Multi-Scale Autoregressive Processes

Michèle Basseville[1]
Albert Benveniste[1]
Institut de Recherche en Informatique et Systèmes Aléatories
Campus Universitaire de Beaulieu
35042 RENNES CÉDEX
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 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.

[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.

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 [13,27] 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 possiblity of developing highly parallel and iterative algorithms based on such representations. Multigrid methods for solving partial differential equations [14,23,28,30] or for performing Monte Carlo experiments [18] 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 [20].

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 [24,26] and the closely related topic of wavelet transforms [4,5,6,7,10,19,22]. 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 [17,18] 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 characterize 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, in Section 3 we obtain a far simpler picture if we consider the generalization of lattice structures, and in particular we find that only one reflection coefficient is added as the order is increased by one. The latter fact leads to the development of a set of scalar recursions that provide us with the reflection coefficients and can be viewed as generalizations of the Schur and Levinson recursions for AR models of time series. These recursions are also developed in Section 3 as are the constraints that the reflection coefficients must satisfy which are somewhat different than for the case of time series. In Section 4 we then present the full vector Levinson recursions that provide us with both whitening and modeling filters for AR processes, and in Section 5 we use the analysis of the preceding sections to provide a complete characterization of the structure of autoregressive processes and a necessary and sufficient condition for an isotropic process to be purely nondeterministic. The paper concludes with a brief discussion in Section 6.

2 Multiscale Representations and Stochastic Processes on Homogenous Trees

2.1 Multiscale Representations and Wavelet Transforms

The multi-scale representation [25,26] 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

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

As $m \rightarrow \infty$ 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 [7], $h(n)$ 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 [7,31]. The simplest example of such a ϕ, h pair is the Haar approximation with

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

and

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

Multiscale represention 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 [7] 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 [3], 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 [1,2,3,12, 16] 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 2.1, 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, onto a 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 [12]:

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 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* [1,16] 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 2.1. 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 $\mathcal{T}$:

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

For example, the points s, t, and u in Figure 2.1 are all equivalent. The equivalence classes of such nodes are referred to as *horocycles*.

These equivalence classes can best be visualized as in Figure 2.2 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 2.2 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 2.3

- 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 intercharge 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 right-most operator is applied first):

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

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

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

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

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 2.3

$$s_1 = \gamma^{-4}t \quad , \quad s_2 = \delta\gamma^{-3}t \quad , \quad s_3 = \alpha\delta\gamma^{-3}t$$

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

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

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

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 ωt_0, where $\omega \in \mathcal{L}$. 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 ($\{\alpha, \beta\}^*\delta(\gamma^{-1})^*$). 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 $\alpha\gamma^{-1}$ or $\beta\gamma^{-1}$. For example $\alpha\gamma^{-1}t$ either equals t or δt 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 $\nu\omega$ need not be in $\mathcal{L}$ for $\nu, \omega \in \mathcal{L}$. However, for future reference we note that, using (2.14)–(2.17) we see that $\delta\omega$ and $\gamma^{-1}\omega$ are both in $\mathcal{L}$ for any $\omega \in \mathcal{L}$.

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

$$|\gamma^{-1}| = |\alpha\ | = |\beta| = 1$$

$$|0| = 0 \quad , \quad |\delta| = 2 \tag{2.20}$$

Thus $|\gamma^{-n}| = n$, $|\omega_{\alpha\beta}| =$ the number of α's and β's in $\omega_{\alpha\beta} \in \{\alpha\beta\}^*$, and $|\omega_{\alpha\beta}\delta\gamma^{-n}| = |\omega_{\alpha\beta}| + 2 + n$.[3] This notion of length will be useful in defining the *order* of dynamic 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 the either involve pure ascents on the tree, i.e.

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

all elements of $\{\gamma^{-1}\}^*$, or elements $\omega_{\alpha\beta}\delta\gamma^{-n}$ of $\{\alpha,\beta\}^*\delta\{\gamma^{-1}\}^*$ in which the descent is no longer than the ascent, i.e. $|\omega_{\alpha\beta}| \leq n$. We use the notation $\omega \preceq 0$ to indicate that ω 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 $\omega \asymp 0$ 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

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

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 $t = m_{\alpha,\beta}\gamma^{-n+1}t$ then $\delta^{(n)}t = m_{\alpha\beta}\delta\gamma^{-n+1}t$. For example, referring to Figure 2.3, $s_6 = \delta^{(4)}t$.

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

$$S = \sum_{\omega\in\mathcal{L}} s_\omega \cdot \omega \tag{2.22}$$

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

$$(Su)_t = \sum_{\omega\in\mathcal{L}} s_\omega u_{\omega t} \tag{2.23}$$

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:

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

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

where we use (2.14)–(2.17) and (2.21) to write $\gamma^{-1}\omega$ and $\delta^{(k)}\omega$ as elements of $\mathcal{L}$.

2.4 Isotropic Processes on Homogenous Trees

Consider a zero-mean stochastic process $Y_t, t \in T$ 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 [1,2,12] 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 $S(dw)$ so that

$$\begin{aligned} r_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega n} S(d\omega) \\ &= \frac{1}{\pi}\int_{0}^{\pi} \cos(\omega n) S(d\omega) \end{aligned} \tag{2.27}$$

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

$$r_n = \int_{-1}^{1} C_n(x)\mu(dx) \tag{2.28}$$

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

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

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

$$\mu(dx) = \frac{1}{\pi}(1-x^2)^{-\frac{1}{2}} \tag{2.30}$$

The analogous theorem for isotropic processes on dyadic trees requires the introduction of the Dunau polynomials [2,12]:

$$P_0(x) = 1 \quad , \quad P_1(x) = x \tag{2.31}$$

$$xP_n(x) = \frac{2}{3}P_{n+1}(x) + \frac{1}{3}P_{n-1}(x) \tag{2.32}$$

Theorem 2.1 [1,2]: *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*

$$r_n = \int_{-1}^{1} P_n(x)\mu(dx) \tag{2.33}$$

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 (2.33) in this case is [12]

$$\mu(dx) = \frac{1}{2\pi}\chi_{\left[-\frac{2\sqrt{2}}{3},\frac{2\sqrt{2}}{3}\right]}(x)\frac{(8-9x^2)^{\frac{1}{2}}}{1-x^2}dx \tag{2.34}$$

where $\chi_A(x)$ is the characteristic function of the set A. A key point here is that 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 $d(t_i, t_{i+1}) = 1$; 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 s, u, and $s \wedge t$ in Figure 2.1, and let

$$X^T = [Y_s, Y_u, Y_{s\wedge t}] \tag{2.35}$$

Then

$$E[XX^T] = \begin{bmatrix} r_0 & r_2 & r_2 \\ r_2 & r_0 & r_2 \\ r_2 & r_2 & r_0 \end{bmatrix} \geq 0 \tag{2.36}$$

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

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

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 [1] has the form

$$Y_t = \sum_{s \in \mathcal{T}} c_{d(s,t)} W_s \tag{2.38}$$

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 $f(s) = f(s')$,

$f(t) = f(t')$. Then

$$\begin{aligned} E[Y_{s'}Y_{t'}] &= \sum_u c_{d(s',u)}c_{d(t',u)} \\ &= \sum_{u'} c_{d(s',f(u'))}c_{d(t',f(u'))} \\ &= \sum_{u'} c_{d(f(s),f(u'))}c_{d(f(t),f(u'))} \\ &= \sum_{u'} c_{d(s,u')}c_{d(t,u')} = E[Y_sY_t] \end{aligned} \tag{2.39}$$

The class of systems of the form of (2.38) 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

$$Y_t = (SW)_t \quad , \quad S = \textstyle\sum_{\omega\in\mathcal{L}} s_\omega \cdot \omega \tag{2.40}$$

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

$$Y_t = \sum_{\substack{\omega \preceq 0 \\ |\omega| \leq p}} a_\omega Y_{\omega t} + \sigma W_t \tag{2.41}$$

where W_t is a white noise with unit variance.

The form of (2.41) deserves some comment. First note that the constraints placed on ω in the summation of (2.41) state that Y_t is a linear combination of the white noise W_t and the values of Y at nodes that are both at distances at most p from $Y(|\omega| \leq p)$ and also on the same or previous horocycles ($\omega \preceq 0$). Thus the model

(2.41) is not strictly "causal" and is indeed an implicit specification since values of Y on the same horocycle depend on each other through (2.41) (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 $\gamma^{-n}t$. 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:

$$Y_t = \sum_{n=0}^{\infty} a_n W_{\gamma^{-n}t} \tag{2.42}$$

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

$$a_n = \sigma a^n \tag{2.43}$$

so that the *only* process with strict past dependence as in (2.42) is the AR(1) process

$$Y_t = aY_{\gamma^{-1}t} + \sigma W_t \tag{2.44}$$

Consider next the AR(2) process, which specializing (2.41), has the form

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

Note first that this is indeed an implicit specification, since if we evaluate (2.45) at δt rather than t we see that

$$Y_{\delta t} = a_1 Y_{\gamma^{1}t} + a_2 Y_{\gamma^{-2}t} + a_3 Y_t + \sigma W_{\delta t} \tag{2.46}$$

We can, of course, solve the pair (2.45), (2.46) to obtain the explicit formulae

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

$$Y_{\delta t} = \left(\frac{a_1}{1-a_3^2}\right) Y_{\gamma^{-1}t} + \left(\frac{a_2}{1-a_3^2}\right) Y_{\gamma^{-2}t} + \sigma V_{\delta t} \tag{2.48}$$

where

$$V_t = \frac{1}{1-a_3^2}\left\{W_t + a_3 W_{\delta t}\right\} \tag{2.49}$$

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 $E\left[V_t V_{\gamma^{-2}t}\right] \neq E\left[V_t V_{\delta t}\right]$). Thus while the explicit representation (2.47)–(2.48) may be of some value in some contexts (e.g. in [17] we use similar nonisotropic models to analyze some estimation problems), the implicit characterization (2.45) is the more natural choice for a generalization of AR modeling.

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 (2.45) $a_3 \neq 0$ 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 (2.45) are related by a 4th-order polynomial relation.

Because of the complex relationship among the a_w's in (2.41), 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 Reflection Coefficients and Levinson and Schur Recursions for Isotropic Trees

As outlined in the preceding section the direct parametrization of isotropic AR models in terms of their coefficients $\{a_\omega\}$ 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 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 [8,9] 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 (2.41) and the fact that unlike the real line, the number of points a distance p from a node on a tree *increases* geometrically with p. For example, from (2.45)–(2.49) we see that Y_t and $Y_{\delta t}$ are closely coupled in the AR(2) model, and thus their prediction might best be considered simultaneously. For higher orders the coupling involves (a geometrically 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, thanks to isotropy, deduce that only one new coefficient is needed as the filter order is increased by one. This leads to the desired scalar recursions. In the next section we use the prediction filter origin of these recursions to construct *lattice forms* for modeling and whitening filters. Because of the variation in filter dimension the lattice segments are somewhat more complex and capture the fact that as we move inward toward a node, dimensionality decreases, while it increases if we expand outward.

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 [9], 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:

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

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 $(s \succ t)$ *and* on the *same* horocycle $(s \sim t)$ 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 3.1–3.4).

In analogy with the time series case, the backward innovations or prediction errors are defined as the variables spanning the new information, $\mathcal{F}_{t,n}$, in $\mathcal{Y}_{t,n}$ not contained in $\mathcal{Y}_{t,n-1}$:

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

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 $|\omega| \preceq 0$ and $|\omega| = n$, define

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

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

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

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

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

Let $E_{t,n}$ denote the column vector of the $E_{t,n}(\omega)$. 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 ω for which $|\omega| < n$, $w \asymp 0$. Specifically any such ω other than 0 must have the form

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

with

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

and with $|\omega| = 2i_k$. For example the points ωt for $\omega = 0, \delta, \delta^{(2)}$, and $\delta\delta^{(2)}$ are illustrated in Figure 3.4[4]. Thus the words ω 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 $F_{t,n}(\omega)$. 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 ω for which $|\omega| = n$ and $\omega \preceq 0$ can be written as $\omega = \tilde{\omega}\gamma^{-k}$ for some $k \leq 0$ and $\tilde{\omega} \asymp 0$. For example, as illustrated in Figure 3.4, for $n = 5$ the set of such ω's is ($\delta^{(2)}\gamma^{-1}$, $\delta\delta^{(2)}\gamma^{-1}$, $\delta\gamma^{-3}$, and γ^{-5}). We order the ω'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 $\tilde{\omega}$.

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 3.1 and the last three are in Figures 3.2–3.4 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 3.1): To begin we have*

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

The only word ω for which $|\omega| = 1$ and $\omega \preceq 0$ is $\omega = \gamma^{-1}$. Therefore

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

Also

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

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

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

[4]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.

$n = 2$ (See Figure 3.1): Here

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

In this case $|\omega| = 2$ and $\omega \preceq 0$ implies that $\omega = \delta$ or γ^{-1}. Thus

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

Similarly,

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

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

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

$n = 3$ (See Figure 3.2) In this case

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

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

Also

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

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

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

$n = 4$ *(See Figure 3.3)*

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

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

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

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

$n = 5$ *(See Figure 3.4)*

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

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

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

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

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 ωt for $|\omega| = n$, $\omega \asymp 0$)—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

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

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

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

For example for $n = 2$ this can be checked from the calculations in Example 3.1 plus the fact that

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

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 $n \leq p$, 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 $n \leq p$). 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 $\mathcal{T}$ 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 $F_{t,n}(\omega)$ and $E_{t,n}(\omega)$. 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 $F_{t,n}(\omega)$ for $|\omega| = n$, $\omega \preceq 0$. There are two natural subclasses for these words ω. In particular either $\omega \prec 0$ of $\omega \asymp 0$.

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

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

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

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

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

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

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

Now for n even we can show that

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

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

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

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

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

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

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

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

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

Similarly

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

Consider next the computation of $F_{t,n}(w)$ for $n \geq 3$ and odd. Note that for n odd it is impossible for $|w| = n$ *and* $w \asymp 0$. Therefore the condition

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

is equivalent to

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

Therefore, proceeding as before,

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

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

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

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

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

where in this case we use the fact that

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

For example for $n = 5$ these both equal

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

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 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 $E\left[E_{t,1}E_{\delta t,1}\right]$ and $E\left[E_{t,1}E_{\gamma^{-2}t,1}\right]$ which should be equal if $E_{t,1}$ is isotropic. From (3.18), (3.19) we find that

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

while

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

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 we state without proof is that $E_{t,n}$, suitably normalized, is isotropic for *all* $n \geq p$ if and only if Y_t is AR(p).

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

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

$$e_{t,n} = 2^{-\left[\frac{n-1}{2}\right]} \sum_{|w|<n, w \asymp 0} E_{t,n}(w) \tag{3.23}$$

$$f_{t,n} = 2^{-\left[\frac{n}{2}\right]} \sum_{|w|=n, w \preceq 0} F_{t,n}(w) \tag{3.24}$$

The following result is critical

Lemma 3.1 *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 $w'' \asymp 0$, we have that*

$$E\left(F_{\gamma^{-1}t,n-1}(w')|E_{t,n-1}\right) = E\left(E_{\delta(\frac{n}{2})t,n-1}(w'')|E_{t,n-1}\right) \quad (3.25)$$

$$= E\left(F_{\gamma^{-1}t,n-1}(w_0)|e_{t,n-1}\right) \quad (3.26)$$

$$= E\left(E_{\delta(\frac{n}{2})t,n-1}(0)|e_{t,n-1}\right) \quad (3.27)$$

(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

$$E\left(E_{t,n-1}(w)|F_{\gamma^{-1}t,n-1}\right) = E\left(E_{t,n-1}(0)|f_{\gamma^{-1}t,n-1}\right) \quad (3.28)$$

(refer to (3.16)).

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

$$E\left(E_{t,n-1}(w')|F_{\gamma^{-1}t,n-1}\right) = E\left(E_{\delta(\frac{n-1}{2})t,n-1}(w'')|F_{\gamma^{-1}t,n-1}\right) \quad (3.29)$$

$$= E\left(E_{t,n-1}(0)|f_{\gamma^{-1}t,n-1}\right) \quad (3.30)$$

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

$$E\left(F_{\gamma^{-1}t,n-1}(\tilde{w})|E_{t,n-1}, E_{\delta(\frac{n-1}{2})t,n-1}\right) = E\left(F_{\gamma^{-1}t,n-1}(w_0)|\frac{1}{2}\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1}\right)\right) \quad (3.31)$$

(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 3.5 and 3.6. Consider $n = 4$ first. Note that the distance relationships of each of the elements of $F_{\gamma^{-1}t,3}$ and of $E_{\delta(2)t,3}$ to $E_{t,3}$ are the same. Furthermore all three of these vectors contain errors in estimates based on $\mathcal{Y}_{\gamma^{-1}t,2}$. Hence because of this symmetry and the isotropy of Y, the projections of any of the elements of $F_{\gamma^{-1}t,3}$ or $E_{\delta(2)t,3}$ onto $E_{t,3}$ must be the same, as stated in (3.25). 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 (3.26), (3.27). Similarly, the two elements of $F_{\gamma^{-1}t,3}$ have identical geometric relations to each of the elements of $E_{t,3}$, so that (3.28) must hold. Similar geometric arguments apply to Figure 3.6 and (3.29)–(3.31) evaluated at $n = 5$. Perhaps the only one deserving comment is (3.31). Note, however, in this case that each of the elements of $F_{\gamma^{-1}t,4}$ has the same geometric relationship to *all* of the elements of $E_{t,4}$ *and* $E_{\delta^{(2)}t,4}$ and therefore the projection onto the combined span of these elements must weight the elements of $E_{t,4}$ and $E_{\delta^{(2)}t,4}$ equally and thus is a function of $\left(e_{t,n-1} + e_{\delta^{(\frac{n-1}{2})}t,n-1}\right)/2$.

Proof of Lemma 3.1: 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 (3.26). The other formulas are then obtained by analogous arguments.

The demonstration of (3.26) depends on the following three lemmas which are proved in Appendix C by exploiting symmetry and the isometry extension lemma.

Lemma 3.2 *The expectation*

$$G_{t,n} = E\left(F_{\gamma^{-1}t,n-1}(w)|E_{t,n-1}\right) \tag{3.32}$$

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

Lemma 3.3 *The expectation*

$$H_{t,n} = E\left(F_{\gamma^{-1}t,n-1}(w)E_{t,n-1}(w')\right) \tag{3.33}$$

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

Lemma 3.4 *The covariance $\sum_{E,n}$ of $E_{t,n}$ has the following structure. Let $\sum(\alpha_0, \ldots, \alpha_d)$ denote a $2^d \times 2^d$ covariance matrix, depending upon scalars $\alpha_0, \ldots, \alpha_d$ and with the following recursively-defined structure:*

$$\sum(\alpha_0) = \alpha_0 \tag{3.34}$$

$$\Sigma(\alpha_0, \cdots, \alpha_d) = \begin{bmatrix} \Sigma(\alpha_0, \cdots, \alpha_{d-1}) & \alpha_d U_{d-1} \\ & \\ \alpha_d U_{d-1} & \Sigma(\alpha_0, \cdots, \alpha_{d-1}) \end{bmatrix} \tag{3.35}$$

where U_{de} 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

$$\Sigma_{E,n} = \Sigma\left(\alpha_0, \cdots, \alpha_{[\frac{n-1}{2}]}\right) \tag{3.36}$$

From Lemma 3.2 we see that we need only show that $G_{t,n}$ depends only on $e_{t,n-1}$. However, from Lemma 3.4 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

$$X = \sum_{\substack{|w'|<n \\ w' \asymp 0}} \lambda_{w'} E_{t,n-1}(w') \tag{3.37}$$

where

$$\sum_{\substack{|w'|<n \\ w' \asymp 0}} \lambda_{w'} = 0 \tag{3.38}$$

Then, since $e_{t,n-1}$ is also as in (3.37) but with all λ_w equal, we have that

$$2^{[\frac{n-1}{2}]} E\left(X e_{t,n-1}\right) = \left(\lambda_{w'_1}, \cdots, \lambda_{w'_{2^{[\frac{n-1}{2}]}}}\right) \Sigma_{E,n} 1 = 0 \tag{3.39}$$

Thus we have an orthogonal decomposition of $\mathcal{E}_{t,n-1}$ into the space spanned by X as in (3.37), (3.38) and the one-dimensional subspace spanned by $e_{t,n-1}$. However, thanks to Lemma 3.3, for any X satisfying (3.37), (3.38)

$$E\left[F_{\gamma^{-1}t,n-1}(w) X\right] = \left(\sum_{w'} \lambda_{w'}\right) H_{t,n} = 0 \tag{3.40}$$

Thus the projection (3.32) is equal to the projection onto $e_{t,n-1}$, proving our result.

Remark: Lemma 3.4 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

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \end{bmatrix} \tag{3.41}$$

Also, as shown in Section 4.2 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 $l = \left[\frac{n-1}{2}\right]$.

3.4 Scalar Recursions for the Barycenters

An immediate consequence of Lemma 3.1, and the definitions of the barycenters, and the computations in Section 3.2 is the following recursions for the barycenters themselves:

For n even:

$$e_{t,n} = e_{t,n-1} - E\left(e_{t,n-1} | f_{\gamma^{-1}t,n-1}\right) \tag{3.42}$$

$$f_{t,n} = \frac{1}{2}\left(f_{\gamma^{-1}t,n-1} + e_{\delta(\frac{n}{2})t,n-1}\right) - \frac{1}{2}E\left(f_{\gamma^{-1}t,n-1} + e_{\delta(\frac{n}{2})t,n-1} | e_{t,n-1}\right) \tag{3.43}$$

For n odd, $n > 1$:

$$e_{t,n} = \frac{1}{2}\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1}\right) - \frac{1}{2}E\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1} | f_{\gamma^{-1}t,n-1}\right) \tag{3.44}$$

$$f_{t,n} = f_{\gamma^{-1}t,n-1} - E\left(f_{\gamma^{-1}t,n-1} | \frac{1}{2}\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1}\right)\right) \tag{3.45}$$

while for $n = 1$,

$$f_{t,1} = F_{t,1} \qquad e_{t,1} = E_{t,1} \tag{3.46}$$

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

It remains now to compute explicitly the projections indicated in (3.42)–(3.45). 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.

Lemma 3.5 *For n even:*

$$e_{t,n} = e_{t,n-1} - k_n f_{\gamma^{-1}t,n-1} \tag{3.47}$$

$$f_{t,n} = \frac{1}{2}\left(f_{\gamma^{-1}t,n-1} + e_{\delta(\frac{n}{2})t,n-1}\right) - k_n e_{t,n-1} \tag{3.48}$$

where

$$\begin{aligned} k_n &= cor(e_{t,n-1}, f_{\gamma^{-1}t,n-1}) \\ &= cor\left(e_{\delta(\frac{n}{2})t,n-1}, e_{t,n-1}\right) \\ &= cor\left(e_{\delta(\frac{n}{2})t,n-1}, f_{\gamma^{-1}t,n-1}\right) \end{aligned} \tag{3.49}$$

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

For n odd:

$$e_{t,n} = \frac{1}{2}\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1}\right) - k_n f_{\gamma^{-1}t,n-1} \tag{3.50}$$

$$f_{t,n} = f_{\gamma^{-1}t,n-1} - \frac{1}{2}k_n\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1}\right) \tag{3.51}$$

where

$$k_n = cor\left(\frac{1}{2}\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1}\right), f_{\gamma^{-1}t,n-1}\right) \tag{3.52}$$

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 3.6 *For n odd:*

$$E\left(e^2_{\delta^{(\frac{n+1}{2})}t,n}\right) = E\left(e^2_{t,n}\right) = E\left(f^2_{\gamma^{-1}t,n}\right) \triangleq \sigma^2_n \tag{3.53}$$

Lemma 3.7 *For n even* $\frac{1}{2}\left(e_{t,n} + e_{\delta^{(\frac{n}{2})}t,n}\right)$ *and* $f_{\gamma^{-1}t,n}$ *have the same variance.*

Proof of Lemma 3.5 We begin with the case of n even. Since $n-1$ is odd, Lemma 3.6 yields

$$E\left(e^2_{t,n-1}\right) = E\left(e^2_{\delta^{(\frac{n}{2})}t,n-1}\right) = E\left(f^2_{\gamma^{-1}t,n-1}\right) \triangleq \sigma^2_{n-1} \tag{3.54}$$

From (3.42)–(3.43) we than see that (3.47)–(3.49) are correct if

$$\begin{aligned} E\left[e_{t,n-1}f_{\gamma^{-1}t,n-1}\right] &= E\left[e_{\delta^{(\frac{n}{2})}t,n-1}e_{t,n-1}\right] \\ &= E\left[e_{\delta^{(\frac{n}{2})}t,n-1}f_{\gamma^{-1}t,n-1}\right] \triangleq g_{n-1} \end{aligned} \tag{3.55}$$

so that

$$k_n = \frac{g_{n-1}}{\sigma^2_{n-1}} \tag{3.56}$$

However, the first equality in (3.55) follows directly from Lemma 3.1 while the second equality results from the first with t replaced by $\delta^{(\frac{n}{2})}t$ and the fact that

$$\mathcal{F}_{\gamma^{-1}t,n-1} = \mathcal{F}_{\gamma^{-1}\delta^{(\frac{n}{2})}t,n-1} \tag{3.57}$$

For n odd the result directly follows from Lemma 3.7 and (3.44),(3.45).

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

$$\sigma^2_{e,n} = E\left(e^2_{t,n}\right) = \left(1-k^2_n\right)\sigma^2_{n-1} \tag{3.58}$$

$$\sigma^2_{f,n} = E\left(f^2_{t,n}\right) = \left(\frac{1+k_n}{2} - k^2_n\right)\sigma^2_{n-1} \tag{3.59}$$

where k_n must satisfy

$$-\frac{1}{2} < k_n < 1 \tag{3.60}$$

For n odd

$$\sigma^2_{e,n} = \sigma^2_{f,n} = \sigma^2_n = \left(1-k^2_n\right)\sigma^2_{f,n-1} \tag{3.61}$$

where

$$-1 < k_n < 1 \tag{3.62}$$

Proof: Equation (3.58) follows directly from (3.47) and (3.49) and the standard formulas for the estimation variance. Equation (3.59) follows in a similar way from (3.48) and (3.49) where the only slightly more complex feature is the use of (3.49) to evaluate the mean-squared value of the term in parentheses in (3.48). Equation (3.61) follows in a similar way from (3.50)–(3.52) and Lemma 3.7. The constraints (3.60) and (3.62) are immediate consequences of the nonnegativity of the various variances.

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 (3.60) of the constraint on k_n for n *even* over the form (3.62) that one also finds in the corresponding theory for time series. Also, as in the case of time series the satisfaction of (3.60) or (3.62) 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 Section 5.

3.5 Schur Recursions and Computation of the Reflection Coefficients

We now need to address the question of the explicit computation of the reflection coefficients. The key to this result is the following

Lemma 3.8 *For n even:*

$$\begin{aligned} E\left[e_{t,n-1} f_{\gamma^{-1}t,n-1}\right] &= E\left[Y_t f_{\gamma^{-1}t,n-1}\right] \\ &= E\left[Y_t e_{\delta^{(\frac{n}{2})}t,n-1}\right] \end{aligned} \tag{3.63}$$

$$\sigma_{n-1}^2 = E\left[e_{t,n-1}^2\right] = E\left[Y_t e_{t,n-1}\right] \tag{3.64}$$

For n odd

$$
\begin{aligned}
E\left(e_{t,n-1} f_{\gamma^{-1}t,n-1}\right) &= E\left(e_{\delta(\frac{n-1}{2})t,n-1} f_{\gamma^{-1}t,n-1}\right) \\
&= E\left[Y_t f_{\gamma^{-1}t,n-1}\right] \qquad (3.65)
\end{aligned}
$$

$$
\begin{aligned}
E\left[f^2_{\gamma^{-1}t,n-1}\right] &= E\left[\frac{1}{4}\left(e_{t,n-1} + e_{\delta(\frac{n-1}{2})t,n-1}\right)^2\right] \\
&= \frac{1}{2}\left[E\left(Y_t e_{t,n-1}\right) + E\left(Y_t e_{\delta(\frac{n-1}{2})t,n-1}\right)\right] \qquad (3.66)
\end{aligned}
$$

Proof: This result is essentially a consequence of other results we have derived previously. For example, for n even, since $f_{\gamma^{-1}t,n-1}$ is orthogonal to $\mathcal{Y}_{\gamma^{-1}t,n-2}$, we have that for $|w| < n$, $w \asymp 0$

$$
\begin{aligned}
E\left[Y_t f_{\gamma^{-1}t,n-1}\right] &= E\left[E_{t,n-1}(0) f_{\gamma^{-1}t,n-1}\right] \\
&= E\left[E_{t,n-1}(w) f_{\gamma^{-1}t,n-1}\right] \qquad (3.67)
\end{aligned}
$$

where the second equality follows from Lemma 3.2. Summing over $|w| < n$ and $w \asymp 0$ and using (3.23) then yields the first equality in (3.63). The second follows in a similar fashion (see also (3.25)). Similarly, since $e_{t,n-1}$ is also orthogonal to $\mathcal{Y}_{\gamma^{-1}t,n-2}$, we have that

$$
E\left[Y_t e_{t,n-1}\right] = E\left[E_{t,n-1}(0) e_{t,n-1}\right] = E\left[E_{t,n}(w) e_{t,n-1}\right] \qquad (3.68)
$$

The last equality here follows from the structure of $\Sigma_{E,n-1}$

$$
\begin{aligned}
E\left[E_{t,n-1}(w) e_{t,n-1}\right] &= [0, \ldots, 1, 0 \cdots 0] \Sigma_{E,n-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \text{eigenvalue associated with } [1, \ldots, 1]^T \qquad (3.69)
\end{aligned}
$$

(here the single 1 in the row vector in (3.69) is in the wth position.) Summing over w and using (3.23) yields (3.64). Equations (3.65) and (3.66) follow in an analogous manner.

It is now possible to write the desired recursions for k_n. Specifically if we multiply (3.47), (3.48), (3.50), (3.51) by Y_t and take expectations we obtain recursions for the quantities needed in the right-hand sides of (3.63)–(3.66). Furthermore, from (3.49) and (3.52) we see that k_n is directly computable from the left-hand sides of (3.63)–(3.66). In order to put the recursions in the most compact and revealing form it is useful to use formal power series. Specifically for $n \geq 0$ define P_n and Q_n as:

$$P_n \triangleq \operatorname{cov}(Y_t, e_{t,n}) \triangleq \sum_{w \preceq 0} E(Y_t e_{wt,n}) \cdot w \tag{3.70}$$

$$Q_n \triangleq \operatorname{cov}(Y_t, f_{t,n}) \triangleq \sum_{w \preceq 0} E(Y_t f_{wt,n}) \cdot w \tag{3.71}$$

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

$$P_0 = Q_0 = \sum_{w \preceq 0} r_{|w|} \cdot w \tag{3.72}$$

Recalling the definitions (2.24), (2.25) of γS 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 :

Proposition: The following formal power series recursions yield the sequence of reflection coefficients.

For n even

$$P_n = P_{n-1} - k_n \gamma Q_{n-1} \tag{3.73}$$

$$Q_n = \frac{1}{2}\left(\gamma Q_{n-1} + \delta^{(\frac{n}{2})} P_{n-1}\right) - k_n P_{n-1} \tag{3.74}$$

where

$$k_n = \frac{\gamma Q_{n-1}(0) + \delta^{(\frac{n}{2})} P_{n-1}(0)}{2 P_{n-1}(0)} \tag{3.75}$$

For n odd

$$P_n = \frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})} P_{n-1}\right) - k_n \gamma Q_{n-1} \tag{3.76}$$

$$Q_n = \gamma Q_{n-1} - k_n \frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})} P_{n-1}\right) \tag{3.77}$$

where

$$k_n = \frac{2\gamma Q_{n-1}(0)}{P_{n-1}(0) + \delta^{(\frac{n-1}{2})} P_{n-1}(0)} \qquad (3.78)$$

Note that for $n = 1$, (3.76)–(3.77) do agree with (3.17)–(3.19) since $P_0 = \delta^{(0)} P_0$, $\gamma Q_0(0) = r_1$ and $P_0(0) = r_0$.

4 Vector Levinson Recursions and Modeling and Whitening Filters

In this section we return to the vector prediction errors $E_{t,n}$, $F_{t,n}$ 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 are sufficient for us to study the question of stability. Consequently we begin with them.

4.1 Filters Involving the Unnormalized Residuals

To begin, let us introduce a variation on notation used to describe the structure of $\Sigma_{E,n}$. In particular we let 1_* denote a unit vector all of whose components are the same:

$$1_* = \frac{1}{\sqrt{\dim 1}} 1 \tag{4.1}$$

We also define the matrix

$$\mathsf{U}_* = 1_* 1_*^T \tag{4.2}$$

which has a single nonzero eigenvalue of 1. Equations (4.1), (4.2) 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:

$$\begin{aligned} \mathsf{U}_* \mathsf{U}_* &= \mathsf{U}_* && (4.3) \\ f^* &= 1_*^T F = \frac{1}{\sqrt{\dim F}} \sum_w F(w) && (4.4) \\ 1f = 1_* f^* &= \mathsf{U}_* F && (4.5) \end{aligned}$$

where $F = \{F(w)\}$ is a vector indexed by certain words w ordered as we have described previously, where f is its barycenter, and where f^* is a normalized version of its barycenter.

The results of the preceding section lead directly to the following recursions for the prediction error vectors:

Theorem 4.1 *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:

$$E_{t,n} = E_{t,n-1} - k_n \cup_* F_{\gamma^{-1}t,n-1} \tag{4.6}$$

$$F_{t,n} = \begin{bmatrix} E_{\delta(\frac{n}{2})t,n-1} \\ F_{\gamma^{-1}t,n-1} \end{bmatrix} - k_n \begin{bmatrix} \cup_* \\ \cup_* \end{bmatrix} E_{t,n-1} \tag{4.7}$$

For n odd, $n > 1$:

$$E_{t,n} = \begin{bmatrix} E_{t,n-1} \\ E_{\delta(\frac{n-1}{2})t,n-1} \end{bmatrix} - k_n \cup_* F_{\gamma^{-1}t,n-1} \tag{4.8}$$

$$F_{t,n} = F_{\gamma^{-1}t,n-1} - k_n \cup_* \begin{bmatrix} E_{t,n-1} \\ E_{\delta(\frac{n-1}{2})t,n-1} \end{bmatrix} \tag{4.9}$$

while for $n = 1$ we have the expressions (3.17)-(3.19).

Proof: As indicated previously, this result is a direct consequence of the analysis in Section 3. For example, from (3.16), Lemma 3.1 (3.28), and (4.5) we have the following chain of equalities for n even:

$$\begin{aligned} E_{t,n} &= E_{t,n-1} - E\left(E_{t,n-1} | F_{\gamma^{-1}t,n-1}\right) \\ &= E_{t,n-1} - \lambda 1 f_{\gamma^{-1}t,n-1} \\ &= E_{t,n-1} - \lambda \cup_* F_{\gamma^{-1}t,n-1} \end{aligned} \tag{4.10}$$

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 (3.47) 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 isometies of the tree.

In Section 3.4 we obtained recursions (3.58), (3.59), (3.61) for the variances of the barycenters of the prediction vectors. Theorem 4.1 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 $E_{t,n}$ 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

   $$\mu_{E,n} = 2^{\frac{n}{2}-1}\sigma^2_{e,n} \tag{4.11}$$

   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

   $$\mu_{F,n} = 2^{n}\sigma^2_{f,n} \tag{4.12}$$

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

2. For n odd,

   $$\Sigma_{E,n} = \Sigma_{F,n} = \Sigma_n \tag{4.13}$$

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

   $$\mu_n = \mu_{E,n} = \mu_{F,n} = 2^{(\frac{n-1}{2})}\sigma^2_n \tag{4.14}$$

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

3. For n even

   $$\Sigma_n \triangleq \Sigma_{F,n} = \text{cov}\begin{pmatrix} E_{t,n} \\ E_{\delta^{\frac{n}{2}}t,n} \end{pmatrix} = \begin{bmatrix} \Sigma_{E,n} & \lambda_n U \\ \lambda_n U & \Sigma_{E,n} \end{bmatrix} \tag{4.15}$$

where $U = 11^T$, and

$$\begin{aligned}\Sigma_{E,n} &= \Sigma_{n-1} - k_n^2 \sigma_{n-1}^2 U && (4.16)\\ \lambda_n &= \left(k_n - k_n^2\right) \sigma_{n-1}^2 && (4.17)\end{aligned}$$

4. For n odd, $n > 1$

$$\Sigma_n = \begin{bmatrix} \Sigma_{E,n-1} & \lambda_{n-1} U \\ \lambda_{n-1} U & \Sigma_{E,n-1} \end{bmatrix} - k_n^2 \sigma_{f,n-1}^2 U \qquad (4.18)$$

5. For $n = 1$

$$\Sigma_1 = \left(1 - k_1^2\right) r_0 \qquad (4.19)$$

Proof: Equations (4.11), (4.12), and (4.14) follow directly from the definition of the barycenter. For example, for n even

$$2^{(\frac{n}{2})-1} e_{t,n} = 1^T E_{t,n} \qquad (4.20)$$

from which (4.11) follows immediately. Equation (4.13) is a consequence of Lemma 3.1. To verify (4.15) let us first evaluate (4.6) at both t and $\delta^{(\frac{n}{2})} t$:

$$\begin{pmatrix} E_{t,n} \\ E_{\delta^{(\frac{n}{2})} t,n} \end{pmatrix} = \begin{pmatrix} E_{t,n-1} \\ E_{\delta^{(\frac{n}{2})} t,n-1} \end{pmatrix} - k_n \begin{pmatrix} \mathsf{U}_* \\ \mathsf{U}_* \end{pmatrix} F_{\gamma^{-1} t,n-1} \qquad (4.21)$$

The first equality in (4.15) is then a direct consequence of Lemma 3.1 (compare (4.7) and (4.21)). The form given in the right-most expression in (4.15) 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 3.1. The specific expression for $\Sigma_{E,n}$ in (4.16) follows directly from the second equality in (4.10), while (4.17) follows from (4.21) and the fact that

$$E\left[E_{t,n-1}(w) E_{t\delta^{(\frac{n}{2})} t,n-1}(w')\right] = k_n \sigma_{n-1}^2 \qquad (4.22)$$

which in turn follows from Lemma 3.1 and (3.49). Finally, (4.18) follows from (4.15) and (4.8), and (4.19) is immediate from (3.17)–(3.19).

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

Corollary: The modeling filter for Y_t is given by the following. For n even

$$\begin{pmatrix} E_{t,n-1} \\ F_{t,n} \end{pmatrix} = \Sigma\left(k_n\right) \begin{pmatrix} E_{t,n} \\ E_{\delta(\frac{n}{2})t,n} \\ F_{\gamma^{-1}t,n-1} \end{pmatrix} \tag{4.23}$$

where

$$\Sigma\left(k_n\right) \triangleq \begin{bmatrix} I & 0 & k_n \cup_* \\ -k_n \cup_* & I & \left(k_n - k_n^2\right) \cup_* \\ -k_n \cup_* & 0 & \left(I - k_n^2 \cup_*\right) \end{bmatrix} \tag{4.24}$$

For n odd, $n > 1$:

$$\begin{pmatrix} E_{t,n-1} \\ E_{\delta(\frac{n-1}{2})t,n-1} \\ F_{t,n} \end{pmatrix} = \Sigma\left(k_n\right) \begin{pmatrix} E_{t,n} \\ F_{\gamma^{-1}t,n-1} \end{pmatrix} \tag{4.25}$$

where

$$\Sigma\left(k_n\right) \triangleq \begin{bmatrix} I & k_n \cup_* \\ -k_n \cup_* & \left(I - k_n^2 \cup_*\right) \end{bmatrix} \tag{4.26}$$

while for $n = 1$:

$$\begin{pmatrix} E_{t,0} \\ F_{t,1} \end{pmatrix} = \begin{pmatrix} 1 & k_1 \\ -k_1 & 1 - k_1^2 \end{pmatrix} \begin{pmatrix} E_{t,1} \\ F_{\gamma^{-1}t,0} \end{pmatrix} \tag{4.27}$$

These equations can be verified by solving (4.6)–(4.9) and (3.17)–(3.19) 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 $n-1$. Thus, as in the case of lattice filters for time series we have a scattering layer-like structure for the generation of $Y_t = E_{t,0}$.

Since $E_{t,N}$ consists of prediction errors for Y_{wt}, $|w| < N$, $w \asymp 0$, the input-output map for the modeling filter is a map from the $2^{[\frac{N-1}{2}]}$-dimensional input $E_{t,N}$ to the $2^{[\frac{N-1}{2}]}$-dimensional vector of outputs $\{Y_{wt} | w < N, w \asymp 0\}$. Figure 4.1 provides a picture of the structure of this multivariable scattering system. Here $\Sigma(k_1)$

is the matrix on the right-hand side of (4.27), and for n odd $\Sigma(k_n)$ is given by (4.26). For n even $\hat{\Sigma}(k_n)$ is a modified version of $\Sigma(k_n)$ as it produces both $E_{t,n-1}$ and $E_{\delta(\frac{n}{2})_{t,n-1}}$ (essentially (4.21) with these two prediction vectors viewed as outputs and $E_{t,n}, E_{\delta(\frac{n}{2})_{t,n}}$ viewed as inputs). Also $\hat{\Sigma}(k_n)$ has both $F_{\gamma^{-1}t,n-1}$ and $F_{\gamma^{-1}\delta(\frac{n}{2})_{t,n-1}}$ as inputs. Note that the inputs to this block are not all linearly independent and thus there are a number of ways to write $\hat{\Sigma}(k_n)$ (essentially due to the fact that $F_{t,n} = F_{\delta(\frac{n}{2})_{t,n}}$). Ordering the inputs as $(E_{t,n}, E_{\delta(\frac{n}{2})_{t,n}}, F_{\gamma^{-1}t,n-1}, F_{\gamma^{-1}\delta(\frac{n}{2})_{t,n-1}})$ and the outputs as $(E_{t,n-1}, E_{\delta(\frac{n}{2})_{t,n-1}}, F_{t,n})$ one choice is

$$\hat{\Sigma}(k_n) = \begin{bmatrix} I & 0 & k_n U_* & 0 \\ 0 & I & 0 & k_n U_* \\ -k_n U_* & I & (k_n - k_n^2)U_* & 0 \\ -k_n U_* & 0 & I - k_n^2 U_* & 0 \end{bmatrix}$$

where all the blocks are of dimension $2^{(\frac{n}{2})-1}$ (note that this form emphasizes the redundancy in the input: given $E_{t,n}, E_{\delta(\frac{n}{2})_{t,n}}, F_{\gamma^{-1}t,n-1}$, all we need is $f_{\gamma^{-1}\delta(\frac{n}{2})_{t,n-1}}$).

Finally, based on this corollary we can now state the following stability result:

Theorem 4.2 *The conditions*

$$-1 < k_n < 1 \quad n \text{ odd} \quad 1 \leq n \leq N \tag{4.28}$$
$$-\frac{1}{2} < k_{2n} < 1 \quad n \text{ even} \quad 1 \leq n \leq N \tag{4.29}$$

are necessary and sufficient for the Nth-order modeling filter specified by reflection coefficients $\{k_n | 1 \leq n \leq N\}$ to be stable, so that a bounded input $E_{t,N}$ yields a bounded output $Y_t = E_{t,0}$.

Proof: This is a variation on the proof of stabilty for systems described by cascades of scattering sections, complicated here by the growing dimensionality of the E's and F's. Let us consider in detail the scattering diagram illustrated in Figure 4.1. Thanks to the fact that the forward transmission matrices are identity matrices and the common eigenstructure of all matrices involved, a necessary and sufficient condition for stability is that all reflection coefficient matrices have eigenvalues of magnitude less than 1 and all reverse transmission matrices have eigenvalues with magnitudes less than or equal to 1. For n odd, the transmission matrices I and $I - k_n^2 U_*$ have eigenvalues of 1 and $1 - k_n^2$, while the reflection matrices $\pm k_n U_*$, have nonzero eigenvalues of $\pm k_n$. From these we can deduce (4.28). For n even, the eigenvalues of

$$\begin{pmatrix} k_n U_* & 0 \\ 0 & k_n U_* \end{pmatrix}$$

are k_n and 0, yielding the constraint $|k_n| < 1$. However the tighter constraint comes from the other reflection coefficient matrix:

$$\begin{pmatrix} k_n U_* & I \\ k_n U_* & 0 \end{pmatrix}$$

The two nonzero eigenvalues of this matrix are the roots of the equation

$$\lambda^2 + k_n \lambda + k_n = 0$$

It is easily checked that these roots have magnitude less than one if and only if (4.29) is satisfied. Similarly the transmission matrix

$$\begin{bmatrix} (k_n - k_n^2) U_* & 0 \\ I - k_n U_* & 0 \end{bmatrix}$$

has nonzero eigenvalue $(k_n - k_n^2)$, which is between ± 1 for k_n satisfying (4.29), completing the proof.

4.2 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 $E_{\delta(\frac{n}{2})t,n-1}$ are correlated (see (4.15)). 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 it symmetric, positive definite square root.

We now can state and prove the following:

Theorem 4.3 *The following are the recursions for the normalized residuals.*

For n even

$$\begin{pmatrix} E_{t,n} \\ E_{\delta(\frac{n}{2})t,n} \end{pmatrix}^* = \Theta(k_n)\left[\begin{pmatrix} E^*_{t,n-1} \\ E^*_{\delta(\frac{n}{2})t,n-1} \end{pmatrix} - k_n \begin{pmatrix} \mathsf{U}_* \\ \mathsf{U}_* \end{pmatrix} F^*_{\gamma^{-1}t,n-1}\right] \quad (4.30)$$

$$F^*_{t,n} = \Theta(k_n)\left[\begin{pmatrix} E^*_{\delta(\frac{n}{2})t,n-1} \\ F^*_{\gamma^{-1}t,n-1} \end{pmatrix} - k_n \begin{pmatrix} \mathsf{U}_* \\ \mathsf{U}_* \end{pmatrix} E^*_{t,n-1}\right] \quad (4.31)$$

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

$$\Theta^{-2}(k_n) = \begin{pmatrix} I - k_n^2\mathsf{U}_* & (k_n - k_n^2)\mathsf{U}_* \\ (k_n - k_n^2)\mathsf{U}_* & I - k_n^2\mathsf{U}_* \end{pmatrix} \quad (4.32)$$

For n odd, $n > 1$

$$E^*_{t,n} = \Theta(k_n)\left[\begin{pmatrix} E_{t,n-1} \\ E_{\delta(\frac{n-1}{2})t,n-1} \end{pmatrix}^* - k_n\, \mathsf{U}_*\, F^*_{\gamma^{-1}t,n-1}\right] \quad (4.33)$$

$$F^*_{t,n} = \Theta(k_n)\left[F^*_{\gamma^{-1}t,n-1} - k_n\, \mathsf{U}_* \begin{pmatrix} E_{t,n-1} \\ E_{\delta(\frac{n-1}{2})t,n-1} \end{pmatrix}^*\right] \quad (4.34)$$

where

$$\Theta^{-T}(k_n)\Theta^{-1}(k_n) = I - k_n^2\mathsf{U}_* \quad (4.35)$$

for $n = 1$

$$E^*_{t,1} = \frac{1}{\sqrt{1-k_1^2}}\left(E^*_{t,0} - k_1 F^*_{\gamma^{-1}t,0}\right) \quad (4.36)$$

$$F^*_{t,1} = \frac{1}{\sqrt{1-k_1^2}}\left(F^*_{\gamma^{-1}t,0} - k_1 E^*_{t,0}\right) \quad (4.37)$$

Remark: Note that for n even we normalize $E_{t,n}$ and $E_{\delta(\frac{n}{2})t,n}$ *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 (4.15)–(4.19) 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: This result is a relatively straightforward computation given (4.11)–(4.19). For n even we begin with (4.7) and (4.21) and premultiply each by

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

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 (4.30) and (4.31) where the matrix $\Theta(k_n)$ is simply the inverse of the sqauare root of the covariance of the term in brackets in (4.30) and in (4.31) (the equality of these covariances follows from (4.15)). Equation (4.32) then follows from (4.11) and (4.15). The case of n odd involves an analogous set of steps, and the $n = 1$ case is immediate.

The preceding result provides us with a recursive procedure for calculating $\Sigma_n^{-1/2}$ (see Appendix D for an alternate efficient procedure). For n even we have

$$\Sigma_n^{-1/2} = \Theta(k_n)\ \text{diag}\left(\Sigma_{n-1}^{-1/2} \quad , \quad \Sigma_{n-1}^{-1/2} \right) \tag{4.38}$$

while for n odd, $n > 1$

$$\Sigma_n^{-1/2} = \Theta(k_n)\Sigma_{n-1}^{-1/2} \tag{4.39}$$

and for $n = 1$

$$\Sigma_1^{-1/2} = \left[(1 - k_1^2)r_0\right]^{-1/2} \tag{4.40}$$

The calculation of $\Theta(k_n)$ is also obtained in a straightforward way using the following two formulae. For any $k > -1$

$$(I + k\mathsf{U}_*)^{-\frac{1}{2}} = I + \left(\frac{1}{\sqrt{1+k}} - 1\right)\mathsf{U}_* \tag{4.41}$$

and for S and T symmetric

$$\begin{pmatrix} S & T \\ T & S \end{pmatrix}^{-1/2} = \frac{1}{2}\begin{pmatrix} X+Y & X-Y \\ X-Y & X+Y \end{pmatrix} \tag{4.42}$$

where

$$\begin{aligned} X &= (S+T)^{-1/2} \\ Y &= (S-T)^{-1/2} \end{aligned} \tag{4.43}$$

Using (4.42), (4.43) we see from (4.32) that to calculate $\Theta(k_n)$ for n even we must calculate

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

and

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

which can be done using (4.41) as long as $-1/2 < k_n < 1$. As mentioned previously and as discussed in Section 5, $k_n = -1/2$ or 1 corresponds to the case of a singular process and perfect prediction. For n odd, from (4.35) we see that we must calculate

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

which exists and can be calculated using (4.41) as long as $k_n \neq \pm 1$ i.e. as long as we are not dealing with a singular process for which perfect prediction is possible.

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

Theorem 4.4 *For n even we have*

$$\begin{bmatrix} E^*_{t,n-1} \\ \\ F^*_{t,n} \end{bmatrix} = \Sigma(k_n) \begin{bmatrix} \begin{pmatrix} E_{t,n} \\ E_{\delta(\frac{n}{2})t,n} \end{pmatrix}^* \\ F^*_{\gamma^{-1}t,n-1} \end{bmatrix} \tag{4.44}$$

where

$$\Sigma(k_n) \triangleq \begin{pmatrix} I + a(k_n)\cup_* & b(k_n)\cup_* & k_n \cup_* \\ -\frac{k_n}{2}\cup_* & I + c(k_n)\cup_* & b(k_n)\cup_* \\ d(k_n)\cup_* & -\frac{k_n}{2}\cup_* & I + a(k_n)\cup_* \end{pmatrix} \tag{4.45}$$

with

$$a(k) = \frac{\sqrt{1+2k}+1}{2}\sqrt{1-k} - 1 \tag{4.46}$$

$$b(k) = \frac{\sqrt{1+2k}-1}{2}\sqrt{1-k} \tag{4.47}$$

$$c(k) = \frac{\sqrt{1+2k}-(1+k)}{2} \tag{4.48}$$

$$d(k) = -c(k) - k \tag{4.49}$$

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

$$\Sigma(k_n)\,\Sigma^T(k_n) = I \tag{4.50}$$

For n *odd,* $n = 1$

$$\begin{bmatrix} \begin{pmatrix} E_{t,n-1} \\ E_{\delta(\frac{n-1}{2})t,n-1} \end{pmatrix}^{*} \\ F^*_{t,n} \end{bmatrix} = \Sigma(k_n) \begin{bmatrix} E^*_{t,n} \\ F^*_{\gamma^{-1}t,n-1} \end{bmatrix} \tag{4.51}$$

where the scattering matrix

$$\Sigma(k_n) \triangleq \begin{pmatrix} (I - k_n^2 \cup_*)^{1/2} & k_n \cup_* \\ -k_n \cup_* & (I - k_n^2 \cup_*)^{1/2} \end{pmatrix} \tag{4.52}$$

satisfies

$$\Sigma(k_n)\,\Sigma^T(k_n) = I \tag{4.53}$$

For $n = 1$*:*

$$\begin{pmatrix} E^*_{t,0} \\ F^*_{t,1} \end{pmatrix} = \Sigma(k_1) \begin{pmatrix} E^*_{t,1} \\ F^*_{\gamma^{-1}t,0} \end{pmatrix} \tag{4.54}$$

and

$$\Sigma(k_1) = \begin{pmatrix} \sqrt{1-k_1^2} & k_1 \\ -k_1 & \sqrt{1-k_1^2} \end{pmatrix} \tag{4.55}$$

also satisfies

$$\Sigma(k_1)\,\Sigma^T(k_1) = I \tag{4.56}$$

Proof: We begin by solving (4.30) for $\left(E^T_{t,n-1}\ E^T_{\delta(\frac{n}{2})t,n-1}\right)^{*T}$ then by subsituting this into (4.31) we obtain

$$\begin{bmatrix} \begin{pmatrix} E_{t,n-1} \\ E_{\delta(\frac{n}{2})t,n-1} \end{pmatrix}^* \\ F^*_{t,n} \end{bmatrix} = \hat{\Sigma}(k_n) \begin{bmatrix} E^*_{t,n} \\ E^*_{\delta(\frac{n}{2})t,n} \\ F^*_{\gamma^{-1}t,n-1} \end{bmatrix} \tag{4.57}$$

where

$$\hat{\Sigma}(k_n) \triangleq \begin{bmatrix} \Theta^{-1}(k_n) & k_n \begin{pmatrix} \mathsf{U}_* \\ \mathsf{U}_* \end{pmatrix} \\ \Theta(k_n) \begin{pmatrix} -k_n\mathsf{U}_* & I \\ -k_n\mathsf{U}_* & 0 \end{pmatrix} \Theta^{-1}(k_n) & \Theta(k_n) \begin{pmatrix} (k_n - k_n^2)\mathsf{U}_* \\ I - k_n^2\mathsf{U}_* \end{pmatrix} \end{bmatrix} \tag{4.58}$$

To obtain the desired relation, we simply drop the calculation of $E^*_{\delta(\frac{n}{2})t,n-1}$ from (4.57). 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 $E^*_{\delta(\frac{n}{2})t,n-1}$ 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 (4.45)–(4.49). That $\Sigma(k_n)$ satisfies (4.50) follows immediately from the fact that the vectors on both sides of (4.44) have identity covariances. The result for n odd, $n > 1$ is obtained in a similar fashion, and the case of $n = 1$ is immediate.

5 Characterization of Autoregressive and Regular Processes

The analysis in the preceding sections allows us to deduce a number of properties of the class of autoregressive isotropic processes. The first result summarizes some immediate consequences which we state without proof:

Proposition 5.1 *If Y_t is an AR(p) isotropic process, then*

1. *The reflection coefficents $k_n = 0$ for $n \geq p+1$, and the forward and backward normalized prediction error vectors $E^*_{t,p}$ and $F^*_{t,p}$ are standard white noise processes (i.e. with unity covariance).*
2. *Let us write the formal power series P_p defined in (3.70) as*

$$P_p = \sum_{\substack{w \in \mathcal{L} \\ w \preceq 0}} p_w \cdot w \tag{5.1}$$

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 $w = w_{\alpha,\beta}\delta\gamma^{-k}$ with

$$w_{\alpha,\beta} \in \{\alpha,\beta\}^* \quad \text{and} \quad |w_{\alpha,\beta}| > \left[\frac{p}{2}\right] - 1 \tag{5.2}$$

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 $[t,-\infty)$. 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.

Figure 5.1 illustrates the cylinder for an AR(2) process. Note that (1) is a generalization of the result in Appendix A that stated that if an isotropic process has its support concentrated on $[t,-\infty)$ it is necessarily AR(1).

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

$$\begin{cases} -\frac{1}{2} < k_n < 1 & \text{for } n \text{ even} \\ -1 < k_n < 1 & \text{for } n \text{ odd} \end{cases} \tag{5.3}$$

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

The proof of this theorem rests on the following which is obtained immediately from the Schur recursions:

Lemma 5.1 *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:*

$$\frac{\partial k_n}{\partial r_m} = 0 \text{ for } n < m \tag{5.4}$$

$$\frac{\partial k_{2n}}{\partial r_{2n}} = \frac{1}{2^{n-1}P_{2n-1}(0)} \neq 0 \tag{5.5}$$

$$\frac{\partial k_{2n+1}}{\partial r_{2n+1}} = \frac{1}{2^{n-1}\left(P_{2n}(0) + \delta^{(n)}P_{2n}(0)\right)} \neq 0 \tag{5.6}$$

where the P_n are the Schur series defined in (3.70)

Proof of Theorem 5.1: 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 stabililty 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

$$\Phi : K = (k_n)_{1 \leq n \leq p} \longrightarrow E(y_t y_s) - E(y_{t'} y_{s'}) \tag{5.7}$$

is identically zero.

The form of the modeling filter shows that Φ is a rational function of K. Thus it is sufficient for us to show that Φ is zero on a nonempty set in $\mathcal{R}^p$. Since we know that $\Phi(K) = 0$ if K is obtained via the transformation Ψ, it is sufficient for us to show that the set of K obtained via the transformation Ψ has a nonempty interior.

Thanks to the form of the Schur recursions we know that Ψ is also a rational function and, thanks to Lemma 5.1, its Jacobian is triangular and always invertible. Thus it is sufficient to show that the set of finite sequences $\{r_n | 0 \leq n \leq p\}$ 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

$$\mathcal{R}(r_0, ..., r_p) \geq 0 \tag{5.8}$$

where $\mathcal{R}(r_0, ..., r_p)$ 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 (e.g. $r_n = \delta_{n0}$) and as a consequence the set of $r_0, \ldots, r_p$ satisfying (5.8) has a nonempty interior.

Finally, the machinery we have developed allows us to characterize the class of regular processes.

Definition 5.1 *An isotropic process Y_t is regular or purely nondeterministic if 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_t in the strict past.*

With the aid of a martingale argument, Y_t is regular if and only if

$$\liminf_{n \to \infty} \lambda_{\inf}(\Sigma_{E,n}) > 0 \tag{5.9}$$

where $\lambda_{\text{inf}}(A)$ denotes the minimum eigenvalue of A. Given the form of Σ_n in (4.13), (4.15), and (4.18), we can deduce that this is equivalent to

$$\liminf_{n\to\infty} \lambda_{\text{inf}}(\Sigma_n) > 0 \tag{5.10}$$

Thanks to the structure of Σ_n determined from (4.38)–(4.40) and the definition of $\Theta^{-1}(k_n)$ in (4.32), (4.35), we can deduce that

$$\lambda_{\text{inf}}(\Sigma_n) = r_0(1-k_1^2)\prod_{i=2}^{n} \lambda_{\text{inf}}\left(\Theta^{-2}(k_i)\right) \tag{5.11}$$

and for i odd, $i > 1$

$$\lambda_{\text{inf}}\left(\Theta^{-2}(k_i)\right) = 1 - k_i^2 \tag{5.12}$$

while for i even

$$\begin{aligned}\lambda_{\text{inf}}\left(\Theta^{-2}(k_i)\right) &= \min\left(1-k_i, 1+k_i-2k_i^2\right) \\ &\sim 1-|k_i| \text{ for } k_i \text{ small}\end{aligned} \tag{5.13}$$

From this we can deduce the following:

Theorem 5.2 *An isotropic process Y_t is regular if and only if its reflection coefficient sequence is such that*

$$\sum_{n=1}^{\infty}\left(k_{2n-1}^2 + |k_{2n}|\right) < \infty \tag{5.14}$$

6 Conclusion

In 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 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. Finally we have justified our claim that the reflection coeffients are a good paramtrization for AR processes and isotropic processes in general by showing that AR processes can be uniquely specified by these coefficients and the regularity of an isotropic process can be characterized in terms of its reflection coefficient sequences.

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 counterparts to be useful in analyzing signals displaying fractal-like or self-similar characteristics, i.e. random signals whose behavior is similar at multiple scales. Figure 6.1 illustrates a sample of an AR(1) process with $k_1 = 0.99$ which displays this self-similar behavior.

References

[1] 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.

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

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

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

[6] R. KRONLAND-MARTINET, J. MORLET, A. GROSSMANN, "Analysis of sound patterns through wavelet transforms," *Intl. J. of Pattern Recognition and Artificial Intelligence*, Special Issue on Expert Systems and Pattern Analysis, 1987.

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

[8] 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.

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

[10] R. KRONLAND-MARTINET, Proc. Colloque "Ondelettes," Marseille, Dec. 1987.

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

[12] 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.

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

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

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

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

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

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

[19] I. DAUBECHIES, "The wavelet transform, time-frequency localization and signal analysis," AT&T Bell Laboratories Report.

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

[21] 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.

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

[24] S.G. MALLAT, "A compact multiresolution representation: the wavelet model," Dept. of Computer and Info. Science—U. of Penn., MS–CIS–87–69, GRASP LAB 113, Aug. 1987.

[25] S.G. MALLAT, "A theory for multiresolution signal decomposition: the wavelet representation," Dept. of Computer and Info. Science—U. of Penn., MS–CIS–87–22, GRASP LAB 103, May 1987.

[26] S.G. MALLAT, "Multiresolution approximation and wavelets," Dept. of Computer and Info. Science—U. of Penn., MS–CIS–87–87, GRASP LAB 80, Sept. 1987.

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

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

[29] Y. MEYER, "L'analyse par ondelettes," *Pour la Science*, Sept. 1987.

[30] D. PADDON AND H. HOLSTEIN, eds., *Multigrid Methods for Integral and Differential Equations*, Clarendon Press, Oxford, England, 1985.

[31] 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 $\{\gamma^{-n}t | n \geq 0\}$, 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

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

B The Relation Among the Parameters of AR(2)

Consider the second-order model (2.45) 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 (2.45) we obtain the relation

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

while from (2.46) we find

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

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

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

Next consider multiplying (2.45) by each of the following: Y_t, $Y_{\delta t}$, $Y_{\gamma^{-1}t}$, $Y_{\gamma^{-2}t}$. We take expectations using (B.1), (B.2) and the fact that $E(Y_{\gamma^{-1}t} W_t) = E(Y_{\gamma^{-2}t} W_t) = 0$ (since we are solving the AR equations "casually"—see (2.47), (2.48)). Assuming that Y is isotropic, we obtain the following *four* linear equations in the *three* unknowns r_0, r_1, r_2:

$$\left\{ \begin{aligned} r_0 &= a_1 r_1 + a_2 r_2 + a_3 r_2 + \sigma \tfrac{a_0}{0^2 - a_3^2} \\ r_1 &= a_1 r_0 + a_2 r_1 + a_3 r_1 \\ r_2 &= a_1 r_1 + a_2 r_0 + a_3 r_2 \\ r_2 &= a_1 r_1 + a_2 r_2 + a_3 r_0 + \sigma \tfrac{3}{0^2 - a_3^2} \end{aligned} \right. \tag{B.4}$$

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 (2.45) by any Y_{wt}, $w \preceq 0$, $|w| > 2$ 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 3 without proof—are the constraints of isotropy and the construction of specific isometries.

C.1 Proof of Lemma 3.2

Let

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

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

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

Define the set of nodes

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

The points $w\gamma^{-1}t$ in (C.2) correspond to the points $s = vt$ 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: $\mathcal{T}_{t,n} \rightarrow \mathcal{T}_{t,n}$such that

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

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 $\phi(x) = t$ 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 note that the words w to be considered consist of

$$W = \bigcup_{p=\frac{n}{2}}^{n-1} W_p \tag{C.5}$$

$$W_{n-1} = \{\gamma^{-n+1}\} \quad , \quad W_{n-2} = \{\delta\gamma^{-n+3}\}$$
$$W_p = \{\alpha, \beta\}^{n-p-2}\delta\gamma^{-p+1}, \ \frac{n}{2} \leq p \leq n-3 \tag{C.6}$$

where

$$\{\alpha, \beta\}^k = \{m \in \{\alpha, \beta\}^* \, ||m| = k\} \tag{C.7}$$

We now describe a set of maps:

1. ϕ_{n-1} interchanges W_{n-1} and W_{n-2} and leaves the rest of $\mathcal{T}_{t,n}$ fixed. That is

$$\begin{aligned} \phi_{n-1}(\gamma^{-n}t) &= \delta\gamma^{-n+2}t \\ \phi_{n-1}(\delta\gamma^{-n+2}t) &= \gamma^{-n}t \\ \phi_{n-1}(s) &= s \text{ for all other } s \in \mathcal{T}_{t,n} \end{aligned} \tag{C.8}$$

2. For $\frac{n}{2} \leq p \leq n-2$, ϕ_p interchanges W_{p-1} and $\bigcup_{q=p}^{n-1} W_q$. Specifically for any such p, ϕ_p makes the following interchanges, leaving the remaining points in $\mathcal{T}_{t,n}$ fixed:

$$\begin{aligned} \delta\gamma^{-p+1}t &\leftrightarrow \gamma^{-p-1}t \\ \alpha^k\delta\gamma^{-p+1}t &\leftrightarrow \gamma^{-p-k-1}t, \ 1 \leq k \leq n-p-1 \\ m_{\alpha\beta}\beta\alpha^k\gamma^{-p+1}t &\leftrightarrow m_{\alpha\beta}\delta\gamma^{-p-k+2}t, \\ & \quad 1 \leq k \leq n-p-2, \ 0 \leq |m_{\alpha\beta}| \leq n-p-k \end{aligned}$$

3. For each $\frac{n}{2} \leq p \leq n-3$ and any $1 \leq k \leq n-p-2$, $\phi_{p,k}$ interchanges points in W_p. Specifically, for any such p and k, ϕ_{pk} makes the following interchanges,

leaving the remaining points in $\mathcal{T}_{t,n}$ fixed:

$$m^2_{\alpha\beta} m^1_{\alpha\beta} \delta \gamma^{-p} t \quad \leftrightarrow \quad m^2_{\alpha\beta} \delta m^1_{\alpha\beta} \delta \gamma^{-p} t$$
$$|m^1_{\alpha\beta}| = k \quad , \quad 0 \leq |m^2_{\alpha\beta}| \leq n - k - p - 2$$

A straightforward, if somewhat tedious computation, verifies that (i) these are all local isometries leaving t fixed and $\mathcal{T}_{t,n-1}$ invariant, and (ii) for any w', w'' in W an isometry satisfying (C.4) can be constructed by composing one or more of the isometries in (1) – (3). This completes the proof of Lemma 3.2.

C.2 Proof of Lemma 3.3

Let

$$H_{t,m}(w, w') = E[F_{\gamma^{-1}t,n-1}(w) E_{t,n-1}(w')] \tag{C.9}$$

Where n is even $|w| = n - 1$, $w \preceq 0$ and $|w'| < n$, $w' \asymp 0$. 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 $|w| = n - 1$, $w \preceq 0$, $\phi(w_1) = w_2$, $\phi(w_2) = w_1$, ϕ leaves $\mathcal{T}_{t,n-1}$ invariant *and* leaves fixed any point of the form $w't$, with $|w'| < n$, $w' \asymp 0$.

2. For any w'_1, w'_2 satisfying $|w'| < n$, $w' \asymp 0$, $\psi(w'_1) = \psi(w'_2)$, ψ leaves $\mathcal{T}_{\gamma^{-1}t,n-1}$ invariant *and* leaves fixed any point of the form $w\gamma^{-1}t$, with $|w| = n - 1$, $w \preceq 0$.

It is straightforward to check that the isometrics ϕ_{n-1}, ϕ_p, $\phi_{p,k}$ and their compositions form a class satisfying (1). To construct the second class, let us recall the representation and ordering of the words w for which $|w| < n$, $w \asymp 0$ (see (3.8) (3.9)), and let w_m denote the mth of these with respect to this ordering, where $0 \leq m \leq 2^{\frac{n}{2}-1} - 1$. We then define the following maps:

- For each $1 \le k \le \frac{n}{2} - 1$ and each $0 \le r \le 2^{\frac{n}{2}-k-1} - 1$, ψ_{kr} makes the following interchanges, leaving the remaining points in $\mathcal{T}_{t,n}$ fixed:

$$\gamma^{-j} w_{r2^k} t \leftrightarrow \delta^{(k-j)} \gamma^{-j} w_{r2^k} t, \; 0 \le j \le k-1$$

Again it is a straightforward computation to check that (i) each such ψ_{kr} is a local isometry (so that it can be extented to a full isometry); (ii) ψ_{kr} leaves $\mathcal{T}_{\gamma^{-1}t,n-1}$ invariant and leaves fixed any point of the form $w\gamma^{-1}t$, with $|w| = n-1$, $w \preceq 0$; and (iii) for any w_1', w_2' satisfying $|w'| < n$, $w' \asymp 0$, we can construct ψ as in (2) as a composition of one or more of the ψ_{kr}. This completes the proof of Lemma 3.3.

C.3 Proof of Lemma 3.4

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

$$J_{t,n}(w_i, w_j) = E[E_{t,n}(w_i) E_{t,n}(w_j)] \tag{C.10}$$

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

$$w_i t \leftrightarrow w_i' t, \; w_j t \leftrightarrow w_j' t$$

Direct calculations shows that compositions of the ψ_{kr} defined in the previous subsection yields such a local isometry.

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

$$J_{t,n}(w_i, w_j) = J_{t,n}(0, w_k), \text{ where } |w_k| = s \tag{C.11}$$

Again an appropriate composition of the ψ_{kr} yields an isometry leaving $\mathcal{T}_{\gamma^{-1}t,n-1}$ invariant and performing the interchange

$$w_i t \leftrightarrow t,\ w_j t \leftrightarrow w_k t \tag{C.12}$$

Which finishes the proof of Lemma 3.4.

C.4 Proof of Lemma 3.6

We wish to show that (3.53) holds for n odd. Consider the first equality in (3.53). As before, an argument using the isotropy of Y_t shows that this equality will follow if we can construct a local isometry, this time of $\mathcal{T}_{t,n+1}$ which leaves $\mathcal{T}_{\gamma^{-1}t,n-1}$ invariant and which interchanges the sets

$$\{w_m t | 0 \leq m \leq 2^{\frac{n-1}{2}-1} - 1\} \tag{C.13}$$

and

$$\{w_m t | 2^{\frac{n-1}{2}-1} \leq m \leq 2^{\frac{n+1}{2}-1} - 1\} \tag{C.14}$$

where as in Section C.2, the w_m are the ordered words such that $|w| < n+1$, $w \asymp 0$. The isometry $\psi_{\frac{n-1}{2},0}$ (defined as in Section 3.2 but with n replaced by $n+1$) has the desired properties.

Consider now the second equality in (3.53). In this case we must construct an isometry that again leaves $\mathcal{T}_{\gamma^{-1}t,n-1}$ invariant and which interchanges the set in (C.13) and the set

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

The following local isometry ϕ has the desired property. Each element s of $\mathcal{T}_{t,n+1}$ can be written uniquely in the form

$$s = m_{\alpha,\beta}\gamma^{-\frac{n+1}{2}-p}t \tag{C.15}$$

where

$$-\frac{n+1}{2} \leq p \leq \frac{n+1}{2} \tag{C.16}$$

$$|m_{\alpha,\beta}| + \frac{n+1}{2} + p \leq n+1 \tag{C.17}$$

The desired isometry is then, in essence, a reflection: for s as in (C.15)

$$\phi(s) = m_{\alpha,\beta}\gamma^{-\frac{n+1}{2}+p}t \tag{C.18}$$

which completes the proof of Lemma 3.6.

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

From (3.35) and (4.42) 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

$$\Sigma_+ = \Sigma(\alpha_0, \ldots, \alpha_{k-1}) + \alpha_k \mathsf{U}_{k-1} = \Sigma(\alpha_0 + \alpha_k, \ldots, \alpha_{k+1} + \alpha_k) \quad \text{(D.1)}$$

$$\Sigma_- = \Sigma(\alpha_0, \ldots, \alpha_{k-1}) - \alpha_k \mathsf{U}_{k-1} = \Sigma(\alpha_0 - \alpha_k, \ldots, \alpha_{k-1} - \alpha_k) \quad \text{(D.2)}$$

If we introduce the following notation

$$\text{Bloc}\,(X, Y) = \frac{1}{2} \begin{pmatrix} X+Y & X-Y \\ X-Y & X+Y \end{pmatrix} \quad \text{(D.3)}$$

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

$$\Sigma^{-1/2}(\alpha_0, \ldots, \alpha_k) = \begin{cases} \alpha_0^{-1/2} & \text{if } k = 0 \\ \text{Bloc}\left(\Sigma_+^{-1/2}, \Sigma_-^{-1/2}\right) & \text{if } k \geq 1 \end{cases} \quad \text{(D.4)}$$

which involves a sequence scalar calculations.

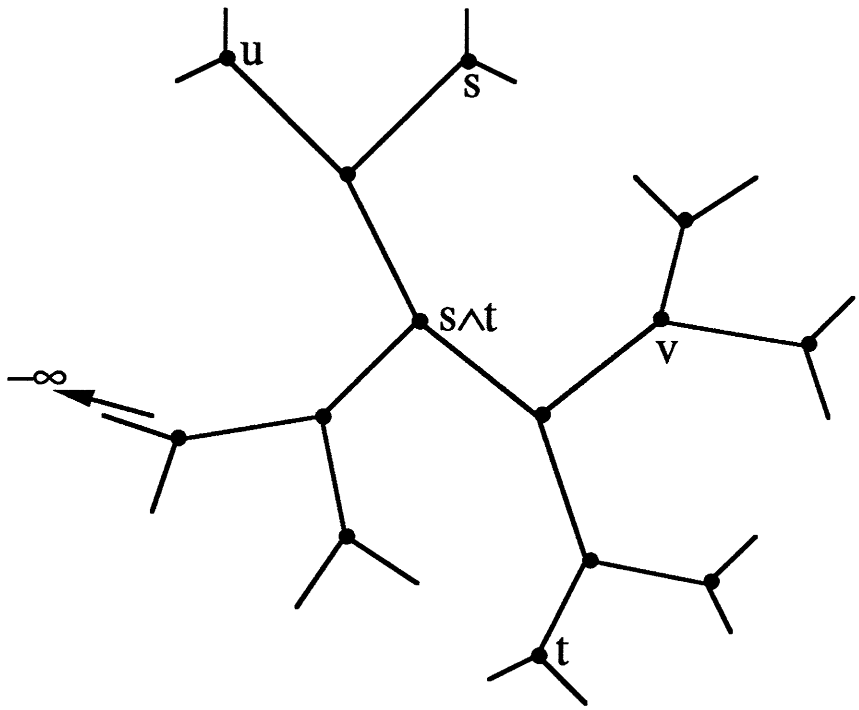


Figure 2.1

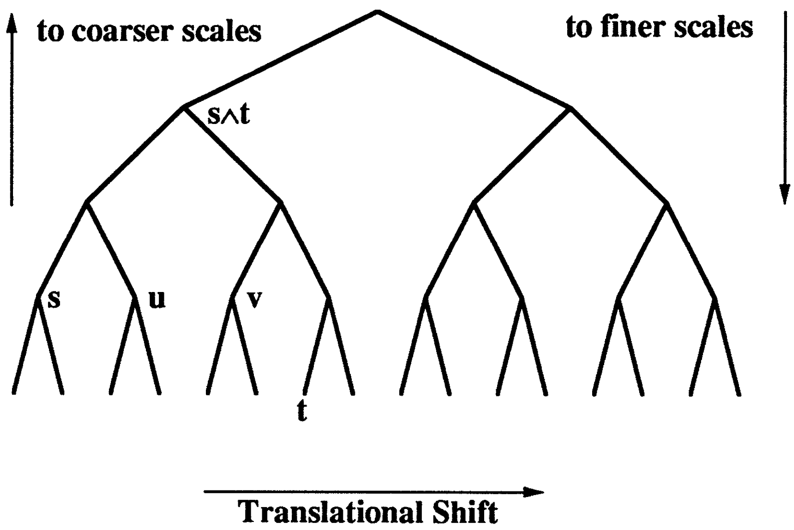

to coarser scales
to finer scales
s∧t
s
u
v
t
Translational Shift

Figure 2.2

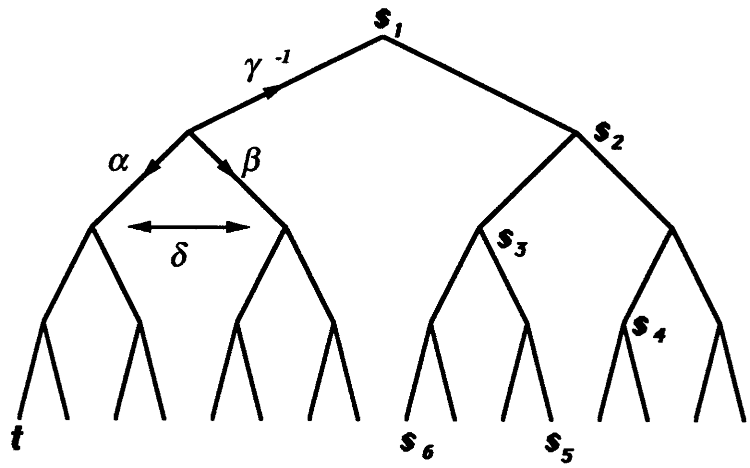
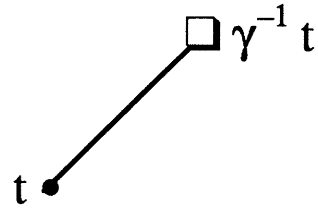


Figure 2.3

- n = 1



- n = 2

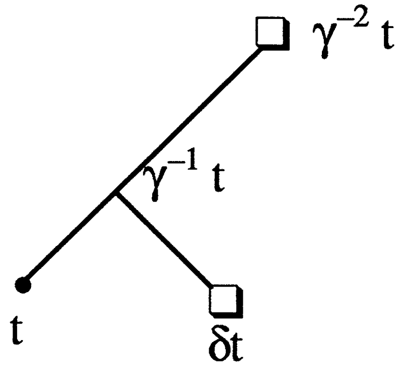


Figure 3.1: Illustrating $E_{t,n}$ (dots) and $F_{t,n}$ (squares) for n=1,2

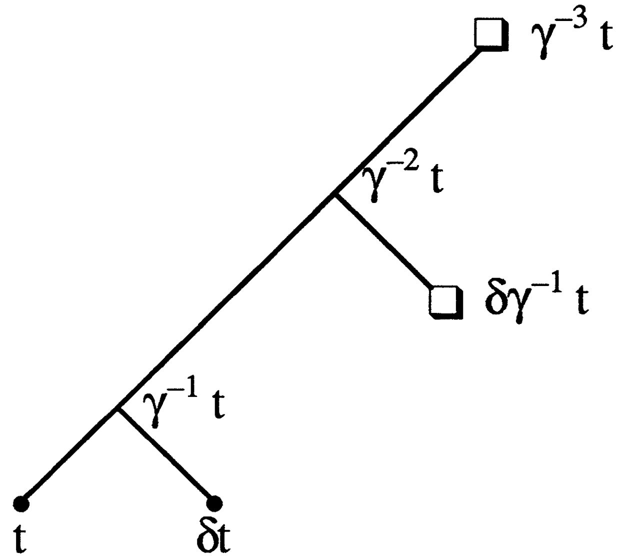


Figure 3.2: Illustrating $E_{t,3}$ (dots) and $F_{t,3}$ (squares)

- n = 4

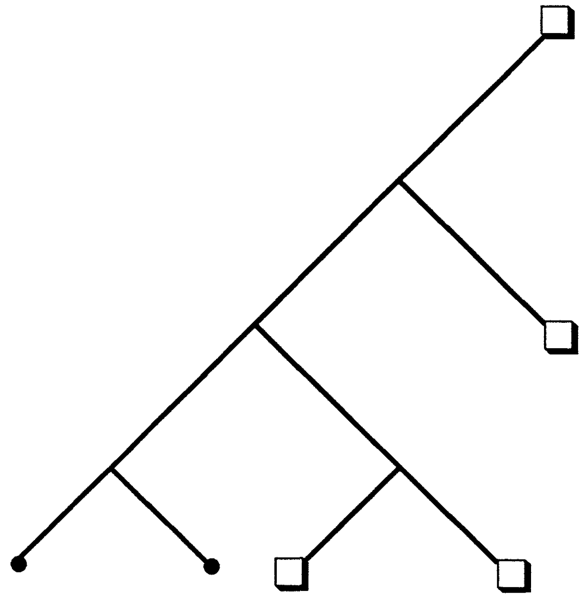


Figure 3.3: Illustrating $E_{t,4}$ and $F_{t,4}$

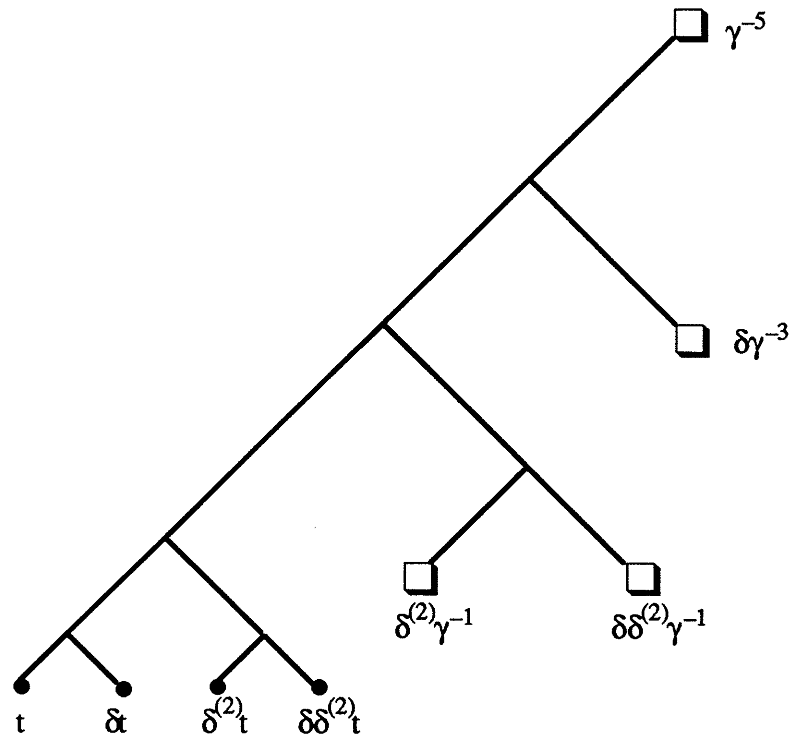


Figure 3.4: Illustrating $E_{t,5}$ and $F_{t,5}$

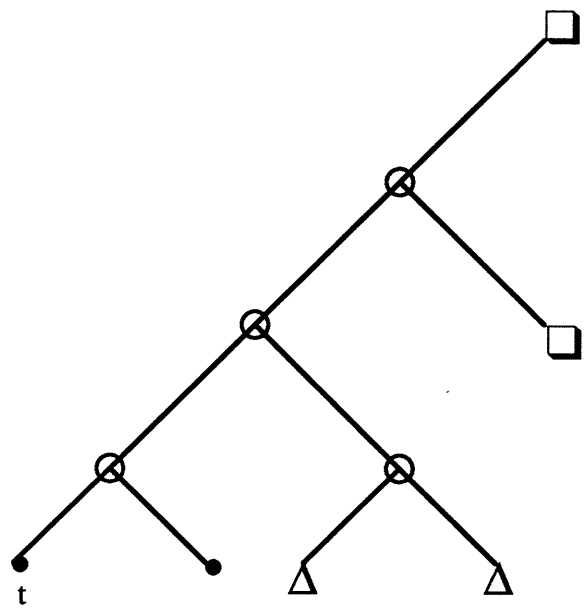


Figure 3.5: Illustrating $E_{t,3}$ (dots), $F_{\gamma^{-1}t,3}$ (squares), $E_{\delta^{(2)}t,3}$ (triangles), and $\mathcal{Y}_{\gamma^{-1}t,2}$ (circles)

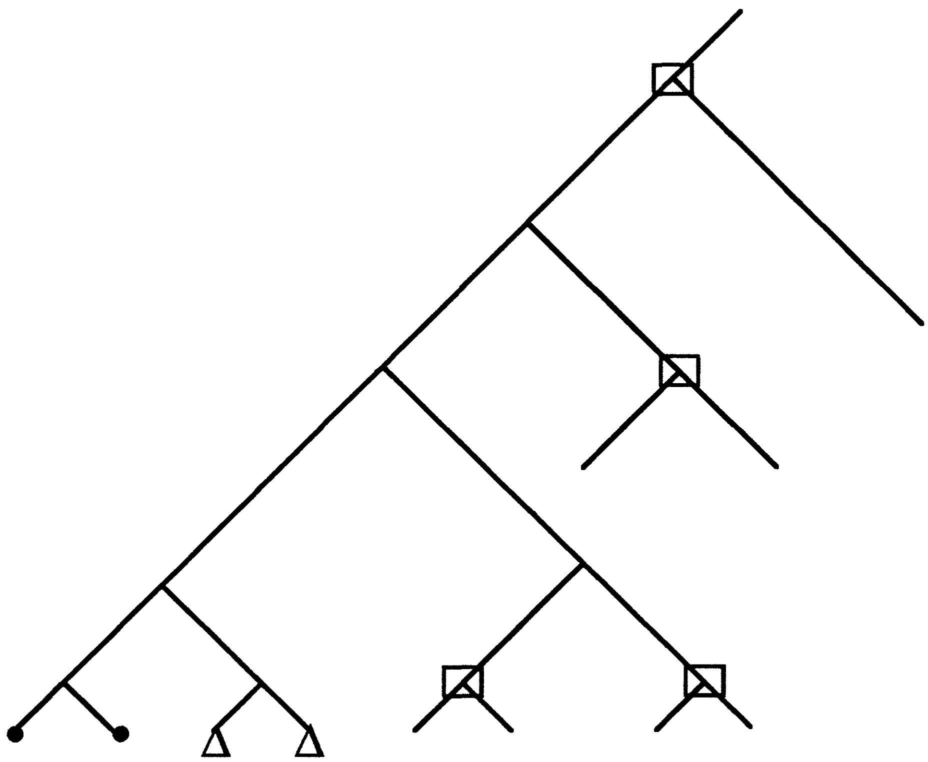


Figure 3.6: Illustrating $E_{t,4}$ (dots), $F_{\gamma^{-1}t,4}$ (squares), $E_{\delta^{(2)}t,4}$ (triangles)

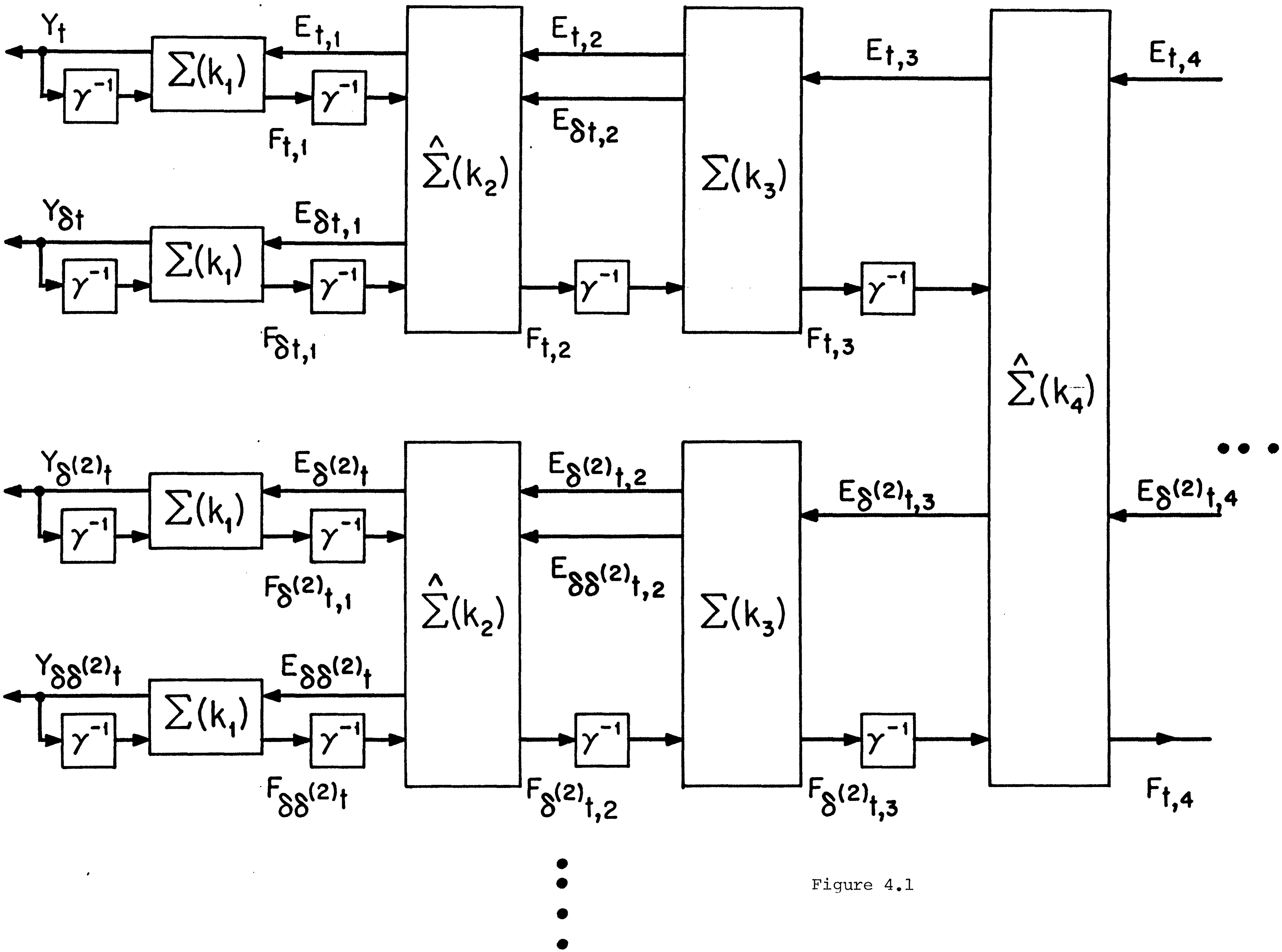
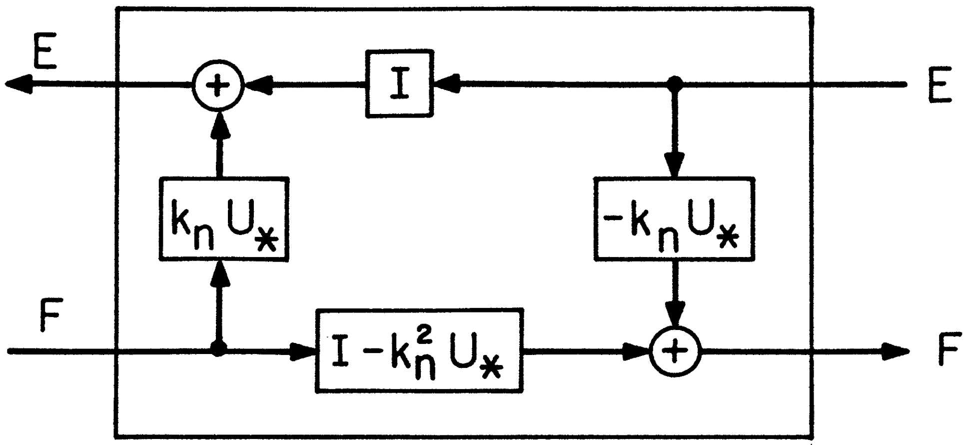
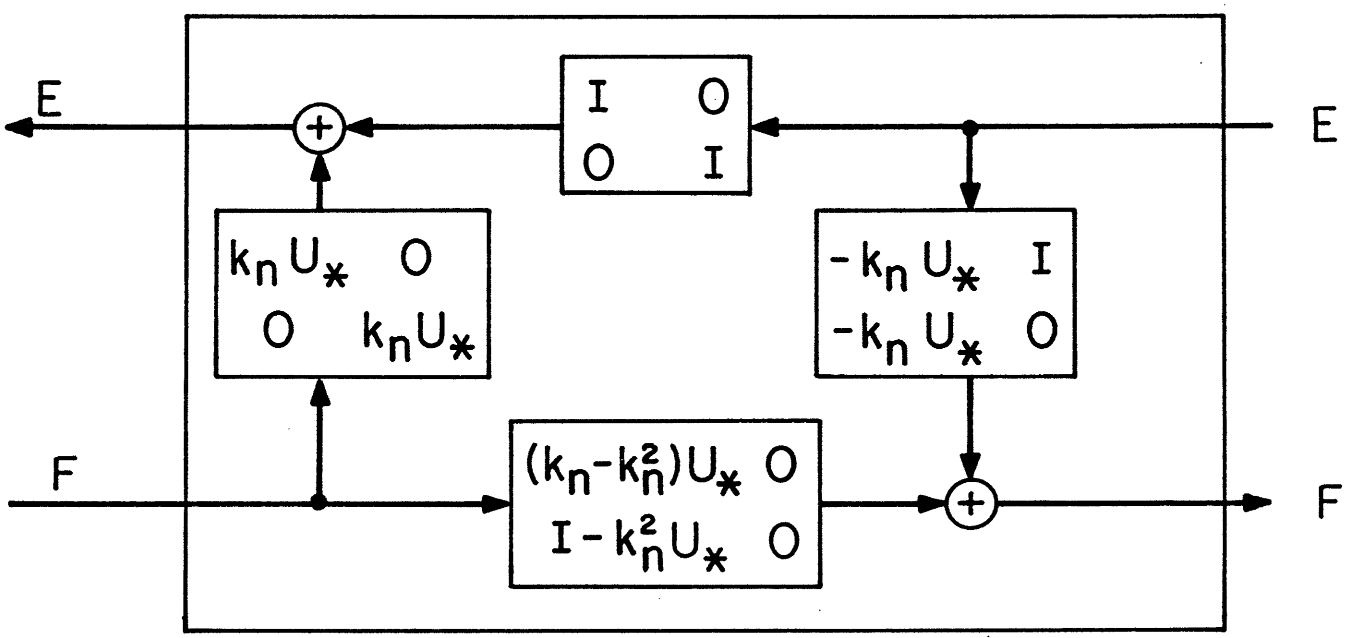

Y_t
Σ(k_1)
γ^-1
E_t,1
F_t,1
Y_δt
E_δt,1
F_δt,1
Y_δ(2)t
E_δ(2)t
F_δ(2)t,1
Y_δδ(2)t
E_δδ(2)t
F_δδ(2)t
Σ̂(k_2)
E_t,2
E_δt,2
F_t,2
E_δ(2)t,2
E_δδ(2)t,2
F_δ(2)t,2
Σ(k_3)
E_t,3
F_t,3
E_δ(2)t,3
F_δ(2)t,3
Σ̂(k_4)
E_t,4
E_δ(2)t,4
F_t,4

Figure 4.1



(a)



(b)

Figure 4.2: Scattering Blocks of Figure 4.1 for (a) n odd; and (b) n even

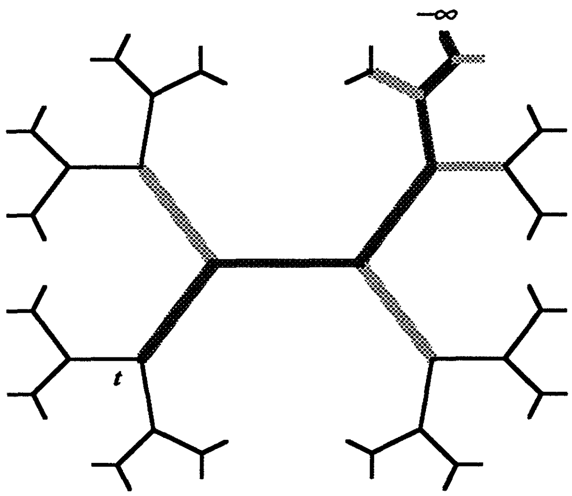


Figure 5.1: Cylinder of radius 1, the support of the impulse response of AR(2)

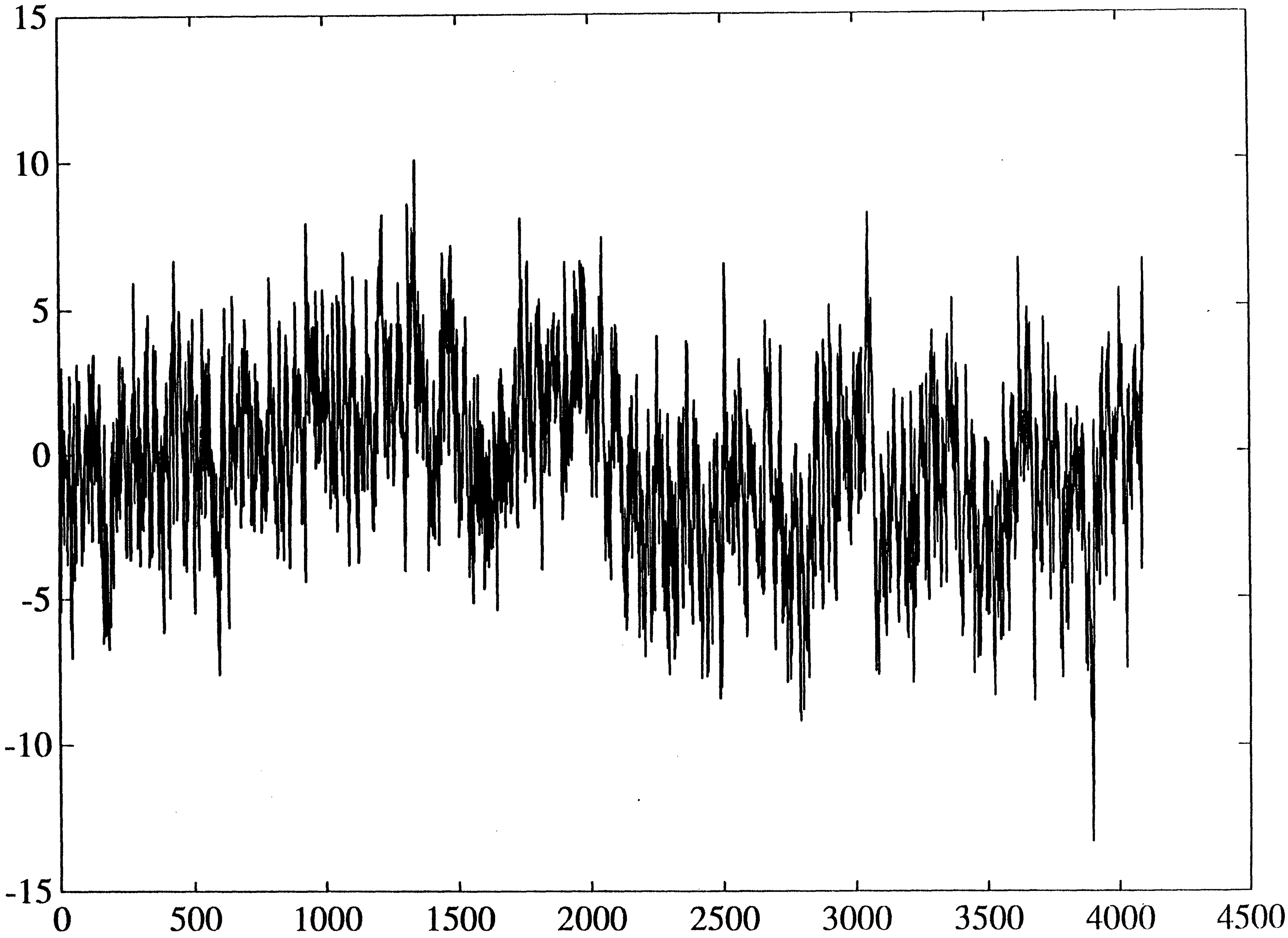


Figure 6.1