$ (See Figure 4): Here

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

In this case $|w| = 2$ and $w \preceq 0$ 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,

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

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

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

$n = 3$ (See Figure 5) In this case

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

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

Also

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

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

$$E_{t,3} = \begin{pmatrix} E_{t,3} & (0) \\ E_{t,3} & (\delta) \end{pmatrix}$$

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

$n = 4$ (See Figure 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 Figure 7)

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

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

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

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

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 $|w| = n$, $w \asymp 0$)—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 are coupled.

As a second point, note that we have already provided a simple interpretation (3.2) of $\mathcal{F}_{t,n}$ 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

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

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

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

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 T but not all of them.

3.2 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.

n even

Consider first $F_{t,n}(w)$ for $|w| = n$, $w \preceq 0$. There are two natural subclasses for these words w. In particular either $w \prec 0$ of $w \succ 0$.

Case 1: Suppose that $w \prec 0$. Then $w = \gamma^{-1}\tilde{w}$ for some $\tilde{w} \preceq 0$ with $|\tilde{w}| = n-1$. We then can perform the following computation, using (3.3) and properties of orthogonal projections:

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

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

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

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

$$F_{t,n}(w) = Y_{t\delta^{(\frac{n}{2})}\tilde{w}} - E\left(Y_{t\delta^{(\frac{n}{2})}\tilde{w}}|\mathcal{Y}_{t\gamma^{-1},n-2}\right) - E\left(Y_{t\delta^{(\frac{n}{2})}\tilde{w}}|\mathcal{Y}_{t,n-1} \ominus \mathcal{Y}_{t\gamma^{-1},n-2}\right) \tag{3.14}$$

Now for n even we can show that

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

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

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

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 $w \asymp 0$. In this case we compute

$$\begin{aligned} E_{t,n}(w) &= Y_{tw} - E\left(Y_{tw}|\mathcal{Y}_{t\gamma^{-1},n-2}\right) - E\left(Y_{tw}|\mathcal{Y}_{t\gamma^{-1},n-1} \ominus \mathcal{Y}_{t\gamma^{-1},n-2}\right) \\ &= E_{t,n-1}(w) - E\left(E_{t,n-1}(w)|F_{t\gamma^{-1},n-1}\right) \end{aligned} \tag{3.16}$$

where the last equality follows from (3.2).

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

$$\begin{aligned} F_{t,1} &= Y_{t\gamma^{-1}} - E\left(Y_{t\gamma^{-1}}|Y_t\right) \\ &= Y_{t\gamma^{-1}} - k_1 Y_t = F_{t\gamma^{-1},0} - k_1 E_{t,0} \end{aligned} \tag{3.17}$$

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

$$k_1 = \frac{E[Y_{t\gamma^{-1}}Y_t]}{E\left[Y^2_{t\gamma^{-1}}\right]} = \frac{r_1}{r_0} \tag{3.18}$$

Similarly

$$\begin{aligned} E_{t,1} &= Y_t - E(Y_t|Y_{t\gamma^{-1}}) \\ &= Y_t - k_1 Y_{t\gamma^{-1}} = E_{t,0} - k_1 F_{t\gamma^{-1},0} \end{aligned} \tag{3.19}$$

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* $w \succ 0$. Therefore the condition

$$|w| = n \text{ and } w \preceq 0$$

is equivalent to

$$w = \gamma^{-1}\tilde{w} \quad , \quad |\tilde{w}| = n-1 \quad , \quad \tilde{w} \preceq 0$$

Therefore, proceeding as before,

$$\begin{aligned} F_{t,n}(w) &= Y_{t\gamma^{-1}\tilde{w}} - E(Y_{t\gamma^{-1}\tilde{w}}|\mathcal{Y}_{t\gamma^{-1},n-2}) - E(Y_{t\gamma^{-1}\tilde{w}}|\mathcal{Y}_{t\gamma^{-1},n-1} \ominus \mathcal{Y}_{t\gamma^{-1},n-2}) \\ &= F_{t\gamma^{-1},n-1}(\tilde{w}) - E\left(F_{t\gamma^{-1},n-1}(\tilde{w})|E_{t,n-1}, E_{t\delta^{(\frac{n-1}{2})},n-1}\right) \end{aligned} \tag{3.20}$$

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 $|w| < n$, $w \succ 0$. 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

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

Case 2: Suppose that $|w| = n-1$. In this case $w = \delta^{(\frac{n-1}{2})}\tilde{w}$ where $\tilde{w} \succ 0$ and computations analogous to those performed previously yield

$$E_{t,n}(w) = E_{t\delta^{(\frac{n-1}{2})},n-1}(\tilde{w}) - E\left(E_{t\delta^{(\frac{n-1}{2})},n-1}(\tilde{w})|F_{t\gamma^{-1},n-1}\right) \tag{3.22}$$

where in this case we use the fact that

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

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 subsection 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

$$E[E_{t,1}E_{t\gamma^{-2},1}] = r_2 - \frac{r_1^2}{r_0} + \frac{r_1(r_1r_2 - r_0r_3)}{r_0^2}$$

In general these expressions are *not* equal so 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).

4 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), (3.21), and (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 result 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.

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

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

$$e_{t,n} = 2^{-[\frac{n-1}{2}]} \sum_{|w|<n, w \succ 0} E_{t,n}(w) \tag{4.1}$$

$$f_{t,n} = 2^{-[\frac{n}{2}]} \sum_{|w|=n, w \preceq 0} F_{t,n}(w) \tag{4.2}$$

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,*

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

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

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

$$= E\left(E_{t\delta^{(\frac{n}{2})},n-1}(0)|e_{t,n-1}\right) \quad (4.5)$$

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

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

(refer to (3.16)).

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

$$E\left(E_{t,n-1}(w')|F_{t\gamma^{-1},n-1}\right) = E\left(E_{t\delta^{(\frac{n-1}{2})},n-1}(w'')|F_{t\gamma^{-1},n-1}\right) \quad (4.7)$$

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

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

$$E\left(F_{t\gamma^{-1},n-1}(\tilde{w})|E_{t,n-1}, E_{t\delta^{(\frac{n-1}{2})},n-1}\right) = E\left(F_{t\gamma^{-1},n-1}(w_0)|\frac{1}{2}\left(e_{t,n-1} + e_{t\delta^{(\frac{n-1}{2})},n-1}\right)\right) \quad (4.9)$$

(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 Figures 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 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 Figure 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 $\left(e_{t,n-1} + e_{t\delta^{(\frac{n-1}{2})},n-1}\right)/2$.

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 C by exploiting symmetry and the isometry extension lemma.

Lemma 4.2 *The best linear estimate*

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

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

Lemma 4.3 *The cross-covariance*

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

is the same for all $|w| = n-1, w \preceq 0$ and all $|w'| < n$ and $w' \succ 0$.

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

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

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

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 $\alpha_0, \alpha_1, \ldots, \alpha_{[\frac{n-1}{2}]}$ so that

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

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 $1_{[\frac{n-1}{2}]}$ is an eigenvector of $\Sigma_{E,n}$. Then, consider any $X \in \mathcal{E}_{t,n-1}$ of the form

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

where

$$\sum_{\substack{|w'|<n \\ w' \asymp 0}} \lambda_{w'} = 0 \tag{4.16}$$

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

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

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)

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

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*. 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 4.1 and in Appendix D, 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 $\Sigma(\alpha_0, \alpha_1, \ldots, \alpha_d)$, with a total complexity of $O(l \log l)$, where $l = \left[\frac{n-1}{2}\right]$.

Finally, let us note an extremely important consequence of Lemma 4.1. Recall that the Levinson recursions developed in the Section 3 involved projections of each 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 $|\omega'| = n - 1$,

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

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

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

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.

4.2 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 3.2 is the following set of recursions for the barycenters themselves:

For n even:

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

$$f_{t,n} = \frac{1}{2}\left(f_{t\gamma^{-1},n-1} + e_{t\delta^{(\frac{n}{2})},n-1}\right) - \frac{1}{2}E\left(f_{t\gamma^{-1},n-1} + e_{t\delta^{(\frac{n}{2})},n-1}|e_{t,n-1}\right) \tag{4.21}$$

For n odd, $n > 1$:

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

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

while for $n = 1$,

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

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 of the sequence k_n which we refer to as the *reflection coefficient* sequence.

Theorem 4.1 *For n even:*

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

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

where

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

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

For n odd:

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

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

where

$$k_n = cor\left(\frac{1}{2}\left(e_{t,n-1} + e_{t\delta(\frac{n-1}{2}),n-1}\right), f_{t\gamma^{-1},n-1}\right) \tag{4.30}$$

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

Lemma 4.5 *For n odd:*

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

Lemma 4.6 *For n even* $\frac{1}{2}\left(e_{t,n} + e_{t\delta(\frac{n}{2}),n}\right)$ *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

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

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

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

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 $t\delta^{(\frac{n}{2})}$ and the fact that

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

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*

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

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

where k_n must satisfy

$$-\frac{1}{2} \leq k_n \leq 1 \tag{4.38}$$

For n odd

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

where

$$-1 \leq k_n \leq 1 \tag{4.40}$$

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.

4.3 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 from 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 [25] [34] 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 2. Specifically for $n \geq 0$ define P_n and Q_n as:

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

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

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

$$P_0 = Q_0 = \sum_{w \preceq 0} r_{|w|} \cdot w \tag{4.43}$$

Recalling the definitions (2.24), (2.25) 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

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

$$Q_n = \frac{1}{2}\left(\gamma[Q_{n-1}] + \delta^{(\frac{n}{2})}[P_{n-1}]\right) - k_n P_{n-1} \tag{4.45}$$

where

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

For n odd

$$P_n = \frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})}[P_{n-1}]\right) - k_n \gamma[Q_{n-1}] \tag{4.47}$$

$$Q_n = \gamma[Q_{n-1}] - k_n \frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})}[P_{n-1}]\right) \tag{4.48}$$

where

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

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).

5 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 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.

ACKNOWLEDGEMENT: the authors would like to express their thanks to Jean-Pierre Conze, Yves Guivarc'h, and Albert Raugi from the Institut de Recherche en Mathématique de Rennes, for bringing to their knowledge the work of Arnaud, Cartier, Dunau, and LeTac on isotropic processes on the homogeneous tree, and also to Ken C. Chou for fruitful discussions.

References

[1] M. BASSEVILLE, A. BENVENISTE, AND A.S. WILLSKY, "Multiscale Autoregressive Processes, Part II: Lattice Structures for Whitening and Modeling", IRISA and LIDS (MIT) report, submitted for publication.

[2] J.P. ARNAUD, "Fonctions sphériques et fonctions définies positives sur l'arbre homogène," *C.R.A.S.* t. 290, série A (14 Jan. 1980), p. 99–101.

[3] J.P. ARNAUD, G. LETAC, "La formule de représentation spectrale d'un processus gaussien stationnaire sur un arbre homogène," *Publi. Labo. Stat. and Proba.*, UA 745, Toulouse.

[4] M. BARNSLEY, *Fractals Everywhere*, Academic Press, San Diego, 1988.

[5] M. BASSEVILLE AND A. BENVENISTE, "Multiscale statistical signal processing," proc. of the ICASSP-89, 2065-2068, Glasgow, 1989.

[6] 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, p. 231–266, Editions du CNRS, 1982.

[7] A. BENVENISTE, "Méthodes d'orthogonalisation en treillis pour le problème de la réalisation stochastique," ibidem, vol. 2, p. 267–308.

[8] J. BERSTEL, C. REUTENAUER, "Les séries rationnelles et leurs langages," Masson 1984, Collection "Etudes et Recherches en Informatique".

[9] A. BRANDT, "Multi-level adaptive solutions to boundary value problems," *Math. Comp.*, Vol. 13, pp. 333–390, 1977.

[10] W. BRIGGS, *A Multigrid Tutorial*, SIAM, Philadelphia, 1987.

[11] P. CARTIER, "Harmonic analysis on trees," *Proc. Sympos. Pure Math.*, Vol. 26, Amer. Math. Soc., Providence, pp. 419–424, 1974.

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

[13] K.C. CHOU, A.S. WILLSKY, A. BENVENISTE, AND M. BASSEVILLE, "Recursive and Iterative Estimation Algorithms for Multi-Resolution Stochastic Processes," Proc. of the IEEE Conf. on Decision and Control, Tampa, Florida, Dec. 1989.

[14] K. CHOU, *A Stochastic Modeling Approach to Multiscale Signal Processing*, MIT, Dept. of Electrical Engineering and Computer Science, Ph.D. Thesis, 1990 (in preparation).

[15] I. DAUBECHIES, A. GROSSMANN, Y. MEYER, "Painless non-orthogonal expansions," *J. Math. Phys.*, vol. 27, p. 1271–1283, 1986.

[16] I. DAUBECHIES, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Math.*, vol. 91, pp. 909–996, 1988.

[17] I. DAUBECHIES, "The wavelet transform, time-frequency localization and signal analysis," AT&T Bell Laboratories Report, to appear in *IEEE Trans. on IT.*

[18] 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, no. 02-83, June 1983.

[19] B. GIDAS, "A renormalization group approach to image processing problems," *IEEE Trans. on Pattern Anal. and Mach. Int.*, vol. 11, pp. 164–180, 1989.

[20] J. GOODMAN AND A. SOKAL, "Multi-grid Monte Carlo I. conceptual foundations," Preprint, Dept. Physics New York University, New York, Nov. 1988; to be published.

[21] P. GOUPILLAUD, A. GROSSMANN, J. MORLET, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, p. 85–102, 1984/85, Elsevier.

[22] 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.

[23] 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. of the Int. Conf. on Wavelets, Marseille, Dec 14-18, 1987, J.M. Combes and A. Grossmann Eds., Springer Verlag, 1989.

[24] W. HACKBUSCH AND U. TROTTENBERG, eds., *Multigrid Methods*, Springer-Verlag, New York, 1982.

[25] 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, Birkhäuser (Basel, Boston, Stuttgart), 1986.

[26] R. KRONLAND-MARTINET, J. MORLET, A. GROSSMANN, "Analysis of sound patterns through wavelet transforms," *Intl. J. of Pattern Recognition and Artificial Intelligence*, vol. 1 No 2, 273–301, Aug. 1987.

[27] S.G. MALLAT, "A compact multiresolution representation: the wavelet model," Proc. of the IEEE Workshop on Computer Vision, Miami, Florida, Dec. 1987.

[28] S.G. MALLAT, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE PAMI-11 No 7*, 674–693, 1989.

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

[30] B. MANDELBROT, *The Fractal Geometry of Nature*, Freeman, New York, 1982.

[31] S. McCormick, *Multigrid Methods*, Vol. 3 of the SIAM Frontiers Series, SIAM, Philadelphia, 1987.

[32] Y. Meyer, *Wavelets and operators,* Proceedings of the Special year in modern Analysis, Urbana 1986/87, published by Cambridge University Press, 1989.

[33] D. Paddon and H. Holstein, eds., *Multigrid Methods for Integral and Differential Equations*, Clarendon Press, Oxford, England, 1985.

[34] E.A. Robinson, S. Treitel, "Maximum Entropy and the Relationship of the Partial Autocorrelation to the Reflection Coefficients of a Layered System," *IEEE Trans. on ASSP*, vol. 28 Nr 2, 224–235, 1980.

[35] M.J. Smith and T.P. Barnwell, "Exact reconstruction techniques for tree-structured subband coders," *IEEE Trans. on ASSP*, vol. 34, pp. 434–441, 1986.

Appendices

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 in 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

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

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

from which we deduce that

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

or equivalently

$$\tfrac{a_n}{a_{n-1}} = \text{constant} \quad , \quad n \geq 1$$

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

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

B The Relation Among the Parameters of AR(2)

Consider the second-order model (2.44) where W_t is unit variance white noise. We would like to show that the coefficients a_1, a_2, and a_3 are related by a fourth-order poylnomial relation that must be satisfied if Y_t is isotropic. To begin note that from (2.44) we obtain the relation

$$E(Y_t W_t) = a_3 E(Y_{t\delta} W_t) + \sigma \tag{B.1}$$

while from (2.45) we find

$$E(Y_{t\delta} W_t) = a_3 E(Y_t W_t) \tag{B.2}$$

from which we deduce that $|a_3| \neq 1$ and

$$\begin{aligned} E(Y_t W_t) &= \sigma \frac{a_0}{a_0^2 - a_3^2} \\ E(Y_{t\delta} W_t) &= \sigma \frac{a_3}{a_0^2 - a_3^2} \end{aligned} \tag{B.3}$$

Next consider multiplying (2.44) by each of the following: Y_t, $Y_{t\delta}$, $Y_{t\gamma^{-1}}$, $Y_{t\gamma^{-2}}$. We take expectations using (B.1), (B.2) and the fact that $E(Y_{t\gamma^{-1}} W_t) = E(Y_{t\gamma^{-2}} W_t) = 0$ (since we are solving the AR equations "causally"—see (2.46), (2.47)). Assuming that Y is isotropic, we obtain the following *four* linear equations in the *three* unknowns r_0, r_1, r_2:

$$\begin{cases} r_0 &= a_1 r_1 + a_2 r_2 + a_3 r_2 + \sigma \frac{a_0}{a_0^2 - a_3^2} \\ r_1 &= a_1 r_0 + a_2 r_1 + a_3 r_1 \\ r_2 &= a_1 r_1 + a_2 r_0 + a_3 r_2 \\ r_2 &= a_1 r_1 + a_2 r_2 + a_3 r_0 + \sigma \frac{a_3}{a_0^2 - a_3^2} \end{cases} \tag{B.4}$$

For this system to have a solution the coefficients a_1, a_2, a_3 must satisfy

$$\begin{vmatrix} -1 & a_1 & a_2 + a_3 & -1 \\ a_1 & a_2 + a_3 - 1 & 0 & 0 \\ a_2 & a_1 & a_3 - 1 & 0 \\ a_3 & a_1 & a_2 - 1 & -a_3 \end{vmatrix} = 0 \tag{B.5}$$

which is a fourth-order polynomial relation. It is straightforward to check that these are the only constraints on the a_i in order for Y to be isotropic (multiply (2.44) by any Y_{tw}, $w \preceq 0$, $|w| > 2$ and take expectations—one obtains a unique expression for each r_n, $n \geq 3$ in terms of the preceding values of r).

C 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 4 without proof—are the constraints of isotropy and the construction of specific isometries.

C.1 Proof of Lemma 4.2

Let

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

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

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

Define the set of nodes

$$\mathcal{T}_{t,n} = \{s = tv; |v| \leq n, v \preceq 0\} \tag{C.3}$$

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

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

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

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 projection operators are identical. Furthermore, from (C.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 Figure 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 leave fixed all nodes above and including the pivot node. Thirdly, 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 Figure 10 map 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 $\mathcal{T}_{t,n}$ 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 $\mathcal{T}_{t,n}$.

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 Figure 11 for $n = 4$. In Figure 11(b) we have redrawn the tree of Figure 11(a) in a more symmetric fashion so that the nature of the flips performed by these isometries is more apparent.

C.2 Proof of Lemma 4.3

Set

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

where n is even $|w| = n-1$, $w \preceq 0$ and $|w'| < n$, $w' \succ 0$. 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 $|w| = n-1$, $w \preceq 0$, $\phi(tw_1) = tw_2$, $\phi(tw_2) = tw_1$, ϕ leaves $\mathcal{T}_{t,n-1}$ invariant *and* leaves fixed any point of the form tw', with $|w'| < n$, $w' \succ 0$.

2. For any w_1', w_2' satisfying $|w'| < n$, $w' \succ 0$, $\psi(t\gamma^{-1}w_1') = t\gamma^{-1}w_2', \psi(t\gamma^{-1}w_2') = t\gamma^{-1}w_1'$, ψ leaves $\mathcal{T}_{t\gamma^{-1},n-1}$ invariant *and* leaves fixed any point of the form $t\gamma^{-1}w$, with $|w| = n-1$, $w \preceq 0$.

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 Figure 12 for the case $n = 4$). Thus the existence of the required isometries is immediate.

C.3 Proof of Lemma 4.4

As in Section C.2, let w_m denote the $2^{[\frac{n-1}{2}]}$ words such that $|w| < n$, $w \succ 0$, and for any two such words let

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

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 $\mathcal{T}_{t,n}$ so that ϕ leaves $\mathcal{T}_{t\gamma^{-1},n-1}$ invariant and performs the interchanges

$$tw_i \leftrightarrow tw_i', \; tw_j \leftrightarrow tw_j'$$

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

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

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

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

This finishes the proof of Lemma 4.4.

C.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 $\mathcal{T}_{t\gamma^{-1},n-1}$ 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 $t\gamma^{-\frac{n+1}{2}}$. 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 $t\gamma^{-\frac{n+1}{2}}$ is along this third direction. Then the pivot about $t\gamma^{-\frac{n+1}{2}}$

with respect to $-\tilde{\infty}$ performs the required interchange and leaves the set of nodes $\mathcal{T}_{t\gamma^{-1},n-1}$ invariant. This is illustrated in Figure 13 for $n = 5$, where for simplicity of presentation we have displayed the tree in a symmetric fashion as in Figure 11(b).

D Calculation of $\Sigma^{-1/2}(\alpha_0 \ldots, \alpha_k)$

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

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

where

$$\begin{aligned} X &= (S+T)^{-1/2} \\ Y &= (S-T)^{-1/2} \end{aligned} \tag{D.2}$$

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

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

$$\Sigma_- = \Sigma(\alpha_0, \ldots, \alpha_{k-1}) - \alpha_k U_{k-1} = \Sigma(\alpha_0 - \alpha_k, \ldots, \alpha_{k-1} - \alpha_k) \tag{D.4}$$

If we introduce the following notation

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

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

$$\Sigma^{-1/2}(\alpha_0, \ldots, \alpha_k) = \begin{cases} \alpha_0^{-1/2} & \text{if } k = 0 \\ \text{Bloc}\left(\Sigma_+^{-1/2}, \Sigma_-^{-1/2}\right) & \text{if } k \geq 1 \end{cases} \tag{D.6}$$

which involves a sequence of scalar calculations.

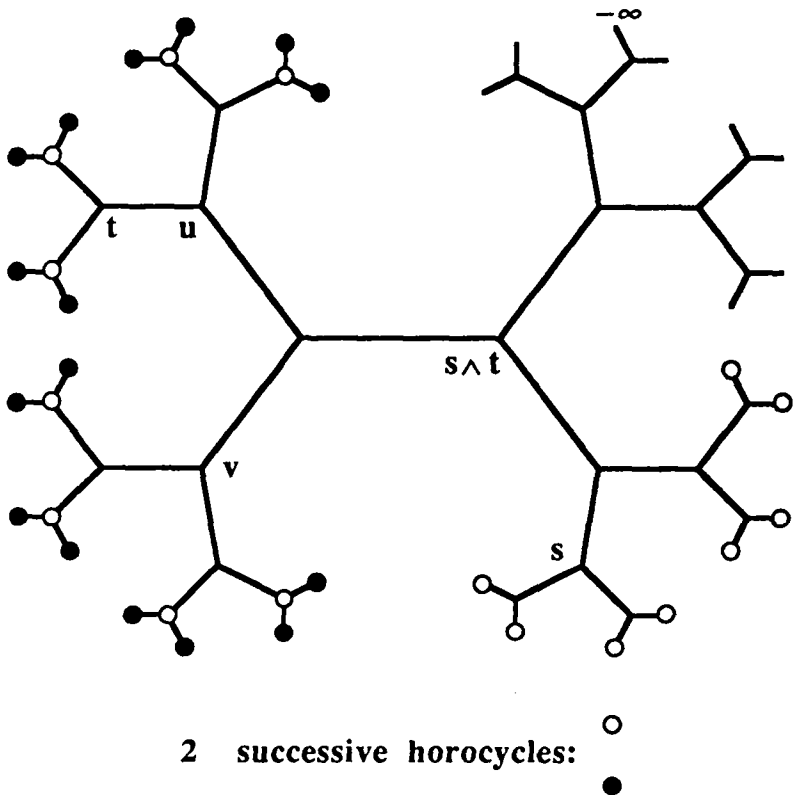

Figure 1: The dyadic homogeneous tree

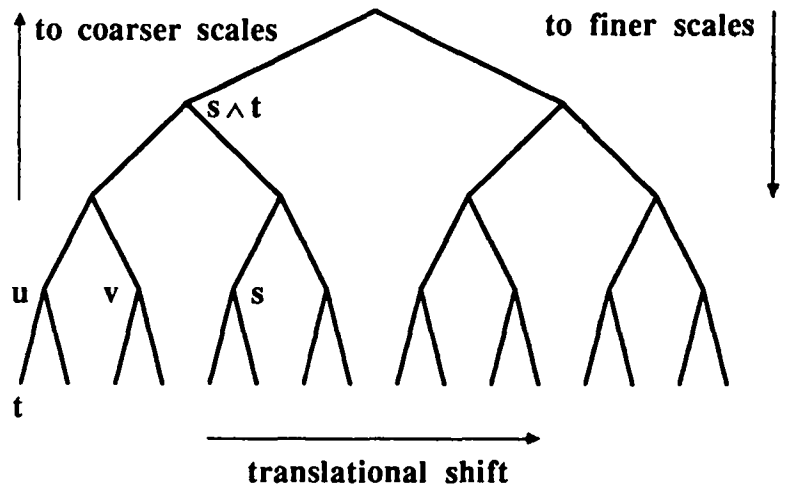

Figure 2: Showing the scale structure

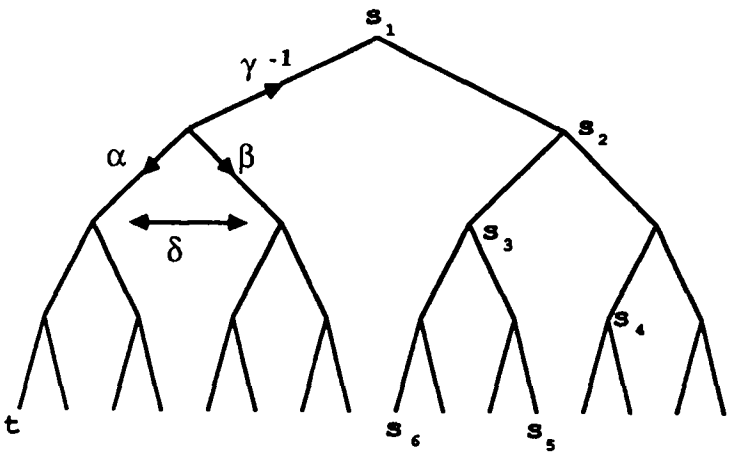

Figure 3: Encoding the moves on the tree

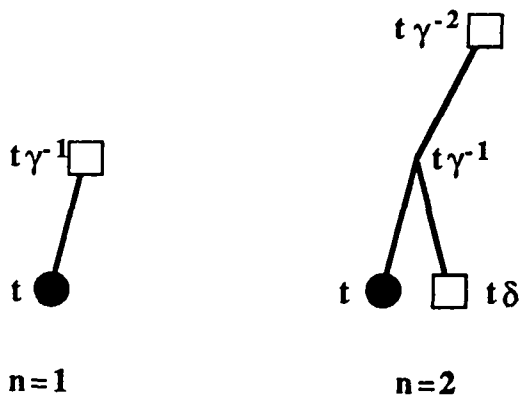

Figure 4: Illustrating $E_{t,n}$ (dots) and $F_{t,n}$ (squares) for $n = 1, 2$

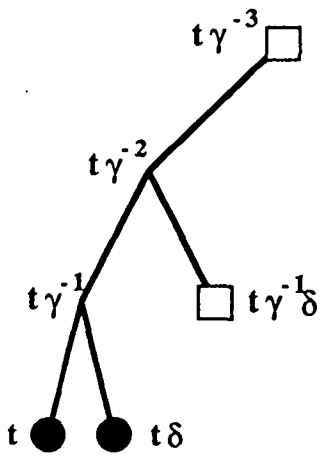

Figure 5: Illustrating $E_{t,3}$ (dots) and $F_{t,3}$ (squares)

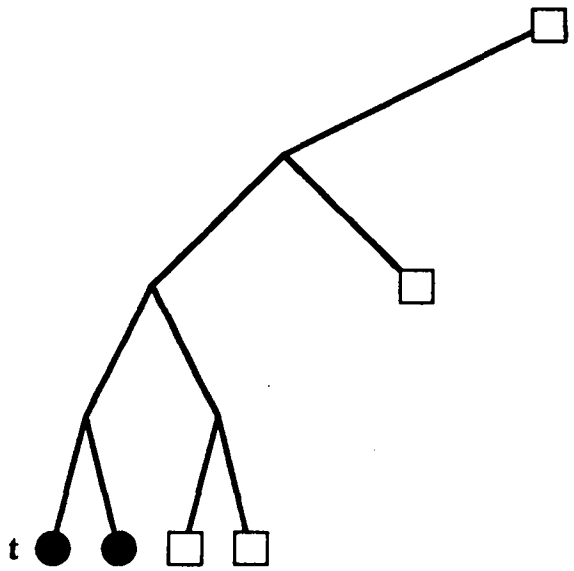

Figure 6: Illustrating $E_{t,4}$ (dots) and $F_{t,4}$ (squares)

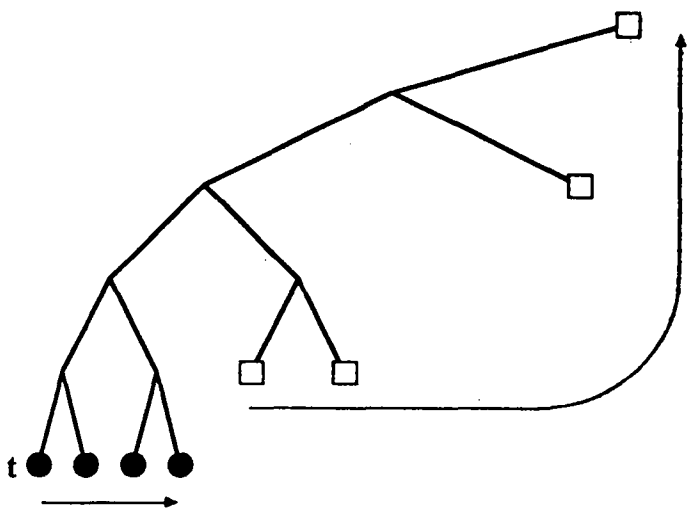

Figure 7: Illustrating $E_{t,5}$ (dots) and $F_{t,5}$ (squares)

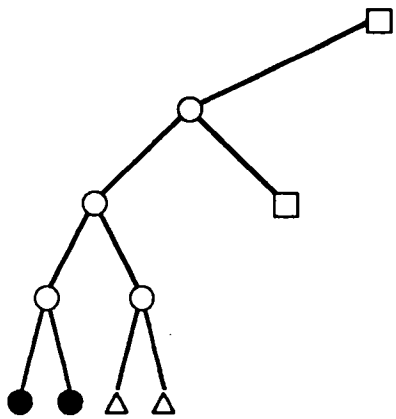

Figure 8: Illustrating $E_{t,3}$ (dots), $F_{t\gamma^{-1},3}$ (squares), $E_{t\delta^{(2)},3}$ (triangles), and $\mathcal{Y}_{t\gamma^{-1},2}$ (circles)

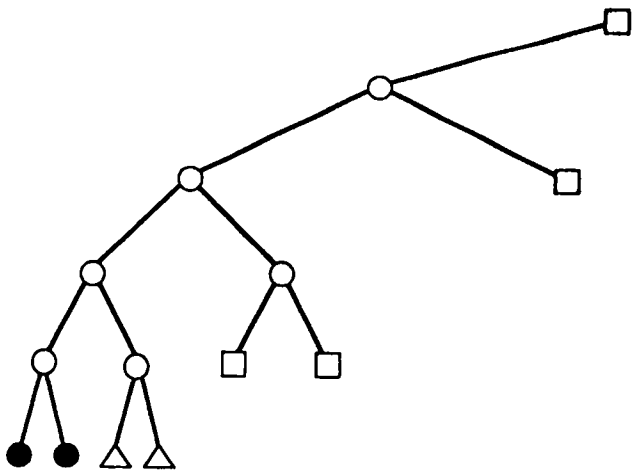

Figure 9: Illustrating $E_{t,4}$ (dots), $E_{t\delta^{(2)},4}$ (triangles), $F_{t\gamma^{-1},4}$ (squares), and $\mathcal{Y}_{t\gamma^{-1},3}$ (circles)

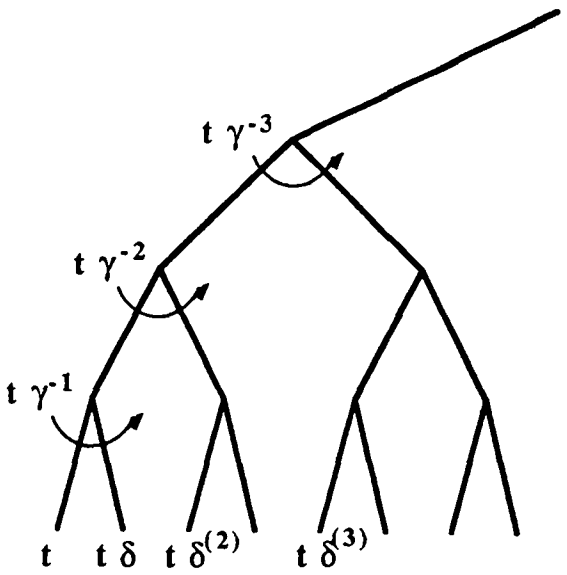

Figure 10: Illustrating local isometries that interchange points on a given horocycle

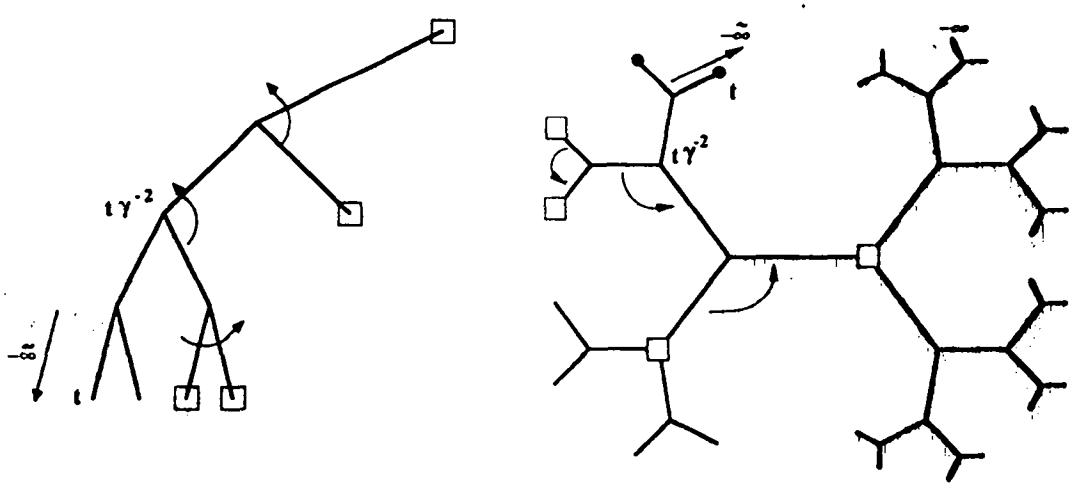

Figure 11: (a) Illustrating (i) the nodes involved in $\mathcal{T}_{t,4}$; (ii) the extreme points of $\mathcal{T}_{t,4}$ (enclosed in boxes); (iii) the alternate choice of $-\tilde{\infty}$ so that these extreme points are on the same horocycle and so that t is closer to $-\tilde{\infty}$ than the closest pivot point $t\gamma^{-2}$ (see the proof of Lemma 4.2 in section C.1); (b) A redrawing of the tree of part (a) in a more symmetric fashion

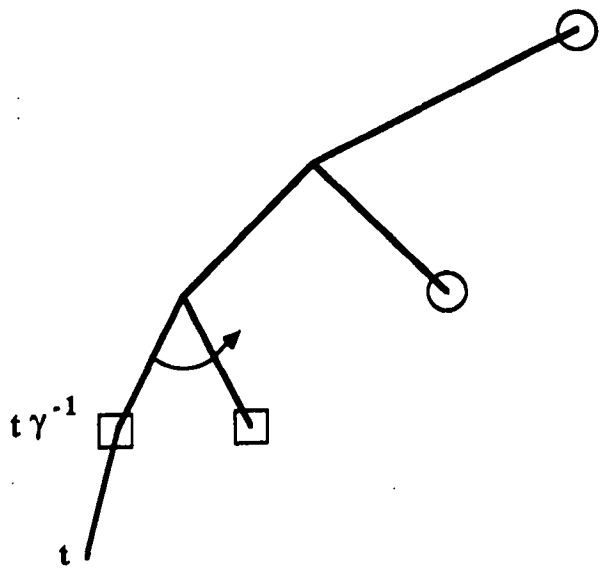

Figure 12: Illustrating (i) the node t together with the nodes comprising $\mathcal{T}_{t\gamma^{-1},3}$; (ii) the nodes (in squares) on the same horocycle to be interchanged by the isometry ψ needed in the proof of lemma 4.3 (see Section 3.2); and (iii) the nodes (in circles) to be left fixed by ψ

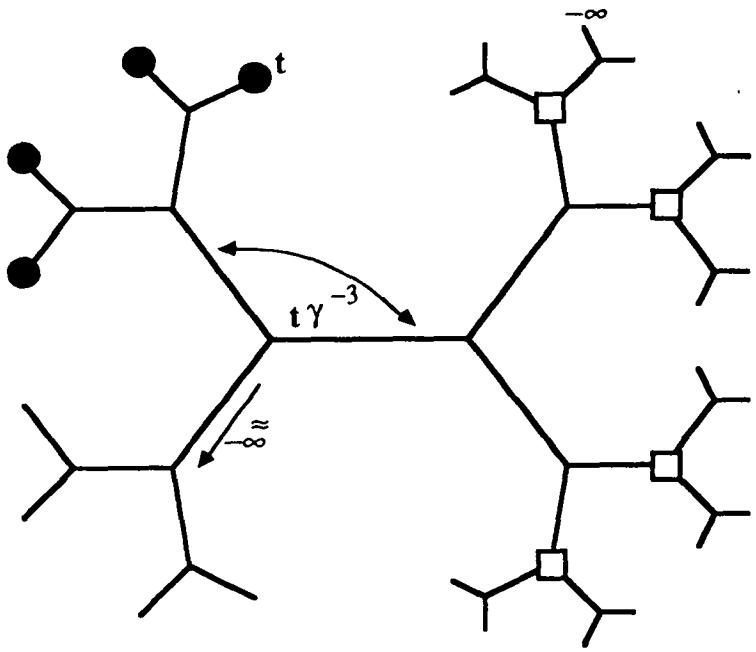

Figure 13: Illustrating $E_{t,5}$ (dots), $F_{t\gamma^{-1},5}$ (squares), and the choice for $-\tilde{\infty}$ **required** in the proof of Lemma 4.5 (section C.4). The required pivot, relative to $-\tilde{\infty}$, to interchange the dots and squares is at $t\gamma^{-3}$

Part II

Multi-Scale Autoregressive Processes Part II: Lattice Structures for Whitening and Modeling

Michèle Basseville[1], Albert Benveniste[1]
IRISA, Campus de Beaulieu
35042 RENNES CEDEX, FRANCE

Alan S. Willsky[2]
Laboratory for Information and Decision Systems, and
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, USA

Abstract

In[3] Part I [1] of this two-part paper we introduced a class of stochastic processes defined on dyadic homogenous trees. The motivation for the study of these processes comes from our desire to develop a theory for multiresolution descriptions of stochastic processes in one and multiple dimensions based on the idea underlying the recently-introduced theory of wavelet transforms. In [1] we described how this objective leads to the study of processes on trees and began the development of a theory of autoregressive (AR) models for isotropic processes on trees. In this second part we complete that investigation by developing lattice structures for the whitening and modeling of isotropic processes on trees. We also present a result relating the stability properties of these models to the reflection coefficient sequence introduced in [1]. In addition this framework allows us to obtain a detailed analysis of the Wold decomposition of processes on trees. One interesting aspect of this is that there is a significantly larger class of singular processes on dyadic trees than on the real line.

EDICS: 5.2, 5.2.1, 6.1.4.

Keywords: Multiresolution signal analysis, statistical models, Schur-Levinson parametrizations.

---

[1]M.B. is also with the Centre National de la Recherche Scientifique (CNRS) and A.B. is also with Institut National de Recherche en Informatique et en Automatique (INRIA). The research of these authors was also supported in part by Grant CNRS GO134

[2]The work of this author was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-88-0032, in part by the National Science Foundation under Grant ECS-8700903, and in part by the US Army Research Office under Contract DAAL03-86-K-0171. In addition some of this research was performed while A.S.W. was a visitor at and received partial support from INRIA.

[3]Authorization is given to publish this abstract separately.

1 Introduction

In Part I [1] of this two-part paper we introduced the class of isotropic processes on homogeneous dyadic trees, and began the analysis of the corresponding class of autoregressive (AR) processes. As developed in [1], the motivation for the study of these processes comes from our desire to provide a statistical framework for multiscale signal processing based on the structure of the recently introduced class of wavelet transforms [4].

In [1], we introduced and described the geometry of homogenous dyadic trees and a natural notion of "past" and "future", where a move into the "past" ("future") corresponds to moving to a coarser (finer) scale description of a signal. The class of isotropic processes on trees was also introduced in [1], and, with our notions of past and future, we defined the class of autoregressive (AR) isotropic processes and began the study of their parametrization. The major result of [1] was to establish that the only suitable parametrization of isotropic processes on the dyadic tree is obtained via *reflection coefficients* following the generalization of the Schur-Levinson parametrization techniques for usual time series. In this second part, we further investigate the properties of isotropic processes in terms of their reflection coefficients. In particular, in this paper we use the analysis in [1] both to construct lattice structures for the whitening and modeling of AR processes on dyadic trees and to analyze in detail these models and the properties of isotropic processes.

This paper relies heavily on the framework and results of [1], and we refer the reader to that paper for reference. In the next section we provide a brief summary of some of the basic notation and constructs from [1]. Section 3 is then devoted to the presentation of whitening and modeling filters for AR isotropic processes. Unnormalized as well as normalized versions of these filters are given. In particular, the normalized modeling filter appears as a tree structured scattering system. Then, in Section 4, several properties of isotropic processes are analyzed in terms of the reflection coefficient sequence. Specifically AR processes are characterized as being the processes with only finitely many non-zero reflection coefficients, purely non-

deterministic processes are characterized in a fairly simple way, a stability result for the modeling filters is presented, and finally it is shown that every finite set of reflection coefficients properly define a unique AR process provided they belong to an easily defined domain. Finally, future issues, both practical and theoretical, are discussed in the conclusion. Many of the results presented here, while paralleling those for time series, are more complex than their time series counterparts due to the significant increase in geometric complexity in going from a homogeneous tree of order 1—i.e. the usual discrete-time index set—to the dyadic tree, which is of order 2. For example, as introduced in [1] and described in detail in Section 4, the prediction error processes associated with lattice filters on dyadic trees are vector processes of dimension that increases with filter order.

2 Dyadic Trees, Isotropic Processes, and Prediction Error Recursions

In this section we review some of the basic concepts and constructs described in [1]. We refer the reader to [1] for details.

A homogeneous dyadic tree $\mathcal{T}$, as illustrated in Figure 1, has a natural notion of distance $d(s,t)$ between any two nodes $s,t \in \mathcal{T}$. By choosing a particular boundary point, denoted by $-\infty$, we can redraw $\mathcal{T}$ as in Figure 2. Here all of the points that are at the same "distance from $-\infty$" appear on the same level or *horocycle*. For multiscale processing we can think of each horocycle as corresponding to describing signals at a particular scale, with finer scales being farther from $-\infty$. Also, as illustrated in the figure the choice of $-\infty$ leads naturally to a backward (fine-to-coarse) shift, γ^{-1}, and two forward (coarse-to-fine) shifts α and β. Also it is useful to introduce the operator δ. As indicated in Figure 2, the transformation δ which interchanges nodes t and $t\delta$ for all $t \in \mathcal{T}$, can be thought of locally as an interchange pivoted at the immediate ancestor $t\gamma^{-1}$. Higher order operators $\delta^{(n)}$ correspond to interchanges pivoted that are more distant ancestors of t (i.e. $t\gamma^{-n}, n > 1$). The node $t\delta^{(2)}$ and $t\delta^{(3)}$ are indicated in the figure.

As developed in [1], all nodes in $\mathcal{T}$ can be coded in terms of shifts from a specified, arbitrary node t_o. Specifically, let

$$\mathcal{L} = (\gamma^{-1})^* \cup (\gamma^{-1})^* \delta \{\alpha, \beta\}^* \cup \{\alpha, \beta\}^* \tag{2.1}$$

Then $\mathcal{T} = \{t_o w | w \in \mathcal{L}\}$. The order $|w|$ of any move $w \in \mathcal{L}$ is defined as

$$|w| = d(t, wt) \tag{2.2}$$

A move w is *causal*, denoted by $w \preceq 0$, if wt is on the same or a coarser horocycle than that on which t is located.

A zero-mean stochastic process $Y_t, t \in \mathcal{T}$, indexed by nodes on the tree is *isotropic* if the correlation between Y at any two nodes depends only on the distance between those nodes, i.e.

$$E[Y_t Y_s] = r_{d(t,s)} \tag{2.3}$$

Equivalently Y_t is isotropic if $Z_t = Y_{f(t)}$ has the same statistics as Y_t for any isometry $f : \mathcal{T} \to \mathcal{T}$, i.e. any one-to-one and onto map of $\mathcal{T}$ onto itself that preserves distances. An AR model of order p has the form

$$Y_t = \sum_{\substack{w \prec 0 \\ |w| \leq p}} a_w Y_{tw} + \sigma W_t \tag{2.4}$$

where W_t is unit variance white noise. Our interest here is in developing AR models for isotropic processes, and as discussed in [1], the constraints of isotropy imply rather complex constraints on the a_w coefficients in (2.4). Note also that the number of these coefficients essentially *doubles* as the order increases by 1.

In [1] we began the process of developing an alternate description of isotropic AR processes in terms of generalizations of the Levinson and Schur recursions for stationary time series. Because of the structure of the dyadic trees—in particular the fact that the number of nodes at a given distance from a specified node increases geometrically with distance—the development of these recursions and the lattice filters to be described here involve prediction error vectors of dimensions increasing with the order of prediction. Specifically, define the nth order past of Y at node t

$$\mathcal{Y}_{t,n} \stackrel{\Delta}{=} \mathcal{H}\{Y_{tw} : w \preceq 0, |w| \leq n\} \tag{2.5}$$

where $\mathcal{H}\{\cdots\}$ denotes the linear span of a set of random variables. Then the nth order backward prediction errors at node t are given by

$$\mathcal{F}_{t,n} = \mathcal{H}\{F_{t,n}(w) : |w| = n, w \preceq 0\} \tag{2.6}$$

where

$$F_{t,n}(w) \stackrel{\Delta}{=} Y_{tw} - E(Y_{tw}|\mathcal{Y}_{t,n-1}) \tag{2.7}$$

We also let $F_{t,n}$ denote the full $2^{[\frac{n}{2}]}$–dimensional vector of $F_{t,n}(w)$, where $[x] =$ largest integer $\leq x$ and where the ordering of the w in (2.6) used in constructing $F_{t,n}$ is described in [1]. Similarly we have the nth order forward prediction errors at node t

$$\mathcal{E}_{t,n} \stackrel{\Delta}{=} \mathcal{H}\{E_{t,n}(w) : |w| < n \text{ and } w \succ 0\} \tag{2.8}$$

where

$$E_{t,n}(w) \triangleq Y_{tw} - E\left(Y_{tw}|\mathcal{Y}_{t\gamma^{-1},n-1}\right) \tag{2.9}$$

The ordering of the w in (2.8) to construct the $2^{[\frac{n-1}{2}]}$–dimensional vector $E_{t,n}$ is described in [1].

In [1] we began the analysis of the recursive computation of these prediction errors as the order, n, increases. What we found was that as in the usual Levinson recursions for time series the forward and backward prediction errors of one order could be expressed in terms of projections onto prediction error vectors of the preceding order. Most importantly, the constraint of isotropy allowed us to show that the required projections onto multidimensional spaces such as $\mathcal{F}_{t,n}$ and $\mathcal{E}_{t,n}$ reduced in fact to projections onto specific scalar random variables, namely the barycenters of the prediction error vectors:

$$e_{t,n} = 2^{-[\frac{n-1}{2}]} \sum_{|w|<n, w\succ 0} E_{t,n}(w) \tag{2.10}$$

$$f_{t,n} = 2^{-[\frac{n}{2}]} \sum_{|w|=n, w\succeq 0} F_{t,n}(w) \tag{2.11}$$

Indeed these projections can be expressed in terms of a *single* reflection coefficient, k_n. Furthermore, as shown in [1], there exist a set of scalar Levinson recursions for the barycenter error processes. In particular for n even

$$e_{t,n} = e_{t,n-1} - k_n f_{t\gamma^{-1},n-1} \tag{2.12}$$

$$f_{t,n} = \frac{1}{2}\left(f_{t\gamma^{-1},n-1} + e_{t\delta(\frac{n}{2}),n-1}\right) - k_n e_{t,n-1} \tag{2.13}$$

where

$$k_n = \mathrm{cor}\left(e_{t,n-1}, f_{t\gamma^{-1},n-1}\right) \tag{2.14}$$

and $\mathrm{cor}\,(x,y) = E(xy)/\left[E(x^2)E(y^2)\right]^{1/2}$. For n odd, $n > 1$:

$$e_{t,n} = \frac{1}{2}\left(e_{t,n-1} + e_{t\delta(\frac{n-1}{2}),n-1}\right) - k_n f_{t\gamma^{-1},n-1} \tag{2.15}$$

$$f_{t,n} = f_{t\gamma^{-1},n-1} - \frac{1}{2}k_n\left(e_{t,n-1} + e_{t\delta(\frac{n-1}{2}),n-1}\right) \tag{2.16}$$

with

$$k_n = \mathrm{cor}\left(\frac{1}{2}\left(e_{t,n-1} + e_{t\delta^{\left(\frac{n-1}{2}\right)},n-1}\right), f_{t\gamma^{-1},n-1}\right) \tag{2.17}$$

Also for $n = 1$, $e_{t,1} = E_{t,1}$, $f_{t,1} = F_{t,1}$, and

$$F_{t,1} = Y_{t\gamma^{-1}} - k_1 Y_t \tag{2.18}$$

$$E_{t,1} = Y_t - k_1 Y_{t\gamma^{-1}} \tag{2.19}$$

where

$$k_1 = \frac{E\left[Y_{t\gamma^{-1}} Y_t\right]}{E\left[Y_{t\gamma^{-1}}^2\right]} = \frac{r_1}{r_0} \tag{2.20}$$

In addition the variances of the prediction errors satisfy the following: for n even

$$\sigma_{e,n}^2 = E\left(e_{t,n}^2\right) = \left(1 - k_n^2\right)\sigma_{n-1}^2 \tag{2.21}$$

$$\sigma_{f,n}^2 = E\left(f_{t,n}^2\right) = \left(\frac{1 + k_n}{2} - k_n^2\right)\sigma_{n-1}^2 \tag{2.22}$$

For n odd

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

where (2.23) holds for $n = 1$ as well, with $\sigma_{f,0}^2 = r_0$. By using these equations it is possible to derive a recursive procedure for computing the k_n that is the counterpart of the recursions in the standard Levinson algorithm for time series. We also have Schur recursions which provide an alternative mechanism for computing the reflection coefficient sequence. Specifically, define the formal power series:

$$P_n \triangleq \mathrm{cov}\left(Y_t, e_{t,n}\right) \triangleq \sum_{w \preceq 0} E\left(Y_t e_{tw,n}\right) \cdot w \tag{2.24}$$

$$Q_n \triangleq \mathrm{cov}\left(Y_t, f_{t,n}\right) \triangleq \sum_{w \preceq 0} E\left(Y_t f_{tw,n}\right) \cdot w \tag{2.25}$$

and recall the following operators on formal power series we introduced in [1]: given

$$S = \sum_{w \in \mathcal{L}} s_w \cdot w$$

we set

$$\begin{aligned}\gamma[S] &= \sum_{w\in\mathcal{L}} s_{w\gamma^{-1}}\cdot w \\ \delta^{(k)}[S] &= \sum_{w\in\mathcal{L}} s_{w\delta^{(k)}}\cdot w\end{aligned}$$

Then for n even

$$P_n = P_{n-1} - k_n\gamma[Q_{n-1}] \tag{2.26}$$

$$Q_n = \frac{1}{2}\left(\gamma[Q_{n-1}] + \delta^{(\frac{n}{2})}[P_{n-1}]\right) - k_n P_{n-1} \tag{2.27}$$

where

$$k_n = \frac{\gamma[Q_{n-1}](0) + \delta^{(\frac{n}{2})}[P_{n-1}](0)}{2P_{n-1}(0)} \tag{2.28}$$

while for n odd

$$P_n = \frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})}[P_{n-1}]\right) - k_n\gamma[Q_{n-1}] \tag{2.29}$$

$$Q_n = \gamma[Q_{n-1}] - k_n\frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})}[P_{n-1}]\right) \tag{2.30}$$

$$k_n = \frac{2\gamma[Q_{n-1}](0)}{P_{n-1}(0) + \delta^{(\frac{n-1}{2})}[P_{n-1}](0)} \tag{2.31}$$

where

$$P_0 = Q_0 = \sum_{w\preceq 0} r_{|w|}\cdot w \tag{2.32}$$

We also note here that, as for time series, there are constraints on the reflection coefficients, which, thanks to the conditions required for isotropy, are slightly more complex for isotropic processes on dyadic trees:

$$\text{for } n \text{ even}, \quad -\frac{1}{2} \le k_n \le 1 \tag{2.33}$$

$$\text{for } n \text{ odd}, \quad -1 \le k_n \le 1 \tag{2.34}$$

As we develop in this paper, these results lead to lattice structures for AR processes on dyadic trees in which only *one* new reflection coefficients is introduced as the

order increases by one. Furthermore the constraints (2.33), (2.34) on these coefficients are quite simple and are decoupled from one another. Thus the lattice filter parametrization of AR processes is far superior to the direct AR model (2.4)

3 Vector Levinson Recursions and Modeling and Whitening Filters

In [1] we showed that the recursive computation of the components of the prediction error vectors $E_{t,n}$ and $F_{t,n}$ involved projections onto the barycenter error processes. In addition we developed scalar Levinson recursions for the barycenters. In this section we combine these results in order to develop whitening and modeling filters for Y_t. As we will see, in order to produce true whitening filters, it will be necessary to perform a further normalization of the innovations. However, the formulas for $E_{t,n}$ and $F_{t,n}$ are simpler, and consequently we begin with them.

3.1 Filters Involving the Unnormalized Residuals

To begin, let us introduce a variation on notation used to describe the structure of the covariance matrix of the prediction error $E_{t,n}$ which we denoted in [1] by $\Sigma_{E,n}$. In particular we let 1_* denote a unit vector all of whose components are the same:

$$1_* = \frac{1}{\sqrt{\dim 1}} 1 \tag{3.1}$$

We also define the matrix

$$U_* = 1_* 1_*^T \tag{3.2}$$

which has a single nonzero eigenvalue of 1. Equations (3.1), (3.2) define a family of vectors and matrices of different dimensions. The dimension used in any of the expressions to follow is that required for the expression to make sense. We also note the following identities:

$$U_* U_* = U_* \tag{3.3}$$

$$f^* = 1_*^T F = \frac{1}{\sqrt{\dim F}} \sum_w F(w) \tag{3.4}$$

$$1f = 1_* f^* = U_* F \tag{3.5}$$

where $F = \{F(w)\}$ is a vector indexed by words w ordered as described in [1], where f is its barycenter, and where f^* is a normalized version of its barycenter.

The results of Section 4 of [1] lead directly to the following recursions for the prediction error vectors:

Theorem 3.1 *The prediction error vectors $E_{t,n}$ and $F_{t,n}$ satisfy the following recursions, where the k_n are the reflection coefficients for the process Y_t:*

For n even:

$$E_{t,n} = E_{t,n-1} - k_n U_* F_{t\gamma^{-1},n-1} \tag{3.6}$$

$$F_{t,n} = \begin{bmatrix} E_{t\delta^{(\frac{n}{2})},n-1} \\ F_{t\gamma^{-1},n-1} \end{bmatrix} - k_n \begin{bmatrix} U_* \\ U_* \end{bmatrix} E_{t,n-1} \tag{3.7}$$

For n odd, $n > 1$:

$$E_{t,n} = \begin{bmatrix} E_{t,n-1} \\ E_{t\delta^{(\frac{n-1}{2})},n-1} \end{bmatrix} - k_n U_* F_{t\gamma^{-1},n-1} \tag{3.8}$$

$$F_{t,n} = F_{t\gamma^{-1},n-1} - k_n U_* \begin{bmatrix} E_{t,n-1} \\ E_{t\delta^{(\frac{n-1}{2})},n-1} \end{bmatrix} \tag{3.9}$$

while for $n = 1$ $F_{t,1}$ and $E_{t,1}$ are scalars satisfying (2.18), (2.19). Here the reflection coefficient sequence k_n is calculated from the correlation function, r_n, of Y_t according to either the Levinson or Schur recursions described in Section 2.

Proof: Equations (2.18), (2.19) for $n = 1$ are exactly (3.17), (3.19) in [1]. As indicated previously, the remainder of this result is also a direct consequence of the analysis in Sections 3 and 4 of [1]. For example, from (3.16), Lemma 4.1 (4.6) of [1], and (3.5) of this paper, we have the following chain of equalities for n even:

$$\begin{aligned} E_{t,n} &= E_{t,n-1} - E\left(E_{t,n-1} | F_{t\gamma^{-1},n-1}\right) \\ &= E_{t,n-1} - \lambda 1 f_{t\gamma^{-1},n-1} \\ &= E_{t,n-1} - \lambda U_* F_{t\gamma^{-1},n-1} \end{aligned} \tag{3.10}$$

where λ is a constant to be determined. If we premultiply this equality by $(\dim E_{t,n-1})\,1^T$, we obtain the formula for the barycenter of $E_{t,n-1}$, and from (2.12) we see that $\lambda = k_n$. The other formulae are obtained in an analogous fashion.

The form of these whitening filters deserves some comment. Note first that the stages of the filter are of growing dimension, reflecting the growing dimension of the $E_{t,n}$ and $F_{t,n}$ as n increases. Nevertheless each stage is characterized by a *single* reflection coefficient. Thus, while the dimension of the innovations vector of order n is on the order of $2^{\frac{n}{2}}$, only n coefficients are needed to specify the whitening filter for its generation. This, of course, is a direct consequence of the constraint of isotropy and the richness of the group of isometries of the tree.

In [1] we obtained recursions (2.21)–(2.23) for the variances of the barycenters of the prediction vectors. Theorem 3.1 above provides us with the recursions for the covariances and correlations for the entire prediction error vectors. We summarize these and other facts about these covariances in the following.

Corollary: Let $\Sigma_{E,n}$, $\Sigma_{F,n}$ denote the covariances of $E_{t,n}$ and $F_{t,n}$, respectively. Then

1. For n even

   (a) The eigenvalue of $\Sigma_{E,n}$ associated with the eigenvector $[1,\ldots,1]$ is

   $$\mu_{E,n} = 2^{\frac{n}{2}-1}\sigma^2_{e,n} \tag{3.11}$$

   where $\sigma^2_{e,n}$ is the variance of $e_{t,n}$.

   (b) The eigenvalue of $\Sigma_{F,n}$ associated with the eigenvector $[1,\ldots,1]$ is

   $$\mu_{F,n} = 2^{\frac{n}{2}}\sigma^2_{f,n} \tag{3.12}$$

   where $\sigma^2_{f,n}$ is the variance of $f_{t,n}$.

2. For n odd,

   $$\Sigma_{E,n} = \Sigma_{F,n} = \Sigma_n \tag{3.13}$$

and the eigenvalue associated with the eigenvector $[1, \ldots, 1]$ is

$$\mu_n = \mu_{E,n} = \mu_{F,n} = 2^{\frac{n-1}{2}} \sigma_n^2 \tag{3.14}$$

where σ_n^2 is the variance of both $e_{t,n}$ and $f_{t,n}$.

3. For n even

$$\Sigma_n \triangleq \Sigma_{F,n} = \operatorname{cov} \begin{pmatrix} E_{t,n} \\ E_{t\delta^{(\frac{n}{2})},n} \end{pmatrix} = \begin{bmatrix} \Sigma_{E,n} & \lambda_n U \\ \lambda_n U & \Sigma_{E,n} \end{bmatrix} \tag{3.15}$$

where $U = 11^T$, and

$$\Sigma_{E,n} = \Sigma_{n-1} - k_n^2 \sigma_{n-1}^2 U \tag{3.16}$$

$$\lambda_n = \left(k_n - k_n^2\right) \sigma_{n-1}^2 \tag{3.17}$$

4. For n odd, $n > 1$

$$\Sigma_n = \begin{bmatrix} \Sigma_{E,n-1} & \lambda_{n-1} U \\ \lambda_{n-1} U & \Sigma_{E,n-1} \end{bmatrix} - k_n^2 \sigma_{f,n-1}^2 U \tag{3.18}$$

5. For $n = 1$

$$\Sigma_1 = \left(1 - k_1^2\right) r_0 \tag{3.19}$$

Proof: Equations (3.11), (3.12), and (3.14) follow directly from the definition of the barycenter. For example, for n even

$$2^{\frac{n}{2}-1} e_{t,n} = 1^T E_{t,n} \tag{3.20}$$

from which (3.11) follows immediately. Equation (3.13) is a consequence of Lemma 4.1 of [1]. To verify (3.15) let us first evaluate (3.6) at both t and $t\delta^{(\frac{n}{2})}$:

$$\begin{pmatrix} E_{t,n} \\ E_{t\delta^{(\frac{n}{2})},n} \end{pmatrix} = \begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(\frac{n}{2})},n-1} \end{pmatrix} - k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} F_{t\gamma^{-1},n-1} \tag{3.21}$$

The first equality in (3.15) is then a direct consequence of Lemma 4.1 of [1] (compare (3.7) and (3.21)). The form given in the right-most expression in (3.15) is also

immediate: the equality of the diagonal blocks is due to isotropy, while the form of the off-diagonal blocks again follows from Lemma 4.1 of [1]. The specific expression for $\Sigma_{E,n}$ in (3.16) follows directly from the second equality in (3.10), while (3.17) follows from (3.21) and the fact that

$$E\left[E_{t,n-1}(w)E_{t\delta(\frac{n}{2}),n-1}(w')\right] = k_n\sigma^2_{n-1} \tag{3.22}$$

which in turn follows from Lemma 4.1 and (4.27) of [1]. Finally, (3.18) follows from (3.15) and (3.8), and (3.19) is immediate from (2.18)–(2.20).

Just as with time series, the whitening filter specification leads directly to a modeling filter for Y_t.

Theorem 3.2 *The modeling filter for Y_t is given by the following. For n even*

$$\begin{pmatrix} E_{t,n-1} \\ \cdots \\ F_{t,n} \end{pmatrix} = S(k_n) \begin{pmatrix} E_{t,n} \\ E_{t\delta(\frac{n}{2}),n} \\ F_{t\gamma^{-1},n-1} \end{pmatrix} \tag{3.23}$$

where[4]

$$S(k_n) \triangleq \begin{bmatrix} I & 0 & k_nU_* \\ \cdots\cdots & \cdot & \cdots\cdots \\ -k_nU_* & I & (k_n - k_n^2)\,U_* \\ -k_nU_* & 0 & (I - k_n^2U_*) \end{bmatrix} \tag{3.24}$$

For n odd, $n \geq 1$:

$$\begin{pmatrix} E_{t,n-1} \\ E_{t\delta(\frac{n-1}{2}),n-1} \\ \cdots\cdots \\ F_{t,n} \end{pmatrix} = S(k_n) \begin{pmatrix} E_{t,n} \\ F_{t\gamma^{-1},n-1} \end{pmatrix} \tag{3.25}$$

where

$$S(k_n) \triangleq \begin{bmatrix} I & k_nU_* \\ \cdots\cdots & \cdots\cdots \\ -k_nU_* & (I - k_n^2U_*) \end{bmatrix} \tag{3.26}$$

[4] In fact, we should properly write $S(k_n, n)$ since the dimension of the blocks depends on n; nevertheless we chosed to write $S(k_n)$ to simplify the notation; this will be done everywhere in the sequel.

while for $n = 1$*:*

$$\begin{pmatrix} Y_t \\ F_{t,1} \end{pmatrix} = \begin{pmatrix} 1 & k_1 \\ -k_1 & 1-k_1^2 \end{pmatrix} \begin{pmatrix} E_{t,1} \\ Y_{t\gamma^{-1}} \end{pmatrix} \triangleq S(k_1) \begin{pmatrix} E_{t,1} \\ Y_{t\gamma^{-1}} \end{pmatrix} \tag{3.27}$$

These equations can be verified by solving (3.6)–(3.9) and (2.18)–(2.20) to obtain expressions for E's of order $n-1$ and F's of order n in terms of E's of order n and F's of order $n-1$. Note should also be made of the dimensions of the various signals and matrices in Theorem 3.2. In particular, for n even the two components on the left-hand side of (3.23) are of dimensions $2^{(n/2)-1}$ and $2^{(n/2)}$, respectively, while all three of the vectors on the right-hand side of (3.23) are of dimension $2^{(n/2)-1}$ and each of the square blocks in (3.24) is $2^{(n/2)-1}$–dimensional. For n odd, $n > 1$, both components of the right-hand side of (3.25) are $2^{(n-2)/2}$–dimensional as is the $F_{t,n}$–block on the left-hand side. The two E–blocks on the left-hand side, however, are $2^{(n-3)/2}$–dimensional, and the four square blocks in (3.26) are $2^{(n-1)/2}$–dimensional. We have included dotted lines in (3.23)–(3.26) to emphasize how these mappings operate. Note also that the first section (3.27) of the modeling filter involves only scalar quantities.

As is the case for time series, the lattice modeling filter of Theorem 3.2 has a scattering layer structure. An important difference here is that the growing dimension of the prediction errors leads to a tree-like structure for the scattering diagram, and because of this, we find that groups of values of Y are calculated together in this structure. In particular, from Theorem 3.2 we can deduce that if we consider a modeling filter of odd length N, then this modeling filter can be viewed as a map from the $2^{(N-1)/2}$–dimensional input vector $E_{t,N}$ to the $2^{(N-1)/2}$–dimensional set of outputs $\{Y_{tw} \mid |w| \leq N, w \asymp 0\}$. For N even, the modeling filter maps the two $2^{(N/2)-1}$–dimensional input vectors $E_{t,N}$, $E_{t,\delta^{(N/2)},N}$ to the $2^{(N/2)}$–dimensional set of outputs $\{Y_{tw} \mid |w| \leq N, w \asymp 0\}$. The case of $N = 6$ is illustrated in Figure 3. In this case the input vectors $E_{t,6}$ and $E_{t\delta^{(3)},6}$, produce the outputs Y_{tw} for $w \asymp 0$, $|w| \leq 6$ (as well as the backward errors $F_{t,6}$ and $F_{t\delta^{(3)},6}$, which aren't actually needed for the recursion). The E–vectors of various orders propagate from left to right, while

the F's propagate from right to left. The small black squares represent γ^{-1} operations and the blocks labeled "1", "2", etc. perform the computations described in Theorem 3.2. The details of the operation of this system, however, requires further explanation.

Let us first look at the situation for n odd, in which case each block labeled "n" performs the calculations given by (3.25) (or (3.27) for $n = 1$). For example, the inputs to the top "3" block in the figure (which has been shaded) are $E_{t,3}$ and $F_{t\gamma^{-1},2}$, while the outputs are $E_{t,2}$, $E_{t\delta,2}$ and $F_{t,3}$. Note that this block is connected to the right to systems generating *both* Y_t and $Y_{t\delta}$, but apparently we do not need a corresponding "3" block at $t\delta$ in addition to the one at t. To understand this, consider writing (3.25) at $t\delta^{(\frac{n-1}{2})}$ rather than at t:

$$\begin{bmatrix} E_{t\delta^{(\frac{n-1}{2})},n-1} \\ E_{t,n-1} \\ \cdots\cdots\cdots \\ F_{t\delta^{(\frac{n-1}{2})},n} \end{bmatrix} = S(k_n) \begin{bmatrix} E_{t\delta^{(\frac{n-1}{2})},n} \\ F_{t\delta^{(\frac{n-1}{2})}\gamma^{-1},n-1} \end{bmatrix} \tag{3.28}$$

where we have used the fact that for any k, $\delta^{(k)}$ is its own inverse. Note that the first two components of the output in (3.28) are simply a permutation of the first two in (3.25). The last outputs in these equations and both inputs apparently differ. However it is easily checked that the outputs $F_{t,n}$ and $F_{t\delta^{(\frac{n-1}{2})},n}$ are identical up to a permutation of the ordering of components, as are the input pair $E_{t,n}$ and $E_{t\delta^{(\frac{n-1}{2})},n}$ and the input pair $F_{t\gamma^{-1},n-1}$ and $F_{t\delta^{(\frac{n-1}{2})}\gamma^{-1},n-1}$ (this latter fact is proved in "cigarillos' lemma" of appendix A and expressed via the "cigarillos" –dotted connections– of Figure 3). Thus there is actually no need to have a "3" block at $t\delta$ as there was at t, or more generally an "n" block at $t\delta^{(\frac{n-1}{2})}$ as well as at t.

For n even, the blocks labeled "n" perform the calculations as specified by (3.23). For example, the inputs to the top "2" box in the figure are $E_{t,2}$, $E_{t\delta,2}$, and $F_{t\gamma^{-1},1}$, while the outputs are $E_{t,1}$ and $F_{t,2}$. Again it is important to examine the analogous computation at a related point. Specifically, consider evaluating (3.23) at the point

$t\delta^{(\frac{n}{2})}$:

$$\begin{bmatrix} E_{t\delta^{(\frac{n}{2})},n-1} \\ \cdots\cdots\cdots \\ F_{t\delta^{(\frac{n}{2})},n} \end{bmatrix} = S(k_n) \begin{bmatrix} E_{t\delta^{(\frac{n}{2})},n} \\ E_{t,n} \\ F_{t\delta^{(\frac{n}{2})}\gamma^{-1},n-1} \end{bmatrix} \tag{3.29}$$

Note first that the first outputs of (3.23) and (3.29), namely, $E_{t,n-1}$ and $E_{t\delta^{(\frac{n}{2})},n-1}$ are in fact distinct, and thus it is necessary to implement the computation (3.29). For example, the second "2" block (also shaded) in Figure 3 computes as one of its outputs $E_{t\delta,1}$. Next note that the other outputs, $F_{t,n}$ and $F_{t\delta^{(\frac{n}{2})},n}$, of (3.23) and (3.29) are *not* identical. However, these signals must pass through a γ^{-1} operation before entering the corresponding "$n+1$" block, and we have already seen in our analysis of the odd case that $F_{t\gamma^{-1},n}$ and $F_{t\delta^{(\frac{n}{2})}\gamma^{-1},n}$ are identical up to a permutation . Thus only one of these is needed for a block at level $n+1$. This is indicated in the figure by a connecting, dotted bar between the γ^{-1} block immediately to the left of pairs of even numbered blocks, with only one of these identical signals continuing backward to the corresponding $n+1$ block. For example the two left-going output signals of the shaded "2" blocks, $F_{t,2}$ and $F_{t\delta,2}$ are merged in this way after the γ^{-1} operation on each.

Examining next the right-hand sides of (3.23) and (3.29) we see that the first two inputs are identical except for a flip in the order. This is captured in the figure, as can be seen for $n=2$, where the two inputs entering from the left of the second "2" block are the same as those for the first "2" block, except in reverse position. It is also not difficult to check that the last inputs $F_{t\gamma^{-1},n-1}$ and $F_{t\delta^{(\frac{n}{2})}\gamma^{-1},n-1}$ are identical up to a permutation of components. While these signals do enter individual blocks we have again indicated that they are the same by a connecting dotted bar between the γ^{-1} blocks immediately to the *right* of pairs of even numbered blocks. For the case of $n=2$, the two left-going *input* signals of the shaded "2" blocks, $F_{t\gamma^{-1},1}$ and $F_{t\delta\gamma^{-1},1}$ are identical and are connected by such a dotted bar.

Figures 4 and 5 describe in more explicit terms the data flow and memory structure for the system of order 6. Specifically suppose that we have finished the computations required at the horocycle indicated with squares in Figure 4. As

indicated in the tree at the top of this figure (via a shaded bar connecting the squares), sets of 4 nodes at this level are coupled together (more generally for an nth-order model $2^{[\frac{n-1}{2}]}$ points are coupled together). The state for this set of four nodes is indicated above the nodes: we have stored the scalar values of Y, F_1, and F_2 at each node, we have stored the 2–vectors F_3 and F_4 for each of two pairs of these nodes, and we have stored a single 4–vector F_5 for the set of 4 nodes. Given these quantities and the two 4–vector E_6 inputs for each of the two sets of 4 descendent nodes (indicated with circles, with a connecting bar for each set), the model performs two parallel computations (which are identical in structure) to produce the required variables to be stored at each of the two sets of descendant nodes. Figure 5 illustrates in more detail how these computations are distributed and performed. Here at each level the variables required as inputs are indicated with "?", while those produced as outputs are indicated with "!". Futhermore those inputs corresponding to the stored state are indicated above each layer of the computation, while below each figure we indicate the inputs received externally ($?E_6$) or from previous layers (all other $?E$'s). We also indicate below each layer the outputs produced, some of which (the $!F$'s at layers 2–6 and the $!Y$'s at level 1) form components of the state at the next horocycle while others of which (the $!E$'s) are used as inputs by succeeding layers. For example at the top level F_5 is stored and two E_6 vectors are received as the only external inputs. This layer, as shown in Figure 3, has two actual sets of outputs. One of these, the F_6 vectors, is not needed for the subsequent computation and indeed is typically not computed in lattice implementations. The other outputs produced are the E_5 vectors which will *not* be stored as part of the state at the next horocycle but which do show up as inputs to the layer 5 blocks.

We have also included node indices in part of Figure 5 to make it easier to connect the computational structure of the figure with the computations described in (3.23)–(3.27). For example, the lower left-hand portion of layer 1 (distinguished by shaded circles and squares) corresponds to the pair of computations corresponding to (3.27) evaluated at t and at $t\delta$. Also, at higher layers, we encounter vector error processes,

and as we have seen, these vectors are not distinct or, in fact, needed at all nodes. For example, consider the portion of the layer 5 computations indicated by shaded squares. This describes the computation of (3.25) for $n = 5$, which requires a single E_5 input at node t, a single stored F_4–vector at $t\gamma^{-1}$ and produces one F_5 vector at node t and *two* E_4 vectors at t and $t\delta^{(2)}$. In this case, as we have pointed out, a single F_4–vector needs to be stored for the pair of square nodes connected by the solid bar in the figure. We have indicated its index, $t\gamma^{-1}$, in the center of the bar. Similarly, the index, t, of the single E_5 and F_5 vectors is indicated in the center of the lower solid bar, while the indices, t and $t\delta^{(2)}$, for the two E_4–vectors are indicated above the appropriate portion of the solid bar. Note that the apparent redundancies, indicated by the shaded bars in Figure 4, are not present in Figure 5, as in this figure we have shown just those variables required to be stored and computed from horocycle to horocycle

As we will see, understanding the structure of the filter described in Figures 3–5 greatly facilitates our analysis of stability.

3.2 Levinson Recursions for the Normalized Residuals

The prediction errors $E_{t,n}$ and $F_{t,n}$ do not quite define isotropic processes. In particular the components of these vectors representing prediction error vectors at a *set* of nodes are correlated. Furthermore for n even we have seen that $E_{t,n}$ and $E_{t\delta^{(\frac{n}{2})},n-1}$ are correlated (see (3.15)). These observations provide the motivation for the normalized recursions developed in this section. In this development we use the superscript * to denote normalized versions of random vectors. Specifically $X^* = \Sigma_x^{-1/2} X$ where Σ_x is the covariance of X and $\Sigma_x^{1/2}$ is its symmetric, positive definite square root.

We now can state and prove the following:

Theorem 3.3 *For n odd the covariance matrix Σ_n defined in (3.13) is invertible if and only if $-1 < k_n < 1$. For n even, Σ_n as defined in (3.15), is invertible if and only if $-1/2 < k_n < 1$. Under these conditions the whitening recursions of*

Theorem 3.1 can be normalized, yielding the following recursions for the normalized residuals:

For n even

$$\begin{pmatrix} E_{t,n} \\ E_{t\delta^{(\frac{n}{2})},n} \end{pmatrix}^* = \Theta(k_n)\left[\begin{pmatrix} E^*_{t,n-1} \\ E^*_{t\delta^{(\frac{n}{2})},n-1} \end{pmatrix} - k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} F^*_{t\gamma^{-1},n-1}\right] \quad (3.30)$$

$$F^*_{t,n} = \Theta(k_n)\left[\begin{pmatrix} E^*_{t\delta^{(\frac{n}{2})},n-1} \\ F^*_{t\gamma^{-1},n-1} \end{pmatrix} - k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} E^*_{t,n-1}\right] \quad (3.31)$$

where $\Theta^{-1}(k_n)$ is the matrix square root satisfying[5]

$$\Theta^{-2}(k_n) = \begin{pmatrix} I - k_n^2 U_* & (k_n - k_n^2)U_* \\ (k_n - k_n^2)U_* & I - k_n^2 U_* \end{pmatrix} \quad (3.32)$$

For n odd, $n > 1$

$$E^*_{t,n} = \Theta(k_n)\left[\begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(\frac{n-1}{2})},n-1} \end{pmatrix}^* - k_n U_* F^*_{t\gamma^{-1},n-1}\right] \quad (3.33)$$

$$F^*_{t,n} = \Theta(k_n)\left[F^*_{t\gamma^{-1},n-1} - k_n U_* \begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(\frac{n-1}{2})},n-1} \end{pmatrix}^*\right] \quad (3.34)$$

where

$$\Theta^{-2}(k_n) = I - k_n^2 U_* \quad (3.35)$$

for $n = 1$

$$E^*_{t,1} = \frac{1}{\sqrt{1-k_1^2}}\left(Y^*_t - k_1 Y^*_{t\gamma^{-1}}\right) \quad (3.36)$$

$$F^*_{t,1} = \frac{1}{\sqrt{1-k_1^2}}\left(Y^*_{t\gamma^{-1}} - k_1 Y^*_t\right) \quad (3.37)$$

[5]Again to be precise we should write $\Theta(k_n, n)$ rather than $\Theta(k_n)$. For simplicity we use the less cumbersome notation.

Remark: Note that for n even we normalize $E_{t,n}$ and $E_{t\delta(\frac{n}{2}),n}$ *together* as one vector, while for n odd, $E_{t,n}$ is normalized individually. This is consistent with the nature of their statistics as described in (3.15)–(3.19) and with the fact that for n even $\dim F_{t,n} = 2\dim E_{t,n}$, while for n odd $\dim F_{t,n} = \dim E_{t,n}$.

Proof: Let us first derive (3.30)–(3.37) assuming the invertibility of Σ_n for each n. This result is a relatively straightforward computation given (3.11)–(3.19). For n even we begin with (3.7) and (3.21) and premultiply each by

$$\operatorname{diag}\left(\Sigma_{n-1}^{-1/2} \quad , \quad \Sigma_{n-1}^{-1/2} \right)$$

Since 1_* is an eigenvector of Σ_{n-1}, Σ_{n-1} and therefore $\Sigma_{n-1}^{-1/2}$ commute with U_*. This immediately yields (3.30) and (3.31) where the matrix $\Theta(k_n)$ is simply the inverse of the square root of the covariance of the term in brackets in (3.30) and in (3.31) (the equality of these covariances follows from (3.15)). Equation (3.32) then follows from (3.11) and (3.15). The case of n odd involves an analogous set of steps, and the $n = 1$ case is immediate.

The preceding analysis provides us both with the conditions for the invertibility of Σ_n and with a recursive procedure for calculating $\Sigma_n^{-1/2}$ (see Appendix D of [1] for an alternate efficient procedure). For n even we have

$$\Sigma_n^{-1/2} = \Theta(k_n) \operatorname{diag}\left(\Sigma_{n-1}^{-1/2} \quad , \quad \Sigma_{n-1}^{-1/2} \right) \tag{3.38}$$

while for n odd, $n > 1$

$$\Sigma_n^{-1/2} = \Theta(k_n)\Sigma_{n-1}^{-1/2} \tag{3.39}$$

and for $n = 1$

$$\Sigma_1^{-1/2} = \left[(1 - k_1^2)r_0\right]^{-1/2} \tag{3.40}$$

Note first that from (3.40) we see that $|k_1|$ must be less than 1 for $\Sigma_1^{-1/2}$ to exist. For $n > 1$ and odd, note that the only nonunity eigenvalues of $I - k_n^2 U_*$ is $1 - k_n^2$, and thus $\Theta(k_n)$ exists for n odd if and only if $|k_n| < 1$. Also in this case we can readily compute $\Theta(k_n)$ using the following formula. For any $k > -1$

$$(I + kU_*)^{-\frac{1}{2}} = I + \left(\frac{1}{\sqrt{1+k}} - 1 \right) U_* \tag{3.41}$$

For n even, we make use of the result that for S and T symmetric

$$\begin{pmatrix} S & T \\ T & S \end{pmatrix}^{-1/2} = \frac{1}{2}\begin{pmatrix} X+Y & X-Y \\ X-Y & X+Y \end{pmatrix} \tag{3.42}$$

where

$$\begin{aligned} X &= (S+T)^{-1/2} \\ Y &= (S-T)^{-1/2} \end{aligned} \tag{3.43}$$

Using (3.42), (3.43) we see from (3.32) that to calculate $\Theta(k_n)$ for n even we must calculate

$$\left(I + \left(k_n - 2k_n^2\right) U_*\right)^{-1/2}$$

and

$$(I - k_n U_*)^{-1/2}$$

which exist if and only if $-1/2 < k_n < 1$, completing our proof.

If $k_n = -1/2$ or 1 for n even or $k_n = \pm 1$ for n odd, the resulting error processes are not full rank. This is the simplest example of a singular process, for which perfect prediction of a linear combination of Y's on a given horocycle can be obtained using only a *finite* set of values of Y on "past" horocycles. In Section 4 we will characterize the full class of singular processes in terms of its infinite reflection coefficent sequence.

Now that we have a normalized form for the residual vectors, we can also describe the normalized version of the modeling filters which provide the basis for generating isotropic Y_t's specified by a finite number of reflection coefficients and driven by white noise

Theorem 3.4 *A normalized modeling filter for the isotropic process Y_t exists if and only if $-1 < k_n < 1$ for n odd and $-1/2 < k_n < 1$ for n even. In this case this filter has the following form. For n even we have*

$$\begin{bmatrix} E^*_{t,n-1} \\ \\ F^*_{t,n} \end{bmatrix} = \Sigma(k_n) \begin{bmatrix} \begin{pmatrix} E_{t,n} \\ E_{t\delta^{(\frac{n}{2})},n} \end{pmatrix}^* \\ F^*_{t\gamma^{-1},n-1} \end{bmatrix} \tag{3.44}$$

where[6]

$$\Sigma(k_n) \triangleq \begin{pmatrix} I + a(k_n)U_* & b(k_n)U_* & k_nU_* \\ -\frac{k_n}{2}U_* & I + c(k_n)U_* & b(k_n)U_* \\ d(k_n)U_* & -\frac{k_n}{2}U_* & I + a(k_n)U_* \end{pmatrix} \tag{3.45}$$

with

$$a(k) = \frac{\sqrt{1+2k}+1}{2}\sqrt{1-k} - 1 \tag{3.46}$$

$$b(k) = \frac{\sqrt{1+2k}-1}{2}\sqrt{1-k} \tag{3.47}$$

$$c(k) = \frac{\sqrt{1+2k}-(1+k)}{2} \tag{3.48}$$

$$d(k) = -c(k) - k \tag{3.49}$$

The matrix $\Sigma(k_n)$ *is referred to as the scattering matrix, and it satisfies*

$$\Sigma(k_n)\Sigma^T(k_n) = I \tag{3.50}$$

For n *odd,* $n \neq 1$

$$\left[\begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(\frac{n-1}{2})},n-1} \end{pmatrix}^* \right] = \Sigma(k_n) \begin{bmatrix} E^*_{t,n} \\ F^*_{t\gamma^{-1},n-1} \end{bmatrix} \tag{3.51}$$

$$F^*_{t,n}$$

where the scattering matrix

$$\Sigma(k_n) \triangleq \begin{pmatrix} (I - k_n^2 U_*)^{1/2} & k_n U_* \\ -k_n U_* & (I - k_n^2 U_*)^{1/2} \end{pmatrix} \tag{3.52}$$

satisfies

$$\Sigma(k_n)\Sigma^T(k_n) = I \tag{3.53}$$

[6]Again we abuse notation and write $\Sigma(k_n)$ rather than $\Sigma(k_n, n)$

For $n = 1$:

$$\begin{pmatrix} Y_t^* \\ F_{t,1}^* \end{pmatrix} = \Sigma(k_1) \begin{pmatrix} E_{t,1}^* \\ Y_{t\gamma^{-1}}^* \end{pmatrix} \tag{3.54}$$

and

$$\Sigma(k_1) = \begin{pmatrix} \sqrt{1-k_1^2} & k_1 \\ -k_1 & \sqrt{1-k_1^2} \end{pmatrix} \tag{3.55}$$

also satisfies

$$\Sigma(k_1)\Sigma^T(k_1) = I \tag{3.56}$$

Proof: We begin by solving (3.30) for $\left(E_{t,n-1}^{*T}\ E_{t\delta^{(\frac{n}{2})},n-1}^{*T}\right)^T$ then by substituting this into (3.31) we obtain

$$\left[\begin{pmatrix} E_{t,n-1}^* \\ E_{t\delta^{(\frac{n}{2})},n-1}^* \end{pmatrix}\right] = \hat{\Sigma}(k_n) \left[\begin{pmatrix} E_{t,n} \\ E_{t\delta^{(\frac{n}{2})},n} \end{pmatrix}^* \right] \tag{3.57}$$
$$\left[F_{t,n}^*\right] \qquad \left[F_{t\gamma^{-1},n-1}^*\right]$$

where

$$\hat{\Sigma}(k_n) \triangleq \begin{bmatrix} \Theta^{-1}(k_n) & k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} \\ \Theta(k_n) \begin{pmatrix} -k_n U_* & I \\ -k_n U_* & 0 \end{pmatrix} \Theta^{-1}(k_n) & \Theta(k_n) \begin{pmatrix} (k_n - k_n^2)U_* \\ I - k_n^2 U_* \end{pmatrix} \end{bmatrix} \tag{3.58}$$

To obtain the desired relation, we simply drop the calculation of $E_{t\delta^{(\frac{n}{2})},n-1}^*$ from (3.57). To do this explicitly we consider $\hat{\Sigma}(k_n)$ as a matrix with three block-columns and four block-rows (one each for $E_{t,n-1}^*$ and $E_{t\delta^{(\frac{n}{2})},n-1}^*$ and two for $F_{t,n}^*$). Thus what we wish to do is to drop the second block-row. A careful calculation using the relations derived previously yields (3.45)–(3.49). That $\Sigma(k_n)$ satisfies (3.50) follows immediately from the fact that the vectors on both sides of (3.44) have identity covariances. The result for n odd, $n > 1$ is obtained in a similar fashion, and the case of $n = 1$ is immediate.

4 Reflection Coefficients and the Properties of Processes and Models

The analysis in [1] and in the preceding sections provides us with a framework in which we can say a great deal about stochastic processes and dynamic systems on trees. In the first subsection we provide a complete characterization of isotropic autoregressive processes, and is Subsection 4.2 we characterize purely nondeterministic processes. In Subsection 4.3 we relate the stability of the lattice models on trees to the reflection coefficients, while in Subsection 4.4 we show that all lattice filters with appropriately-constrained reflection coefficients yield AR processes, showing the one-to-one correspondence between these filters and processes. In each case there are similarities to the analysis for stationary time series. However, the more complex structure of the dyadic tree leads to some important and substantive differences.

4.1 Characterization of Autoregressive Processes.

A well-known and essentially trivial result for time series is that if Y_t is a pth order autoregressive process, then the reflection coefficients $k_n = 0$ for $n \geq p+1$. Furthermore the pth-order forward and backward prediction errors—which are also identical to the nth order prediction errors for $n \geq p+1$—form white noise sequences. The following result, whihc states the counterpart of this result for isotropic processes on trees, requires some prefatory comment. Specifically, thanks to the vector nature of our models—i.e. the fact that a *group* of Y's on a given horocycle are generated together from a group of the E's—the prediction error processes whose whiteness we consider consist of *sampled* versions of the (normalized) E and F processes, with one "sample" taken per "group." In particular from our discussion at the end of Section 3.1 and from the definition (2.8), (2.9) of $E_{t,n}$, we find that the components of $E_{t,n}$ and $E_{tw,n}$ are permutations of one another if $w \asymp 0$ and $|w| \leq n-1$. Thus we need only consider one of these vectors for each group on each horocycle. Note

that this means that we are choosing only one out of $2^{[\frac{n-1}{2}]}$ error vectors, but each vector is exactly of dimension $2^{[\frac{n-1}{2}]}$ so that we do have the correct number of total degrees of freedom—one per node on the tree.

Turning to the backward residuals, we find from the discussion in Section 3.1 and the definition (2.6), (2.7) of $F_{t,n}$ that the components of $F_{t,n}$ and $F_{tw,n}$ for $w \succ 0$ and $|w| \leq n-1$ are permutations of one another. On the other hand, as pointed out (for $w = \delta^{(\frac{n}{2})}$) in Section 3.1, if n is even, so that it is possible to find $w \succ 0$ with $|w| = n$, $F_{t,n}$ and $F_{tw,n}$ do *not* have identical sets of components. Furthermore, it is easily checked that these vectors are *not* uncorrelated. However, as is also pointed out in Section 3.1, the signals $F_{t\gamma^{-1},n}$ and $F_{tw\gamma^{-1},n}$ *do* have identical component sets, and it is only these "delayed" signals that play a role in the modeling filter. Thus for our purposes here we need choose only one vector from the set $\{F_{tw} | w \succ 0, |w| \leq n\}$. In this case we are choosing one $2^{[\frac{n}{2}]}$–dimensional vector from a set of $2^{[\frac{n}{2}]}$ such vectors, again producing the correct number of degrees of freedom.

Finally, as we have noted in Section 3.2, it is necessary to normalize the prediction error processes. For the backward prediction errors, this simply means that we will consider the $F^*_{t,n}$ rather than the $F_{t,n}$. Similarly for n odd we consider the $E^*_{t,n}$. However for n even our normalization involves the *combined* normalization of $E_{t,n}$ and $E_{t\delta^{(\frac{n}{2})},n}$ (e.g. referring to Figure 3, the two inputs $E_{t,6}$ and $E_{t\delta^{(3)},6}$ are normalized together). Thus for n even, instead of choosing one vector from $\{E_{tw,n} \ : \ w \succ 0, |w| \leq n-1\}$ $\{E_{t\delta^{(\frac{n}{2})}w,n} \ : \ w \succ 0, |w| \leq n-1\}$ we choose *one* vector (of twice the dimension) from

$$\left\{ \begin{pmatrix} E_{tw,n} \\ E_{t\delta^{(\frac{n}{2})}w,n} \end{pmatrix}^* | w \succ 0, |w| \leq n-1 \right\}$$

Proposition 4.1 *If Y_t is an AR(p) isotropic process, then the reflection coefficients $k_n = 0$ for $n \geq p+1$. Furthermore, the forward and backward normalized prediction error vectors of order p and greater form standard white noise processes (i.e. with unity covariance). More precisely, let t_0 be an arbitrary node on the tree, and*

consider an infinite sequence of predecessors and successors to t_0*:*

$$T = \{t_0\gamma^{-k} | k \geq 0\} \bigcup \{t_0\alpha^k | k \geq 1\}$$

Then for any $n \geq p$*, the set of backward prediction error vectors*

$$\{F^*_{s\delta(j),n} | s \in T, j > \frac{n}{2}\}$$

forms a standard white noise process. Similarly, for $n \geq p$ *and odd, the set*

$$\{E^*_{s\delta(j),n} | s \in T, j > \frac{n}{2}\}$$

forms a standard white noise process, while for $n \geq p$ *and even, the set*

$$\left\{ \begin{pmatrix} E_{s\delta(j),n} \\ E_{s\delta(j)\delta(\frac{n}{2}),n} \end{pmatrix}^* \middle| s \in T, j > \frac{n}{2} \right\}$$

forms a standard white noise process.

The construction of T and the choices of points forming the sets of prediction errors in Proposition 4.1 represents one particular way of choosing one prediction error vector from each of the sets described before the statement of the Proposition.

Proof of Proposition 4.1: We focus explicitly on the E's, as an analogous proof holds for the F's. Note first that, thanks to the normalization, all of the E^* variables do have unity covariance. Also, thanks to the sampling done in forming the E^* sets, it is straightforward to check that the whiteness will be proven if we can show that for $n \geq p$ (and *either* even or odd) the unnormalized prediction error $E_{t,n}$ is uncorrelated with $E_{tw,n}$ (denoted $E_{t,n} \perp E_{tw,n}$) for $w \prec 0$ and for $w \succ 0, |w| > n$.

Showing that this is true for $w \prec 0$ is essentially the same as the proof in the time series case. Specifically, if $k_n = 0$ for $n \geq p$, then,

$$\begin{aligned} E_{t,2m} &= E_{t,2m-1} \text{ if } 2m > p \\ E_{t,2m+1} &= \begin{bmatrix} E_{t,2m} \\ E_{t\delta(m),2m} \end{bmatrix} \text{ if } 2m \geq p \end{aligned} \tag{4.1}$$

so that, $n \geq p$,

$$E_{t,n} \perp \mathcal{Y}_{t\gamma^{-1},\infty} \tag{4.2}$$

by definition of the forward prediction errors. Hence from (2.9) we see that for $n \geq p$,

$$E_{t,n} \perp E_{tw,n} \text{ for } w \prec 0 \tag{4.3}$$

Hence it remains to prove that

$$E_{t,n} \perp E_{t\delta(j),n} \quad \text{for} \quad j > \frac{n}{2} \tag{4.4}$$

This proof, which involves the construction of isometries much as in several of the proofs in [1], is sketched in Appendix C.

For a time series model the constraint of causality severely restricts the support of its impulse response. For example any AR time series model has an AR impulse response whose support is the nonnegative integers. For processes on trees, however, there is considerable flexibility in the possible choice of support for a causal impulse response. However, as the following states, the constraints of isotropy allow us to determine precisely the support for AR models.

Proposition 4.2 *Let Y_t be an AR(p) isotropic process. Let us write the formal power series P_p defined in (2.24) as*

$$P_p = \sum_{\substack{w \in \mathcal{L} \\ w \preceq 0}} p_w \cdot w \tag{4.5}$$

If $p = 0$, $p_w = 0$ if $w \neq 0$. If $p = 1$, $p_w = 0$ unless $w = \gamma^{-k}$ for some $k \geq 0$. If $p \geq 2$, then $p_w = 0$ for all words of the form $w = \gamma^{-k}\delta w_{\alpha\beta}$ with

$$w_{\alpha\beta} \in \{\alpha, \beta\}^* \quad \text{and} \quad |w_{\alpha\beta}| > \left[\frac{p}{2}\right] - 1 \tag{4.6}$$

In other words, P_p has its support in a cylinder of radius $\left[\frac{p}{2}\right]$ around the path $\{\gamma^{-k}\}$ toward $-\infty$. From this we also have that the modeling filter of an AR(p) process has its support in the same cylinder of radius $\left[\frac{p}{2}\right]$ around $[t, -\infty) = \{t\gamma^{-k} | k \geq 0\}$. Conversely, any process such that the modeling filter has its support contained in the cylinder of radius $\left[\frac{p}{2}\right]$ is necessarily an AR(p) process.

Comment: the proof of this result is straightforward, although tedious, and is left to the reader. Concerning point 2, Figure 6 illustrates the cylinders for low order AR processes. Note that Proposition 4.2 is a generalization of the result in Appendix A of [1] which states that if an isotropic process has its support concentrated on $[t, -\infty)$, then it is necessarily AR(1).

4.2 Characterization of Regular (or Purely Nondeterministic) Processes.

Definition 4.1 *We shall say that an isotropic process Y_t is regular or purely nondeterministic if*

$$\sigma^2 > 0 \tag{4.7}$$

holds, where

$$\sigma^2 \triangleq \inf \| \left(\sum_{w \asymp 0} \mu_w Y_{tw} \right) - E\left(\left(\sum_{w \asymp 0} \mu_w Y_{tw} \right) | \mathcal{Y}_{t\gamma^{-1},\infty} \right) \|^2 \tag{4.8}$$

and the infimum ranges over all collections of scalars $(\mu_w)_{w \asymp 0}$ where only finitely many of the μ_w are nonzero and the condition $\sum \mu_w^2 = 1$ is satisfied.

In other words, no nonzero linear combination of the values of Y_t on any given horocycle can be predicted exactly with the aid of knowledge of Y in the strict past, $\mathcal{Y}_{t\gamma^{-1},\infty}$ and the associated prediction error is uniformly bounded from below. We shall now characterize regular processes in a fairly simple way using reflection coefficients.

Theorem 4.1 *(i) The following formulae hold for every isotropic process:*

$$\sigma^2 = \liminf_{n \to \infty} \lambda_{\inf}(\Sigma_{2n+1}) \tag{4.9}$$

$$\lambda_{\inf}(\Sigma_{2n+1}) = r_0(1 - k_1^2) \prod_{p=1}^{n} (1 - k_{2p+1}^2) \min\left\{1 + k_{2p} - 2k_{2p}^2, 1 - k_{2p}\right\} \tag{4.10}$$

where $\lambda_{\inf}(A)$ denotes the smallest eigenvalue of the matrix A, and Σ_{2n+1} is defined in (3.13).

(ii) An isotropic process Y_t is regular if and only if its reflection coefficient sequence is such that $|k_{2n+1}| < 1, -\frac{1}{2} < k_{2n} < +1$, and, furthermore,

$$\sum_{n=1}^{\infty} \left(k_{2n-1}^2 + |k_{2n}|\right) < \infty \tag{4.11}$$

Comment: The corresponding characterization of regular processes in the case of time series is (cf. for instance [2]):

$$|k_n| < 1 \ \forall n, \ \sum_{n=1}^{\infty} k_n^2 < \infty \tag{4.12}$$

Proof: Note first that the singularity of the process if $|k_{2n+1}| = 1$ or if $k_n = -1/2$ or 1 follows directly from the resulting degeneracy of the prediction error covariance (Theorem 3.3). Condition (4.11) is an immediate consequence of point (i), since for k small $\min(1-k, 1+k-2k^2) \sim 1 - |k|$. Thus we shall only prove (i). First, let us prove (4.9) by showing that σ^2 is both $\geq$ and $\leq$ the right hand side of (4.9). To every $(\mu_w)_{w \asymp 0}$ as in Definition 4.1 we associate a sequence of vectors (M_n) of increasing dimension. Specifically, we begin by forming an infinite-dimensional vector by ordering the μ_w according to the ordering on the $w \asymp 0$ defined in Section 3.1 of [1]. For each n we then take the vector $\tilde{M}_n$ to be the truncated version of this infinite vector by keeping only the initial segment consisting of those μ_w's such that w is involved in the definition of $E_{t,2n+1}$. We then set $M_n = \frac{\tilde{M}_n}{\|\tilde{M}_n\|}$ if $\tilde{M}_n \neq 0$, and equal to some arbitrary unit vector otherwise (here, $\| \ldots \|$ denotes the usual Euclidian norm).

We obviously have

$$\tilde{M}_n = M_n \text{ for } n \text{ large enough} \tag{4.13}$$

Hence, thanks to the limit theorem for square integrable martingales ([5], [6]), we can write, for the considered family (μ_w):

$$\| \sum_{w \asymp 0} \mu_w Y_{tw} - E\left(\sum_{w \asymp 0} \mu_w Y_{tw} | \mathcal{Y}_{t\gamma^{-1},\infty}\right) \|^2$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \| \sum_{w\asymp 0} \mu_w Y_{tw} - E\left(\sum_{w\asymp 0} \mu_w Y_{tw} | \mathcal{Y}_{t\gamma^{-1},2n} \right) \|^2 \\
&= \lim_{n\to\infty} M_n^T \Sigma_{2n+1} M_n \\
&\geq \liminf_{n\to\infty} \lambda_{\inf}(\Sigma_{2n+1}) \qquad (4.14)
\end{aligned}
$$

where the second equality uses (4.13), and the inequality is due to the fact that M_n is a unit vector. Since the last expression in (4.14) does not involve the considered family (μ_w), we immediately get the inequality $\geq$ in (4.9). Now, fix $\epsilon > 0$ and select n_o large enough so that $\lambda_{\inf}(\Sigma_{2n_o+1}) - \epsilon$ is smaller than the right-hand side of (4.9). Then, take for M_{n_o} a unit eigenvector of Σ_{2n_o+1} associated with its smallest eigenvalue. We then obtain following inequalities which, since ϵ is arbitrary, yields the inequality $\leq$ in (4.9):

$$
\begin{aligned}
&\liminf_{n\to\infty} \lambda_{\inf}(\Sigma_{2n+1}) + \epsilon \\
&\geq \lambda_{\inf}(\Sigma_{2n_o+1}) \\
&= M_{n_o}^T \Sigma_{2n_o+1} M_{n_o} \\
&\geq \| \sum_{w\asymp 0} \mu_w^o Y_{tw} - E\left(\sum_{w\asymp 0} \mu_w^o Y_{tw} | \mathcal{Y}_{t\gamma^{-1},\infty} \right) \|^2 \\
&\geq \sigma^2
\end{aligned}
$$

where (μ_w^o) is the family associated with M_{n_o}.

It remains to prove (4.10). Using (3.38)–(3.40), we can write

$$
\Sigma_{2n+1}^{-\frac{1}{2}} = \Theta(k_{2n+1})\Theta(k_{2n}) \begin{pmatrix} \Sigma_{2n-1}^{-\frac{1}{2}} & 0 \\ 0 & \Sigma_{2n-1}^{-\frac{1}{2}} \end{pmatrix}
$$

But the three matrices on the right handside of this formula all have the Haar system as eigenvectors (cf. (4.19) of [1]). Hence we can diagonalize all of these matrices simultaneously:

$$
\Lambda\left(\Sigma_{2n+1}^{-\frac{1}{2}}\right) = \Lambda\left(\Theta_{2n+1}^*\right) \Lambda(\Theta_{2n}^*) \begin{pmatrix} \Lambda\left(\Sigma_{2n-1}^{-\frac{1}{2}}\right) & 0 \\ 0 & \Lambda\left(\Sigma_{2n-1}^{-\frac{1}{2}}\right) \end{pmatrix}
$$

holds, where $\Lambda(A)$ denotes the diagonal matrix of the eigenvalues of A. Using the definition of $\Theta^{-1}(k_n)$ in (3.32), (3.35), we can deduce that

$$\Lambda(\Sigma_{2n+1}) =$$
$$\mathrm{diag}(1-k_{2n+1}^2, 1...1)\ \mathrm{diag}(1+k_{2n}-2k_{2n}^2, 1-k_{2n}, 1...1) \begin{pmatrix} \Lambda(\Sigma_{2n-1}) & 0 \\ 0 & \Lambda(\Sigma_{2n-1}) \end{pmatrix}$$

so that, by expanding the product and using (3.36), we finally get (4.10). This finishes the proof of the theorem.

Note that the condition (4.11) is much more easily violated by a valid reflection coefficient sequence than the corresponding expression (4.12) for time series, pointing to the fact that there are apparently a far richer class of singular processes on trees than on the real line. This is apparently related to the characterization of spectral measures for isotropic processes and to the large size of the boundary of the dyadic tree (see the comments concerning (2.33) in [1]).

4.3 A stability criterion

A well-known result for all-pole lattice filters is that such a filter is stable if and only if all of the reflection coefficients have magnitude less than 1. In this section we state and prove Theorem 4.2, which is the counterpart of this result for the lattice filters introduced in this paper. Before stating this result, let us clarify what we mean by "stability". Figures 3–5 depict (for a 6th–order example) the structure of the unnormalized filter. This filter describes how the computation of Y_t propagates from horocycle to horocycle, with $E_{t,n}$ (for n odd) or $(E_{t,n}, E_{t\delta(\frac{n}{2}),n})$ (for n even) as input and the corresponding block of Y's on the same horocycle as output. It is the stability of *this* filter that we wish to study.

Theorem 4.2 *Under the conditions*

$$-1 < k_n < 1 \quad n \text{ odd} \quad 1 \le n \le N \tag{4.15}$$
$$-\frac{1}{2} < k_n < 1 \quad n \text{ even} \quad 1 \le n \le N \tag{4.16}$$

the Nth-order unnormalized modeling filter specified by (3.23)-(3.27) is stable, so that a bounded input $E_{t,N}$ (for N odd) or $(E_{t,N}, E_{t\delta^{(\frac{N}{2})},N})$ (for N even) yields a bounded output Y_t. Similarly the normalized modeling filter specified in theorem 3.4 is also stable under these conditions so that a bounded input $E^*_{t,N}$ (for N odd) or $(E_{t,N}, E_{t\delta^{\frac{N}{2}},N})^*$ (for N even) yields a bounded output Y^*_t.

Proof: Let us first show that by taking advantage of the structure of the filter computations we can simplify the required analysis and can, in fact, reduce it to a question of stability analysis for a standard temporal system. To begin, in Figure 7 we have depicted one of the two parallel computations depicted in Figure 5, where we have used notation that emphasizes the sequential nature of the computations. Here the indices "m" and "$m-1$" index horocycles so that the "$m-1$" quantities are stored and the "m" quantities are computed from the input $(E_{61}(m), E_{62}(m))$ which is distinguished by a solid box at level 6 in the figure (note that the reverse-going output from this final level, $F_{61}(m)$ is distinguished by a dashed box). The subscripts for the signals in Figure 7 code the various error and output processes at each level. The first subscript for the E– and F–vectors indicate the order of the error vector, while the second subscript (and the only subscript for the Y's) indexes the vectors along a segment of a horocycle. The precise correspondence between the normalized version of quantities in Figure 7 and those in Figure 5 can be directly determined by matching up signals and node indices in Figure 5 with signals and horocycles index (m and $m-1$) in Figure 7. For example

$$\begin{aligned} Y_t, Y_{t\delta}, Y_{t\delta^{(2)}}, Y_{t\delta^{(2)}\delta} &\longleftrightarrow Y_1(m), Y_2(m), Y_3(m), Y_4(m) \\ Y_{t\gamma^{-1}}, Y_{t\delta^{(2)}\gamma^{-1}} &\longleftrightarrow Y_1(m-1), Y_2(m-1) \\ F_{t,3}, F_{t\delta^{(2)},3} &\longleftrightarrow F_{31}(m), F_{32}(m) \\ E_{t,4}, E_{t\delta^{(2)},4} &\longleftrightarrow (E_{41}(m), E_{42}(m)) \end{aligned}$$

As we emphasized in Section 3.1, and as illustrated graphically in Figures 3–5 and 7, each stage of the computation is pyramidal in structure. For example, the state of a set of nodes at a given horocycle, together with the inputs, provide

the state at *two* descendent sets of nodes at the next horocycle. Since the computations in generating each of these descendent sets are identical in structure, we need follow only one of these paths in order to examine stability. For example, for our 6th-order example, we need only establish stability of the dynamics from input $(E_{61}(m), E_{62}(m))$ to $(Y_1(m), Y_2(m), Y_3(m), Y_4(m))$. However we can take this considerably farther. In particular, because of the pyramidal symmetries, we need only consider the stability of the map from $(E_{61}(m), E_{62}(m))$ to $Y_1(m)$ as the structure of the map to $Y_2(m), Y_3(m)$, and $Y_4(m)$ are identical. More generally, starting from any node t_0 on the tree, we need only consider the stability of the dynamics involved in generating $\{Y_{t_0\alpha^n} | n \geq 0\}$, since the dynamics for any other path from horocycle to horocycle has identical structure.

Using the notation of Figure 7, we now see that we must examine the stability of the system illustrated in Figure 8 for the 6th order case, where the small solid squares now denote standard z^{-1} operations (i.e. $z^{-1}x(m) = x(m-1)$). Here the $S(k_n)$ matrices are exactly as defined in Theorem 3.2. We can now apply standard time domain methods to this system.[7]

Note first that under conditions (4.15)-(4.16), the $\Theta(k_n)$ matrices defined in Theorem 3.3 and the covariance matrices Σ_n are invertible so that we can equivalently study the stability of the normalized form of the modeling filter. Note also that checking that a system function $H(z)$ has all its poles strictly inside the unit circle is equivalent to checking the same condition for the system function $H(z^2)$. Thus to test for stability we can modify the system of Figure 8 by adding a unit delay in every left-to-right-going path, and by replacing the $S(k_i)$ blocks by the scattering matrices $\Sigma(k_i)$ of Theorem 3.4. For example, in the 6th order case we can equivalently check the stablity of the system in Figure 9. Recall that for an Nth-order filter we proved in theorem 3.4 that

$$\Sigma(k_i)^T \Sigma(k_i) = I, \qquad i = 1, \ldots, N \tag{4.17}$$

[7]We would like to acknowledge Bernard C. Levy for suggesting this line of proof

for any set of coefficients $k_1, \ldots, k_N$ that are reflection coefficients of some isotropic process. But the entries of the matrices $\Sigma(k_i)^T\Sigma(k_i) - I, i = 1, \ldots, N$ are rational functions of the k_n's that have no poles inside the domain specified by the conditions (4.15), (4.16). Hence we may use the lemma C.2 of appendix C to extend the property (4.17) to *the whole domain specified by the conditions* (4.15), (4.16).

Using (4.17) and the notation of Figure 9 we have that

$$\|\xi_i(m)\|^2 + \|\eta_i(m)\|^2 = \|\xi_{i+1}(m-1)\|^2 + \|\eta_{i-1}(m-1)\|^2 \tag{4.18}$$

where we have the boundary conditions

$$\xi_{N+1}(m) = u(m) \tag{4.19}$$

$$\eta_0(m) = \xi_1(m-1) \tag{4.20}$$

To study stability we set $u(m) \equiv 0$ and define the following positive-definite function of the state of the system

$$V(m) = \sum_{i=1}^{N} \|\xi_i(m)\|^2 + \|\eta_{i-1}(m)\|^2 \tag{4.21}$$

Then from (4.18)–(4.21) we obtain

$$V(m) - V(m-1) = -\|\eta_N(m)\|^2 \tag{4.22}$$

It can be readily checked that the system is observable from $\eta_N(m)$, as long as (4.15) and (4.16) are satisfied, so that $V(m)$ is a Lyapunov function proving asymptotic stability.

4.4 Every finite family of reflection coefficients defines an isotropic AR process

Our analysis to this point has shown how to construct a sequence of reflection coefficients $\{k_n\}$ from an isotropic covariance sequence $\{r_n\}$. Furthermore we have seen that the $\{k_n\}$'s have particular bounds and that if $\{r_n\}$ comes from an AR(p) process, only the first p of the reflection coefficients are nonzero. The following

result states that the converse holds, i.e. that any finite k_n sequence satisfying the required constraints corresponds to a unique AR covariance sequence. This result substantiates our previous statement that the reflection coefficients provide a good parameterization of AR processes.

Theorem 4.3 *Given a finite sequence of reflection coefficients k_n, $1 \leq n \leq p$ such that*

$$\begin{cases} -\frac{1}{2} < k_n < 1 & \text{for } n \text{ even} \\ -1 < k_n < 1 & \text{for } n \text{ odd} \end{cases} \tag{4.23}$$

there exists a unique isotropic covariance sequence which has as its reflection coefficient sequence the given k_n followed by all zeroes.

Proof: Consider the modeling filter of order p specified by the given set of reflection coefficients. What we must show is that the output of this filter, y_t, is well defined (i.e. has finite covariance) and isotropic when the input is a standard white noise process. That it is well-defined follows from the stability result in Theorem 4.2. Thus we need only show that y_t is isotropic. More specifically, let (s,t) and (s',t') be any two pairs of points such that $d(s,t) = d(s',t')$. The theorem will be proved if we can show that the function

$$\Phi : K = (k_n)_{1 \leq n \leq p} \longrightarrow E(y_t y_s) - E(y_{t'} y_{s'}) \tag{4.24}$$

is identically zero for all k_n's satisfying the condition (4.23). But the formulae for the modeling filter (Theorem 3.2) show that Φ is a rational function of K which is analytic inside the domain specified by the conditions (4.23). Also Φ is identically zero for all sequences K arising from valid isotropic covariances via the Schur recursions (2.26)–(2.31). Then the theorem is an immediate consequence of the Lemma C.2 of Appendix C.

5 Conclusion

In [1] and 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. However, as we have developed, the generalization of lattice structures provides a much better parametrization of AR processes in terms of a sequence of reflection coefficients.

In developing this theory we have seen that it is necessary to consider forward and backward prediction errors of dimension that grows geometrically with filter order. Nevertheless, thanks to isotropy, only one reflection coefficient is required for each stage of the whitening and modeling filters for an isotropic process. Indeed as shown in [1], isotropy allowed us to develop a generalization of the Levinson and Schur scalar recursions for the local averages or barycenters of the prediction errors, which also yield the reflection coefficients. In this paper we have justified our claim that the reflection coefficients are a good parametrization for AR processes and isotropic processes in general. In particular we have developed whitening and modeling filters for AR processes that can be completely specified in terms of these coefficients. In addition we have shown that there is a one-to-one correspondence between finite reflection coefficient sequences and AR processes, have characterized the stability of lattice filters in terms of the reflection coefficients and have shown how the regularity of an isotropic process can be characterized in terms of its reflection coefficient sequence.

It is our belief that the theory developed in this paper provides an extremely useful framework for the development of multiscale statistical signal processing algorithms. In particular we expect this framework and its multidimensional coun-

terparts to be useful in analyzing signals displaying fractal-like or self-similar characteristics, i.e. random signals whose behavior is similar at multiple scales. When restricted to a given level of resolution, a sample of an isotropic process can be drawn like an ordinary signal. We show in Figure 10 a sample of an AR(3) process with $k_1 = k_2 = k_3 = 0.99$. The next figure shows approximations of this signal at successively coarser scales using the multiresolution analysis via wavelets of Mallat-Daubechies, as presented in [4] (see also the formula (2.1) and subsequent discussion in [1] for a brief explanation of this). These approximations display the self-similar statistical characteristics we expect of this class of models.

There are several promising directions for further research building on our formalism. In particular, as discussed in [3], stochastic process models on dyadic trees seem to form a useful setting for the development of algorithms for fusing data from sensors with several different resolutions. A second natural subject for investigation is the use of our formalism to analyze, recognize and classify the self-similar properties of signals by first performing a wavelet decomposition, producing a set of samples indexed by a dyadic tree and then by viewing this as a realization of an isotropic process from which a sequence of reflection coefficients can be estimated. In addition, several system-theoretic topics can be identified. In particular, the class of isotropic processes studied here is less than completely satisfactory. For example, for stationary time series we know that the forward and backward prediction errors of *all* orders are stationary. On the other hand the analogous statement is *not* true for isotropic processes on trees, leading to the question of finding a larger class of processes that contains these prediction errors: this point is investigated in [7] where the fundamentals of a system theory on the dyadic tree are investigated. Also, as pointed out in [1] the dyadic tree structure is most naturally associated with the Haar wavelet transform, while more general transforms lead to processes defined on weighted lattices. Work on each of these topics is proceeding and will be reported in the near future.

ACKNOWLEDGEMENT: the authors would like to express their thanks to Ofer Zeitouni, Ramine Nikoukhah, and especially to Bernard C. Lévy for extremely fruitful discussions and stimulating suggestions.

References

[1] M. BASSEVILLE, A. BENVENISTE, A.S. WILLSKY "Multi-Scale Autoregressive Processes, Part I: Schur-Levinson Parametrizations" submitted for publication.

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

[3] K.C. CHOU, A.S. WILLSKY, A. BENVENISTE, AND M. BASSEVILLE, "Recursive and Iterative Estimation Algorithms for Multi-Resolution Stochastic Processes," Proc. of the IEEE Conf. on Decision and Control, Tampa, Florida, Dec. 1989.

[4] I. DAUBECHIES, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Math.*, vol. 91, pp. 909–996, 1988.

[5] C.C. HEYDE, D. HALL, *Martingale Limit Theory and Applications,* Academic Press, 1980.

[6] J. NEVEU, *Martingales à temps discret,* Masson, Paris, 1972.

[7] A. BENVENISTE, R. NIKOUKHAH, A.S. WILLSKY, "Multiscale System Theory", submitted to CDC'90, IRISA res. rep., 1990.

Appendices

A Cigarillos' lemma

We shall use the following notation

$$wY_t = Y_{tw}$$

where w is a word. Note that we have

$$vY_{tw} = Y_{twv} = wvY_t$$

Furthermore, in the sequel, $[q]$ denotes the greatest integer smaller than q, and we shall write for short $\delta^{[q]}$ instead of $\delta^{([q])}$. Using these notations, we have the following result:

Lemma A.1 *For each n, the following formulae hold:*

$$\gamma^{-1}F_{.,n} = \delta^{\left[\frac{n+1}{2}\right]}\gamma^{-1}F_{.,n} \quad \textit{up to a permutation}$$

Proof: Recall that, for $w \preceq 0, |w| = n$

$$F_{t,n}(w) = Y_{tw} - E\left(Y_{tw} \mid \mathcal{Y}_{t,n-1}\right)$$

whence

$$\gamma^{-1}F_{t,n}(w) = Y_{t\gamma^{-1}w} - E\left(Y_{t\gamma^{-1}w} \mid \mathcal{Y}_{t\gamma^{-1},n-1}\right)$$

and

$$\gamma^{-1}F_{t\delta^{\left[\frac{n+1}{2}\right]},n}(w) = Y_{t\delta^{\left[\frac{n+1}{2}\right]}\gamma^{-1}w} - E\left(Y_{t\delta^{\left[\frac{n+1}{2}\right]}\gamma^{-1}w} \mid \mathcal{Y}_{t\delta^{\left[\frac{n+1}{2}\right]}\gamma^{-1},n-1}\right)$$

so that, to prove the lemma, it is enough to show the following formulae:

$$|w| = n, w' = \delta^{\left[\frac{n+1}{2}\right]}w \;\Rightarrow\; |w'| = n \tag{A.1}$$

$$|w| \leq n-1, w' = \delta^{\left[\frac{n+1}{2}\right]}w \;\Rightarrow\; |w'| \leq n-1 \tag{A.2}$$

Proof of (A.1): Set $w = \gamma^{-l}\delta^{(k)}, l + 2k = n$. Then

$$w' = \gamma^{-l}\delta^{\left(\left[\frac{n-1}{2}\right]-l\right)_+}\ \delta^{(k)} \tag{A.3}$$

where $x_+ = max(x, 0)$. To prove (A.1) it suffices to verify that

$$k \geq \left[\frac{n-1}{2}\right] - l$$

holds in (A.3), which amounts to verify that $n \geq \left[\frac{n-1}{2}\right] + k$, and this is a consequence of the inequalities $k \leq \left[\frac{n}{2}\right]$ and $n \geq \left[\frac{n}{2}\right] + \left[\frac{n-1}{2}\right]$.

Proof of (A.2): Again set $w = \gamma^{-l}\delta^{(k)}, l+2k \leq n-1$. Then $w' = \delta^{\left[\frac{n-1}{2}\right]}w$. Now if $k \geq \left[\frac{n-1}{2}\right] - l$ holds, then (A.2) follows. Otherwise $l + 2\left(\left[\frac{n-1}{2}\right] - l\right) = 2\left[\frac{n-1}{2}\right] - l \leq n - 1$ also proves (A.2).

B Proof of Equation 4.4

Take any $j > \frac{n}{2}$. Suppose that we can find as an isometry $\Psi : \mathcal{T} \to \mathcal{T}$ so that

1. $\Psi(t) = t$
2. Ψ maps the set $\{t\gamma^{-1}w | w \preceq, |w| \leq n-1\}$ onto itself.
3. Ψ maps the points $\{t\delta^{(j)}\} \cup \{t\delta^{(j)}\gamma^{-1}w | w \preceq 0, |w| \leq n-1\}$ onto a set of points each of which is $\preceq t\gamma^{-1}$.

Let $Y_t^{\Psi} = Y_{\Psi(t)}$ and define E^{Ψ} similarly. Then, thanks to isotropy, Y^{Ψ} has the same statistics as Y. Thus from (4.2)

$$E_{t,n}^{\Psi} \perp \mathcal{Y}_{t\gamma^{-1},\infty}^{\Psi} \tag{B.1}$$

However thanks to properties (1) and (2) of Ψ

$$E_{t,n} = E_{t,n}^{\Psi}$$

while thanks to property (3) and (2.9)

$$E_{t\delta^{(j)},n} \in \mathcal{H}\left\{Y_{t\gamma^{-1}w}^{\Psi} | w \preceq 0\right\} \tag{B.2}$$

Equations (B.1) and (B.2) then imply (4.4).

The required isometry is of the pivot type used in the proofs in Appendix C of [1]. As illustrated in Figure 13, the pivot for this isometry is the point $t\gamma^{-(1+n/2)}$ and the direction of "rotation" is as indicated in the figure. It is straightforward to check that this isometry has the required properties.

C Some Useful Lemmas

The first lemma is an immediate consequence of the Schur recursions (2.26)–(2.31):

Lemma C.1 *Consider the transformation Ψ which maps an isotropic covariance sequence $\{r_n\}$ to the corresponding reflection coefficient sequence. The Jacobian of this transformation satisfies the following:*

$$\frac{\partial k_n}{\partial r_m} = 0 \text{ for } n < m \tag{C.1}$$

$$\frac{\partial k_{2n}}{\partial r_{2n}} = \frac{1}{2^{n-1}P_{2n-1}(0)} \neq 0 \tag{C.2}$$

$$\frac{\partial k_{2n+1}}{\partial r_{2n+1}} = \frac{1}{2^{n-1}\left(P_{2n}(0) + \delta^{(n)}\left[P_{2n}\right](0)\right)} \neq 0 \tag{C.3}$$

where the P_n are the Schur series defined in (2.26).

Next we write $K \triangleq (k_n)_{1 \leq n \leq p}$ to denote a vector in R^p, and we let $\mathcal{S}$ denote the set of such vectors so that

$$\begin{aligned} -1 < \quad & k_{2n+1} \quad < +1 \\ -\frac{1}{2} < \quad & k_{2n} \quad < +1 \end{aligned}$$

Lemma C.2 *Consider a function Φ from R^p into R satisfying the following properties:*

1. *$\Phi(K) = 0$ if K is the reflection coefficient sequence of an isotropic process,*
2. *Φ is analytic inside $\mathcal{S}$.*

Then, $\Phi \equiv 0$ in $\mathcal{S}$.

Proof: Since Φ is analytic in $\mathcal{S}$, it is sufficient to show that Φ is zero on a set with nonempty interior in $\mathcal{S}$. Since we know that $\Phi(K) = 0$ if K is in the image of the map Ψ introduced in Lemma C.1, it is sufficient for us to show that the image of Ψ has a nonempty interior.

Thanks to the form of the Schur recursion formulae (2.26)–(2.31), we know that Ψ is also a rational function and, thanks to Lemma C.1, its Jacobian is triangular and always invertible. Thus it is sufficient to show that the set of finite sequences $\{r_n | 0 \leq n \leq N\}$ that can be extended to a covariance function of an isotropic process has a nonempty interior. However, this property is characterized by a *finite* family of conditions of the form

$$\mathcal{R}(r_0, \ldots, r_p) \geq 0 \tag{C.4}$$

where $\mathcal{R}(r_0, \ldots, r_p)$ denotes a matrix whose elements are chosen from the $r_0, \ldots, r_p$. The set of $(p+1)$–tuples satisfying these conditions with strict inequality is nonempty (for instance, $r_n = \delta_{n0}$ is the covariance sequence of white noise) and as a consequence the set of $r_0, \ldots, r_p$ satisfying (C.4) has a nonempty interior. This finishes the proof of the lemma.

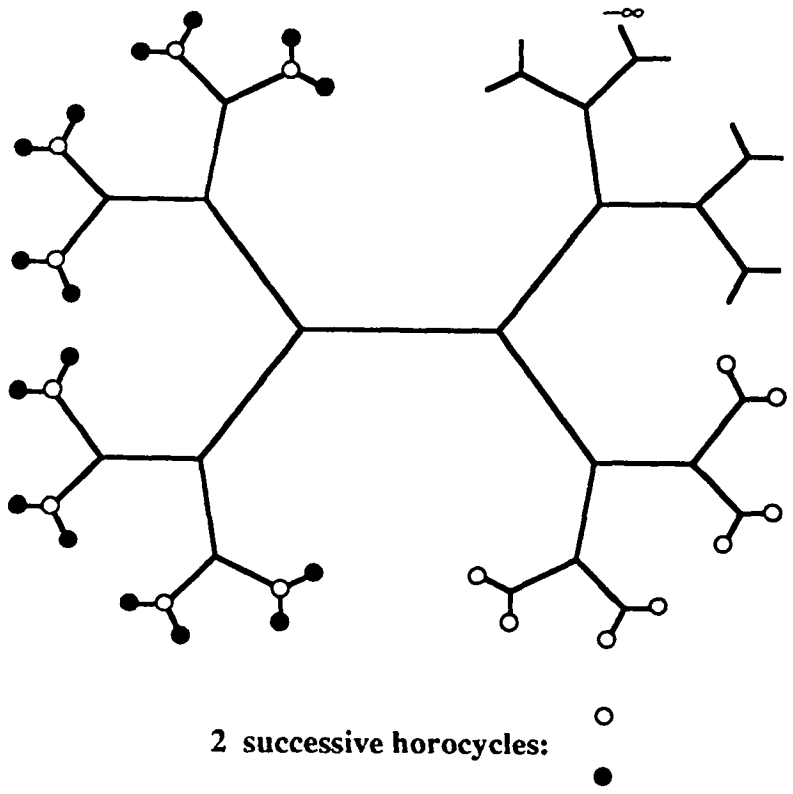

Figure 1: A homogenous dyadic tree, with a choice for the boundary point, $-\infty$, indicated.

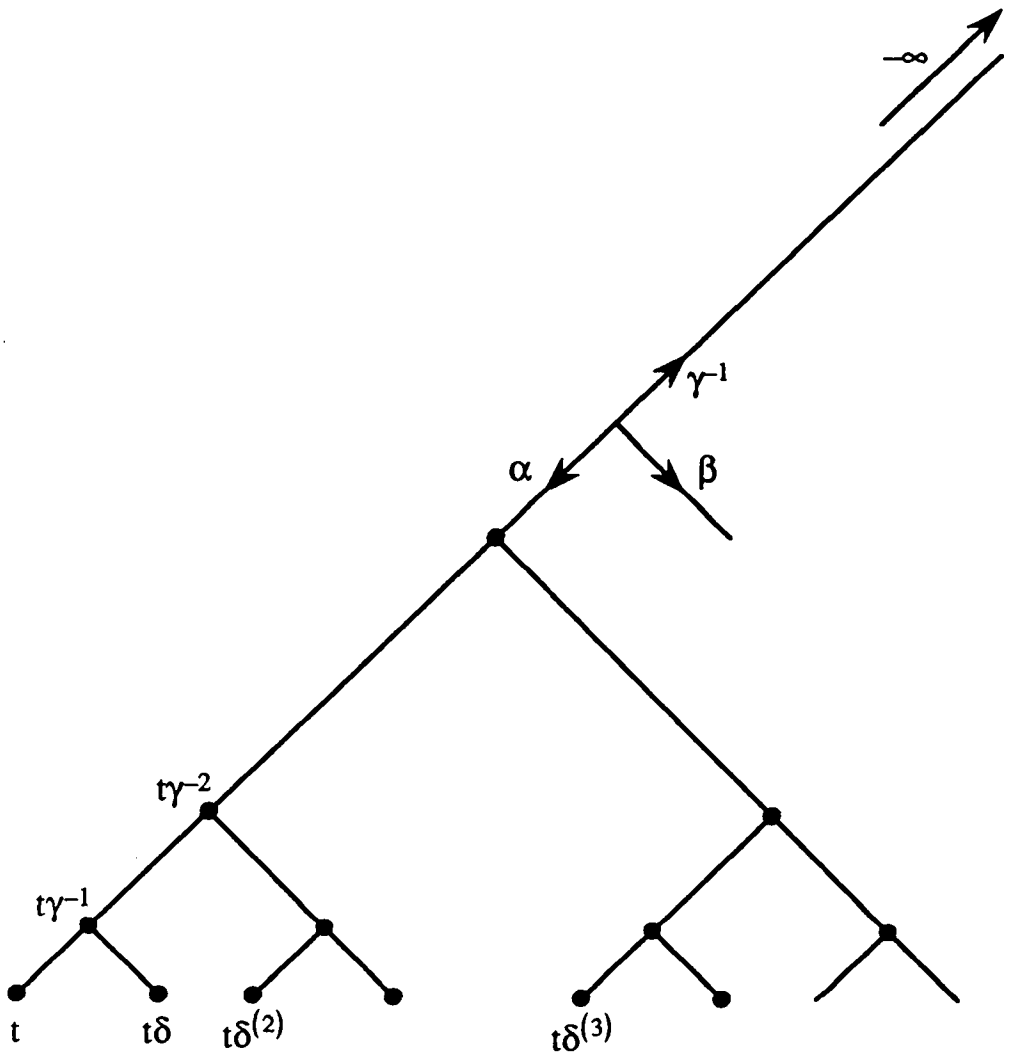

Figure 2: Redrawing the dyadic tree with a particular choice of boundary point $-\infty$.

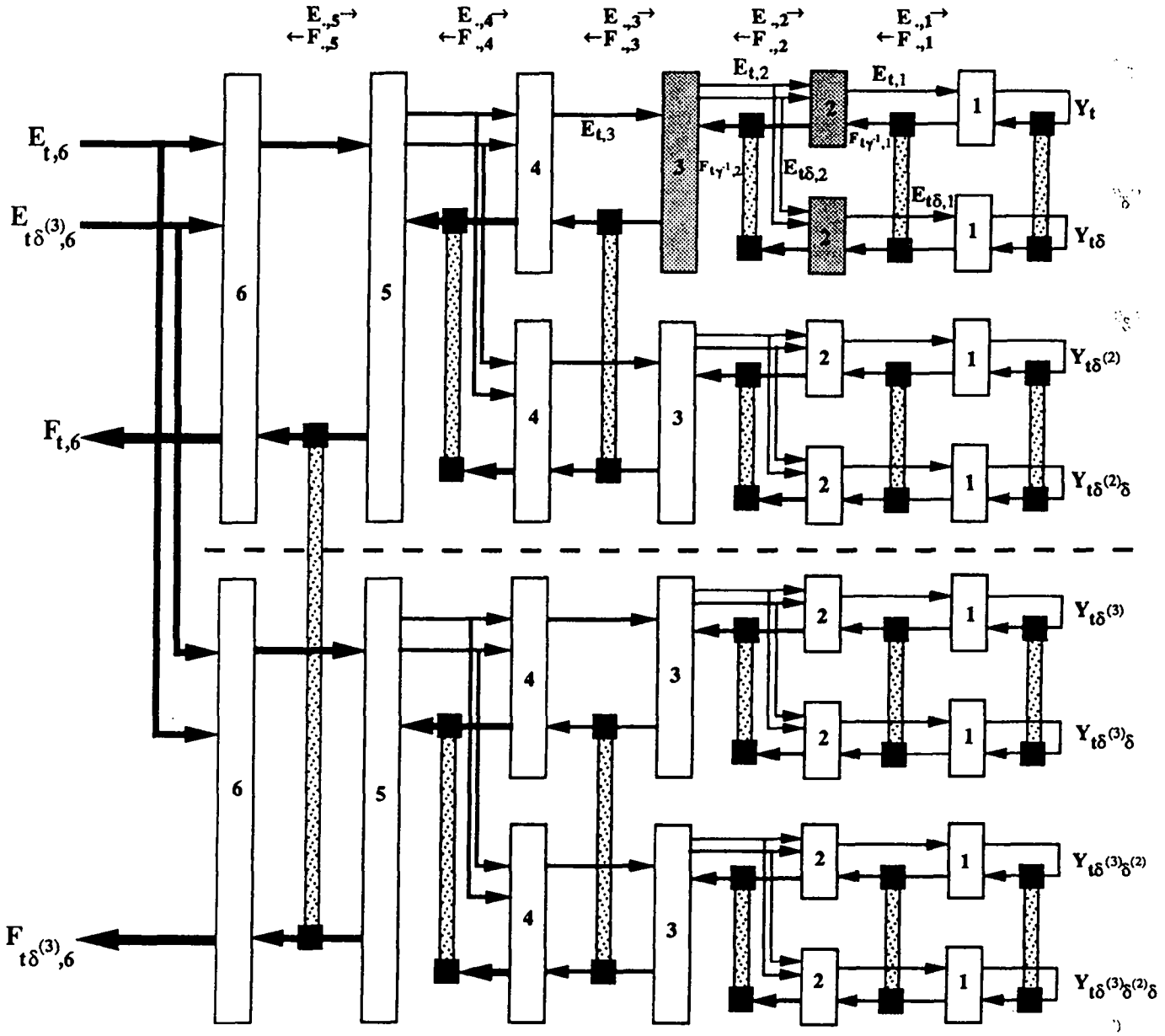

Figure 3: Illustrating the scattering/lattice structure of the modeling filter of Theorem 3.2 for a 6th-order model. Each block labeled "n" performs the computation in (3.23) (for n even), (3.25) (for n odd, $n > 1$), or (3.27) (for $n = 1$). The small solid squares denote γ^{-1} operations, and the dotted connections between such squares (the *cigarillos*) indicate signals (at the outputs of these squares), that are identical up to a permutation of components. As indicated at the top of the figure, the signals flowing through this system are the E and F error processes of successive orders, with the E's flowing left-to-right and the F's right-to-left.

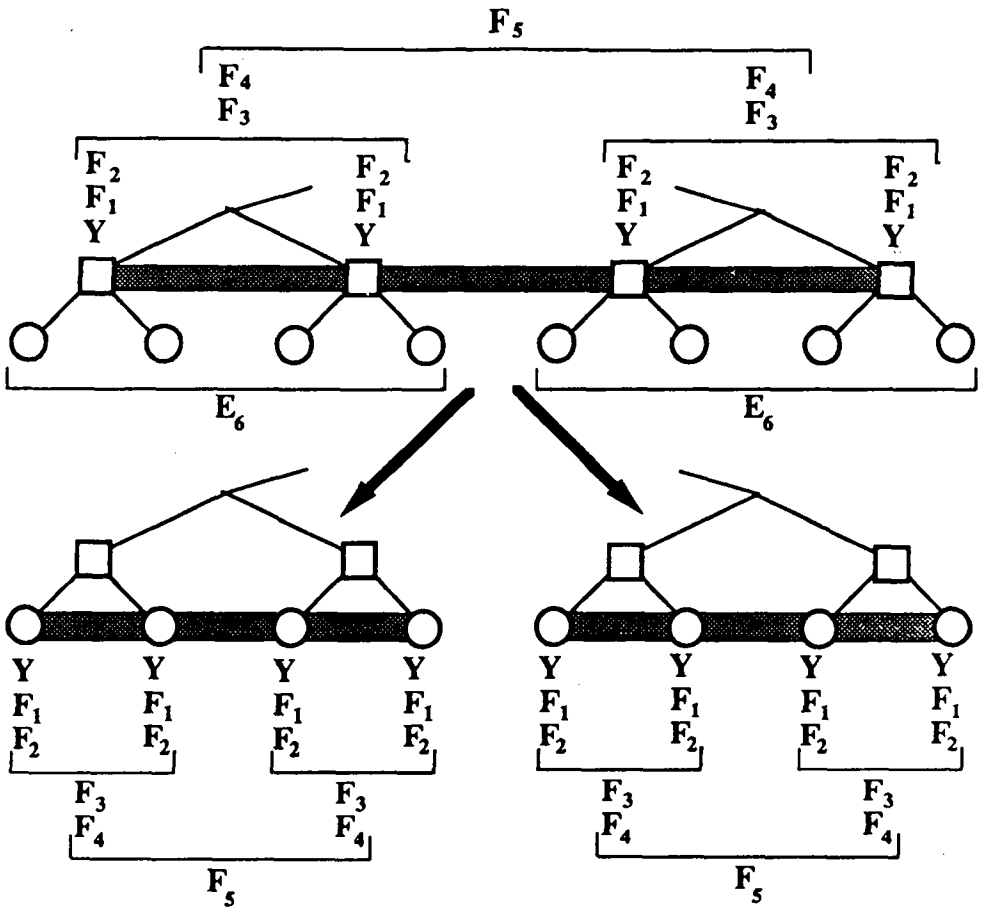

Figure 4: Illustrating the propagation of state information for the filter in Figure 3. The stored information (indicated above the top portion of the figure) for a set of four nodes at the horocycle indicated by squares is used, together with the input E_6 vectors, to compute the two corresponding sets of information at the two descendent groups of four points at the next horocycle.

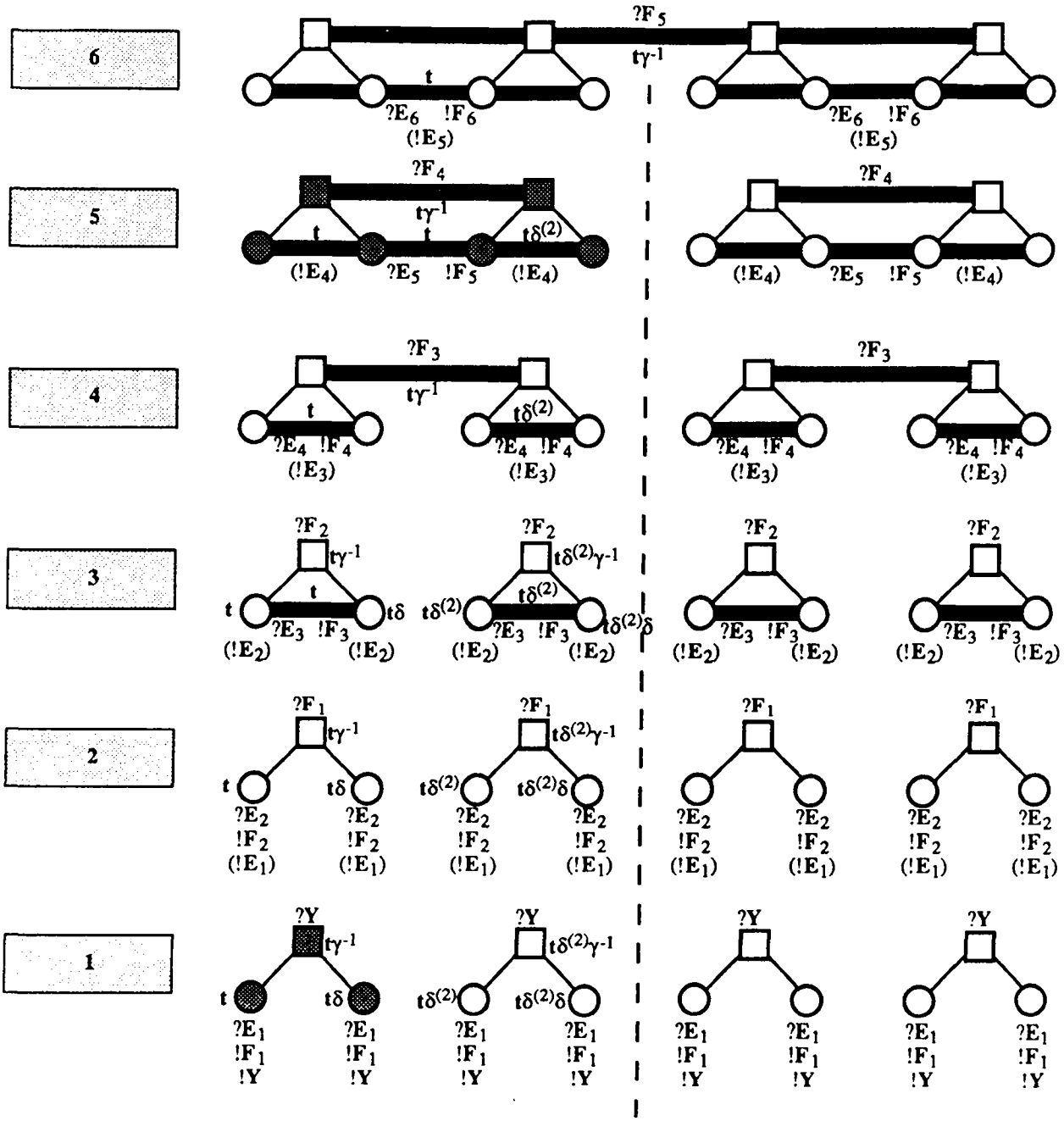

Figure 5: Illustrating the detailed computational flow for the propagation of state information for the filter described in Figures 3 and 4.

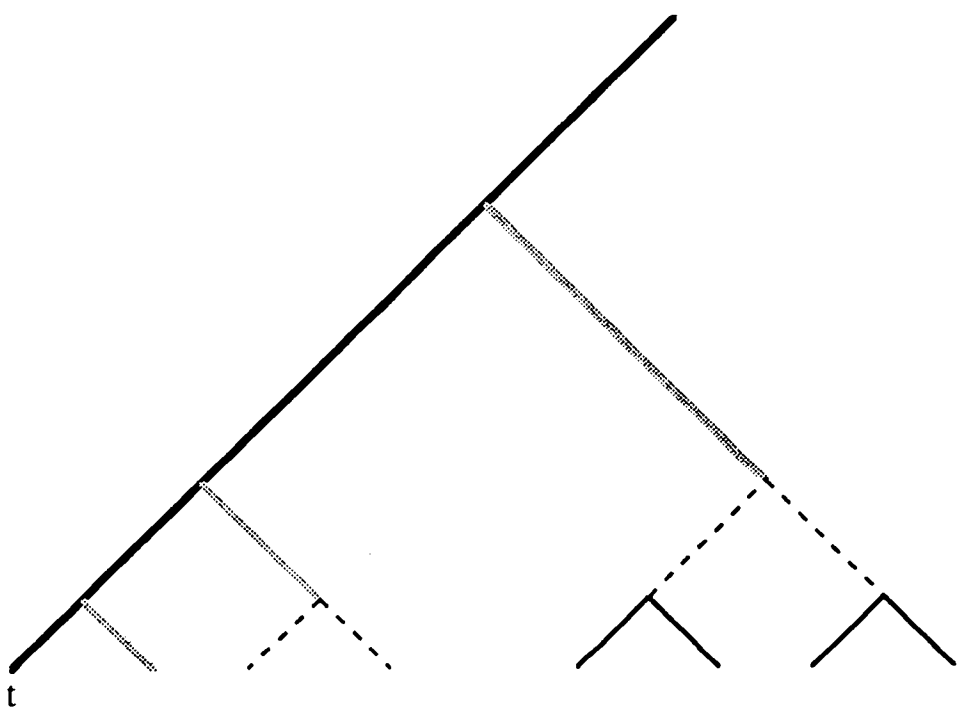

Figure 6: Illustrating the cylinder of radius 0 (support of AR(1)—dark solid line); cylinder of radius 1 (support of AR(2), AR(3)–dark solid and gray shaded lines); and cylinder of radius 2 (support of AR(4), AR(5)–dark solid, gray shaded, and dashed lines).

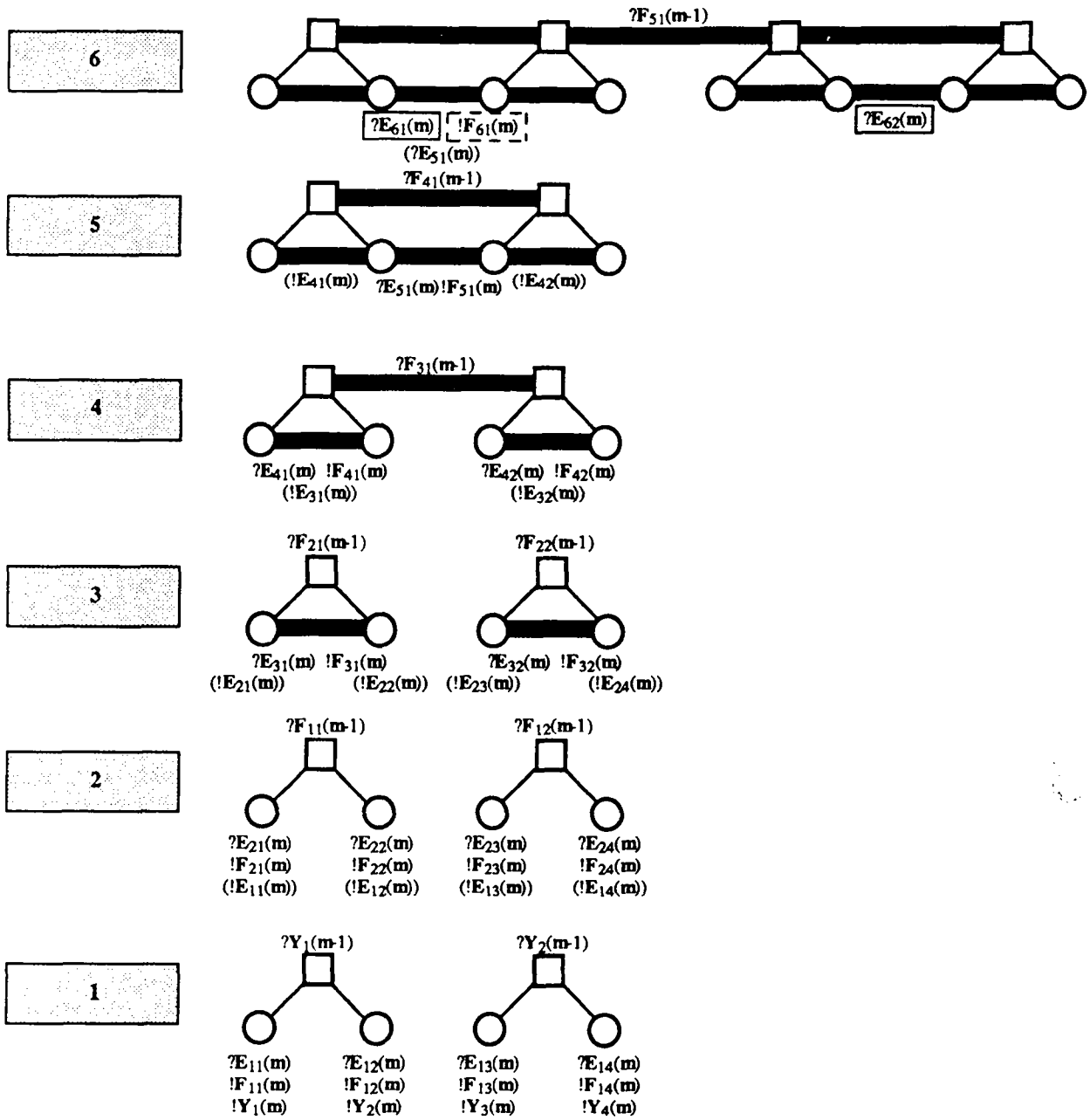

Figure 7: Illustrating one of the two parallel, sequential computations for the model of Figures 3–5.

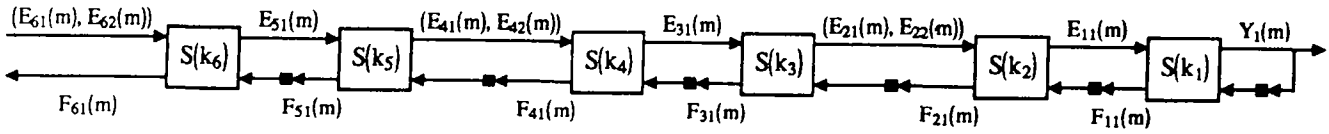

Figure 8: Illustrating one computational path from horocycle to horocycle. It is this standard time domain system whose stability is equivalent to that of the unnormalized modeling filter.

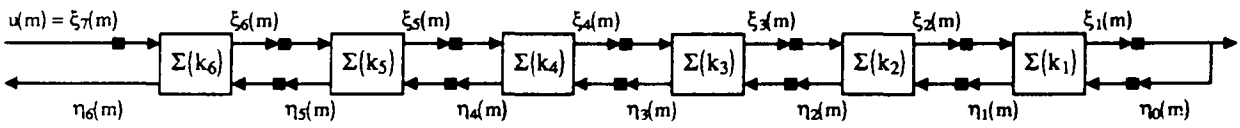

Figure 9: Equivalent system whose stability is investigated in the proof of Theorem 4.2.

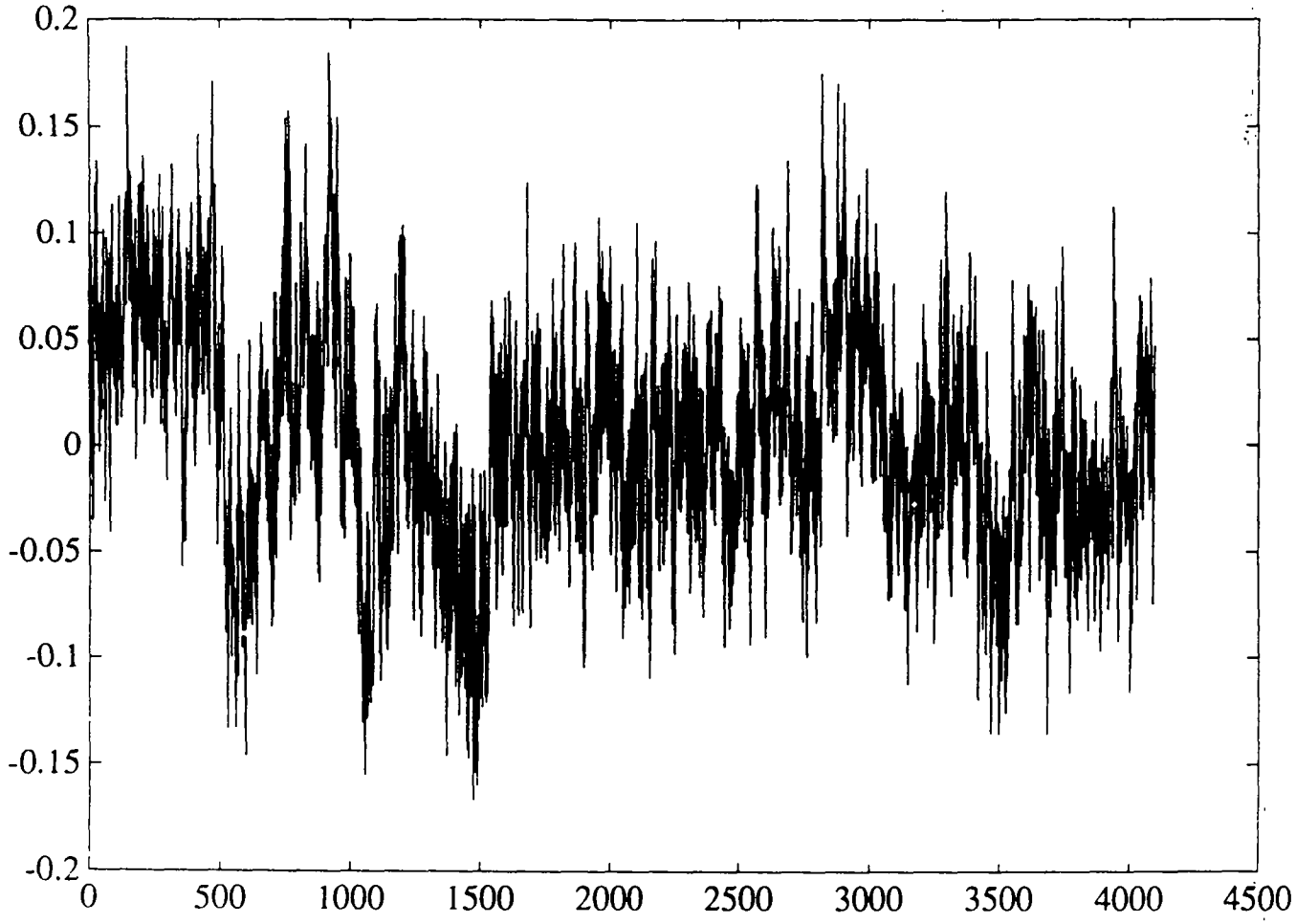

Figure 10: A sample of an AR(3) process at a given horocycle.

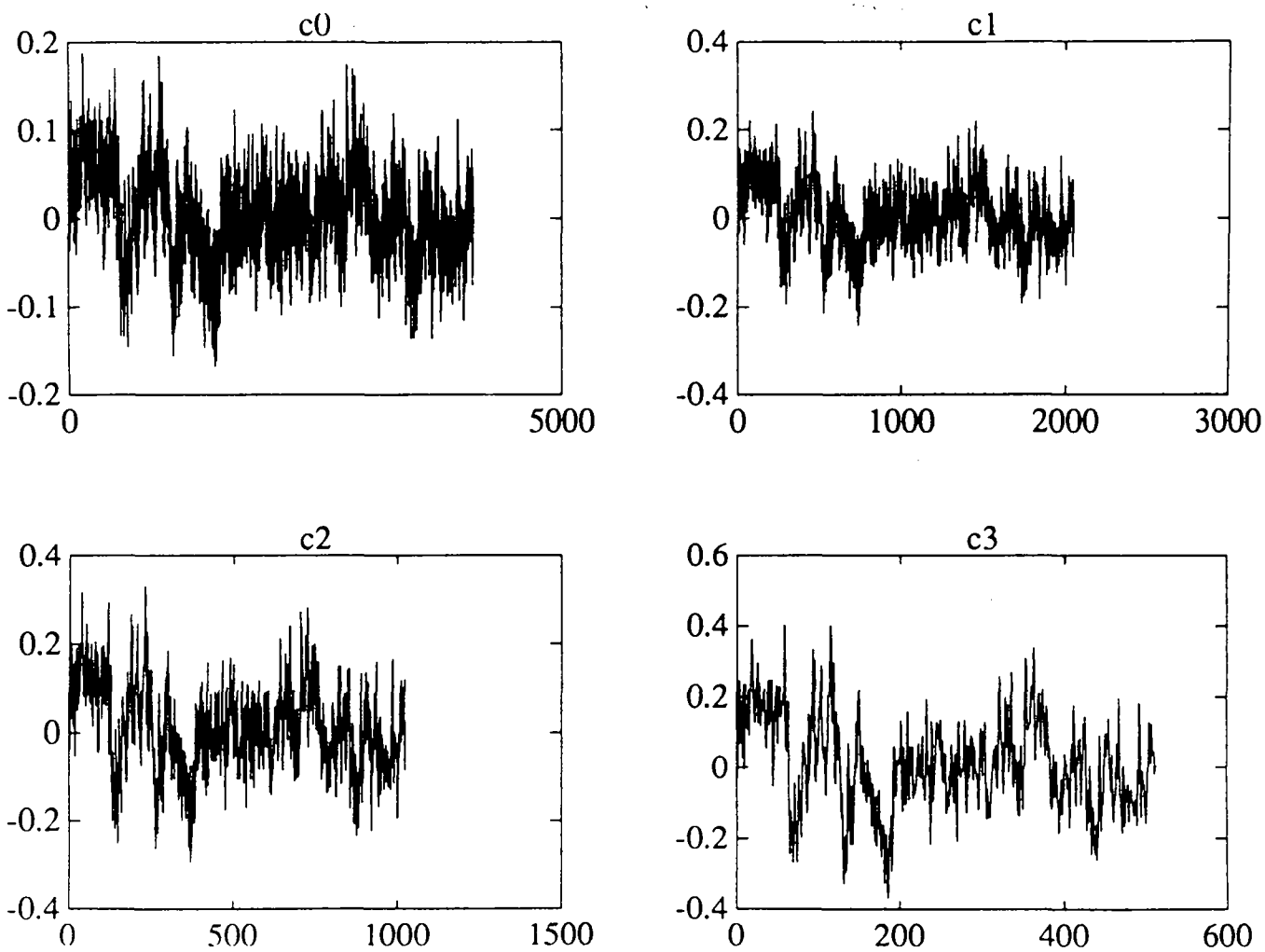

Figure 11: The Mallat-Daubechies multiresolution approximation of the signal of Figure 10.

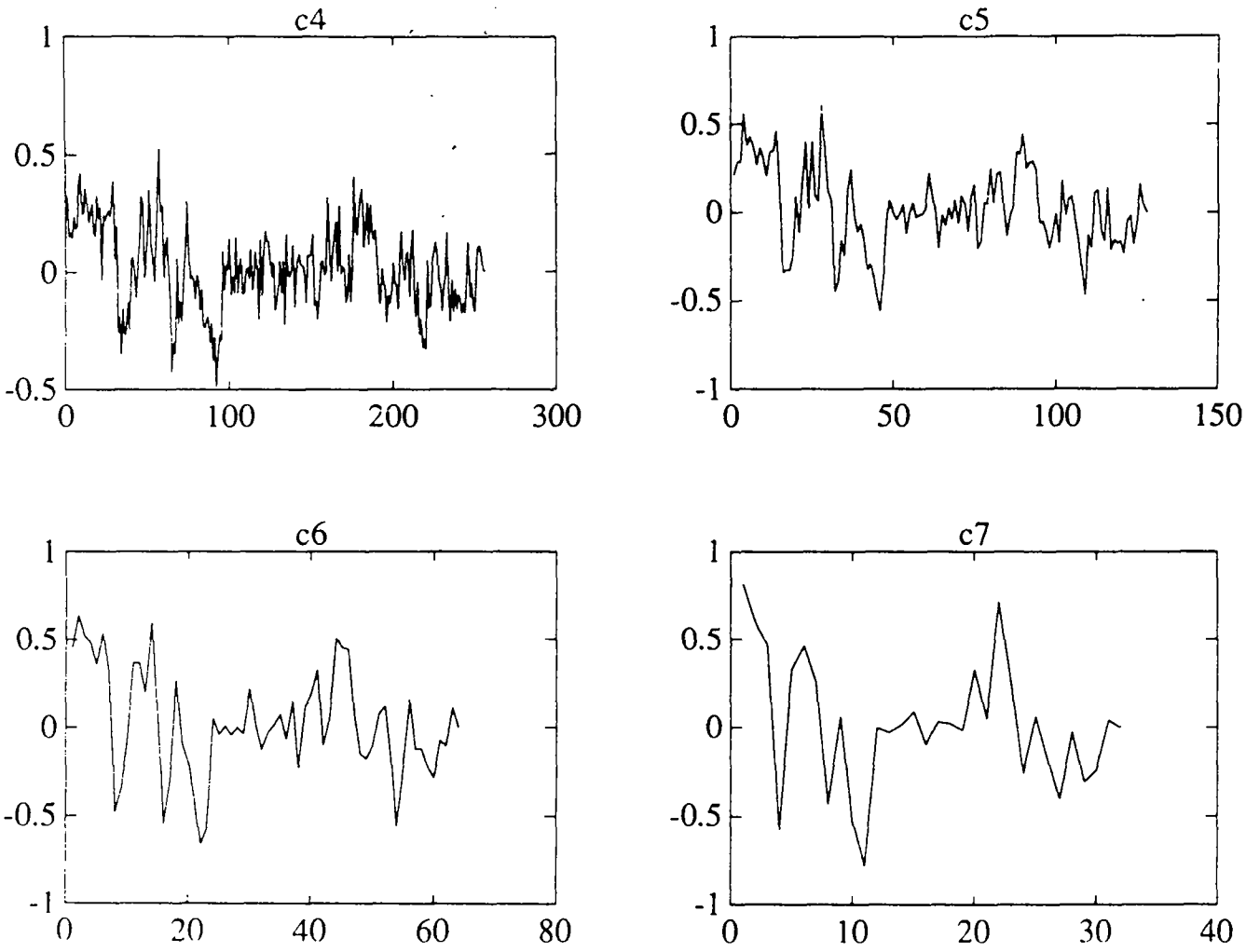

Figure 12: The Mallat-Daubechies multiresolution approximation of the signal of Figure 10, continued.

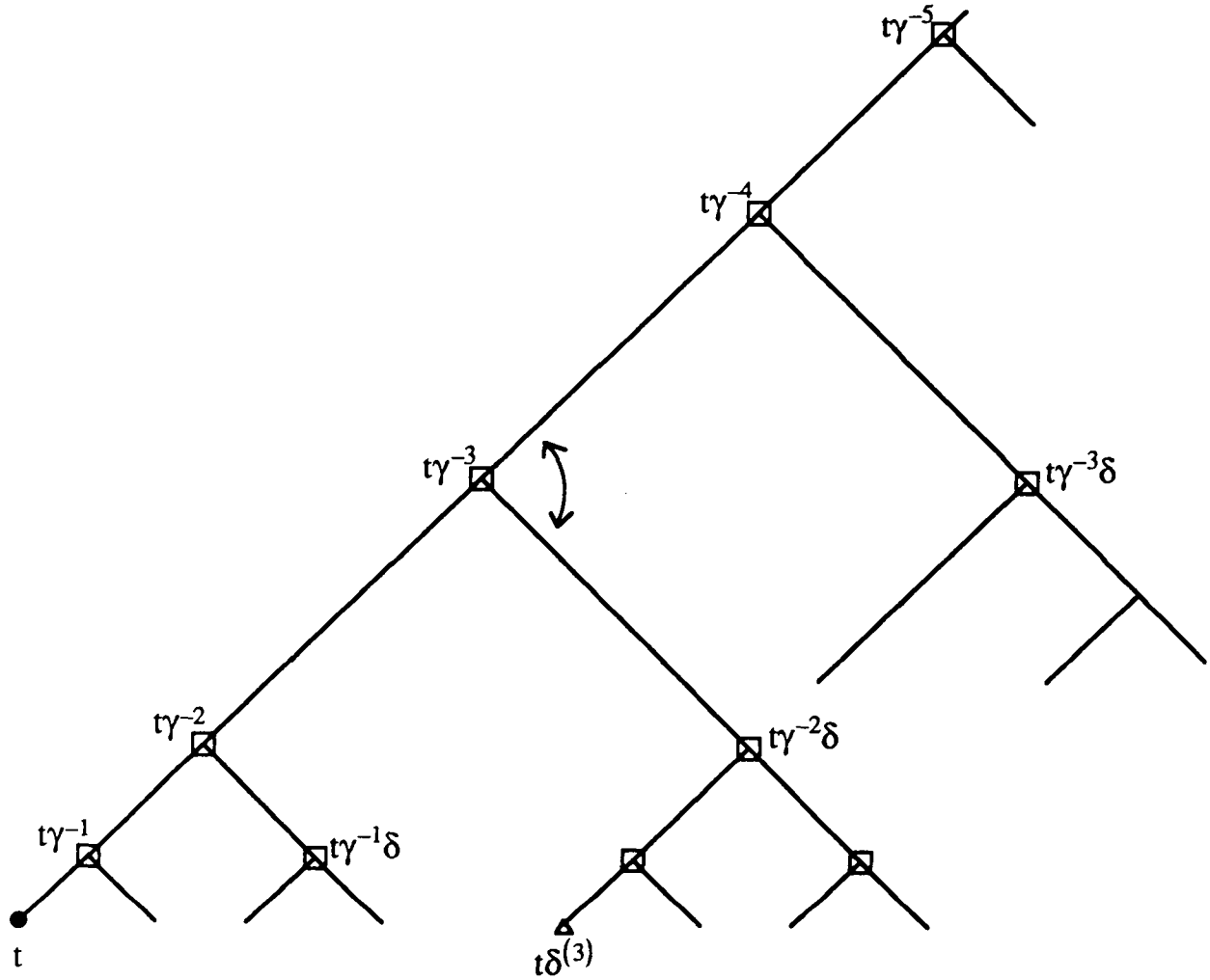

Figure 13: Illustrating the isometry used in Appendix A for the case $n = 5$ and $j = 3$. Here the pivot point is $t\gamma^{-3}$, so that the part of the tree toward t from $t\gamma^{-3}$ is left unchanged. The "rotation" exchanges the points $t\gamma^{-4}$ and $t\gamma^{-2}\delta$ and maps their successors accordingly. The set of nodes indicated with $\square$, which is in this case both $\{t\gamma^{-1}w|w \preceq 0, |w| \leq n-1\}$ and $\{t\delta^j\gamma^{-1}w|w \preceq 0, |w| \leq n-1\}$ is left invariant by this isometry. Also the point $t\delta^{(3)}$ is mapped onto one of the immediate successors of $t\gamma^{-3}\delta$, both of which are $\preceq t\gamma^{-1}$.

LISTE DES DERNIERES PUBLICATIONS INTERNES PARUES A L'IRISA

PI 518 **MULTISCALE SYSTEM THEORY**
Albert BENVENISTE, Ramine NIKOUKHAH, Alan S. WILLSKY
Février 1990, 30 Pages.

PI 519 **PANDORE : A SYSTEM TO MANAGE DATA DISTRIBUTION**
Françoise ANDRE, Jean-Louis PAZAT, Henry THOMAS
Février 1990, 14 Pages.

PI 520 **SCHEDULING AFFINE PARAMETERIZED RECURRENCES BY MEANS OF VARIABLE DEPENDENT TIMING FUNCTIONS**
Christophe MAURAS, Patrice QUINTON,
Sanjay RAJOPADHYE, Yannick SAOUTER
Février 1990, 14 Pages.

PI 521 **COMPUTABILITY OF RECURRENCE EQUATIONS**
Yannick SAOUTER, Patrice QUINTON
Février 1990, 28 Pages.

PI 522 **PROGRAMMING BY MULTISET TRANSFORMATION**
Jean-Pierre BANATRE, Daniel LE METAYER
Mars 1990, 26 Pages.

PI 523 **GOTHIC MEMORY MANAGEMENT : A MULTIPROCESSOR SHARED SINGLE LEVEL STORE**
Béatrice MICHEL
Mars 1990, 20 Pages.

PI 524 **ORDER NOTIONS AND ATOMIC MULTICAST IN DISTRIBUTED SYSTEMS : A SHORT SURVEY**
Michel RAYNAL
Mars 1990, 18 Pages.

PI 525 **MULTI-SCALE AUTOREGRESSIVE PROCESSES**
Michèle BASSEVILLE, Albert BENVENISTE
Mars 1990, 136 Pages.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique

ISSN 0249-6399