A Generalized Levinson Algorithm for Covariance Extension With Application to Multiscale Autoregressive Modeling

Austin B. Frakt, Hanoch Lev-Ari, *Senior Member, IEEE,* and Alan S. Willsky, *Fellow, IEEE*

***Abstract*—Efficient computation of extensions of banded, partially known covariance matrices is provided by the classical Levinson algorithm. One contribution of this paper is the introduction of a generalization of this algorithm that is applicable to a substantially broader class of extension problems. This generalized algorithm can compute unknown covariance elements in any order that satisfies certain graph-theoretic properties, which we describe. This flexibility, which is *not* provided by the classical Levinson algorithm, is then harnessed in a second contribution of this paper, the identification of a multiscale autoregressive (MAR) model for the maximum-entropy (ME) extension of a banded, partially known covariance matrix. The computational complexity of MAR model identification is an order of magnitude below that of explicitly computing a full covariance extension and is comparable to that required to build a standard autoregressive (AR) model using the classical Levinson algorithm.**

***Index Terms*—Chordal graphs, covariance completion, covariance extension, Levinson's algorithm, maximum entropy, multiscale autoregressive (MAR) models, reflection coefficients.**

I. Introduction

COVARIANCE extension is a classical and important problem in statistics. Perhaps the simplest instantiation is the problem of inferring or characterizing unknown autocorrelation values from a finite number of consecutive known ones [2]. That is, given knowledge only of the diagonal bands of a covariance matrix,[1] what are all possible valid values for the unknown elements? Moreover, how can these elements be computed efficiently? Answers to these questions are given by the classical Levinson algorithm [1], [26], [51]. All possible extensions are parameterized by so-called *reflection coefficients* and any extension can be computed efficiently using the Levinson recursion. One contribution of this paper is a generalization of Levinson's algorithm that is applicable to a substantially broader class of extension problems.

Among all possible extensions of a partially known covariance matrix is the *maximum-entropy* (ME) extension, the study of which has received a great deal of attention. In particular, much has been written about its applications in spectral estimation [2], [3], [32], [46], [51] and very large scale integration (VLSI) modeling [12], its connection with autoregressive (AR) (all-pole) models [1], [24], [29], [41], [51], and its link to lattice structures for finite impulse response filters [20], [26], [39]. A well-known fact about the ME extension of a partially known, banded covariance matrix is that its *inverse* is a banded matrix with the same bandwidth [25], [41]. This fact and the correspondence between inverse-covariance zeros and conditional decorrelation [11], [50] implies that the ME extension of a banded, partially known covariance matrix is wide-sense Markov[2] of order given by the bandwidth.

It is well known that an AR time-series model for the ME extension of a banded, partially known covariance matrix can be computed efficiently using Levinson's algorithm [1], [26], [51]. Such a model is an *implicit* and compactly parameterized characterization of the extension. As mentioned, Levinson's algorithm can also be used to compute the ME extension *explicitly*. In doing so, Levinson's algorithm determines unknown covariance elements one diagonal band at a time, working outward from the main diagonal. An important open question is: what are the other possible orders in which covariance elements may be computed with the same efficiency as Levinson's algorithm? This question is not merely an academic one as our original motivation for studying it is the problem of multiscale autoregressive (MAR) model [4], [5] identification from incomplete covariance information.

MAR models generalize state-space models of time series, evolving in *scale* rather than in time. They have been effectively applied to a wide variety of signal and image processing problems [8], [9], [13]–[17], [27], [30], [34]–[38], [42], [48], [49] and their success stems, in part, from the efficiency of the statistical inference algorithms to which they lead [4], [44]. Recent approaches to the MAR model identification problem [10], [18], [19], [31] rely on *complete* knowledge of the second-order statistics of the process to be modeled. This represents a significant limitation because, for large, real-world problems, such complete knowledge is unlikely and impractical.

Manuscript received February 20, 2001; revised October 6, 2002. This work was supported by the Office of Naval Research under Grant N00014-00-1-0089 and by DoD MURI under Grant DAAD19-00-1-0466 through the Army Research Office.

A. B. Frakt is with Health Services Research and Evaluation, Abt Associates, Inc., Cambridge, MA 02138-1168 USA (e-mail: austin_frakt@abtassoc.com).

H. Lev-Ari is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: levari@ece.neu.edu).

A. S. Willsky is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: willsky@mit.edu).

Communicated by J. A. O'Sullivan, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2002.807315

[1]It is assumed that the known bands are adjacent to and include the main diagonal.

[2]Hereafter, wide-sense is assumed.

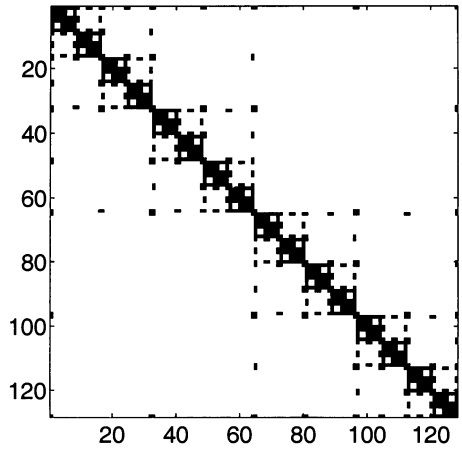

Fig. 1. Black indicates the entries needed for building a MAR model for the ME extension of a 128×128, tri-diagonal, partially known covariance matrix.

Suppose instead that the information available is a banded, partially known covariance matrix, so that its ME extension is Markov. As shown in [43], any Markov process has a MAR representation. This immediately leads to the question of computing MAR model parameters directly from the partial covariance matrix, much as the Levinson algorithm computes the parameters of an AR time series model for this same Markov process. However, as we review in Section II, while constructing this MAR model requires explicit calculation of only $O(N)$ elements of the full $N \times N$ ME extension (the same complexity as for the AR model for a nonstationary Markov process), the location of those elements in the ME extension is highly nonstandard; some appear in bands arbitrarily far from the main diagonal. For instance, to build a MAR model for the ME extension of a 128×128, tri-diagonal, partially known covariance matrix requires determining only the elements shown in black in Fig. 1. In principle, one could use Levinson's algorithm to compute a full extension and then simply extract the required elements. However, this is computationally wasteful since only $O(N)$ out of the total of N^2 covariance elements are needed. Using our generalized Levinson algorithm, we will show how to compute *only* the $O(N)$ that are required for MAR model identification. A consequence is that the computational complexity of building a MAR model is comparable to that which is obtained by the classical Levinson algorithm in designing an AR time-series model.

The remainder of this paper is organized as follows. Our motivating application—building MAR models from incomplete second-order characterizations of random processes—is introduced in Section II. Some background concepts upon which we will rely are provided in Section III. Our generalized Levinson algorithm is presented in Section IV, and its application to MAR model identification is made in Section V. Closing remarks are made in Section VI.

II. Motivating Application

In this section, we review the aspects of MAR models that are of relevance to this paper and introduce the problem of MAR model identification from incomplete covariance

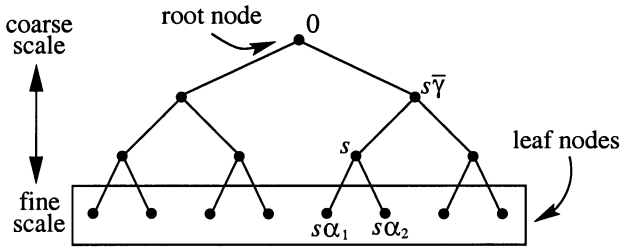

Fig. 2. A dyadic tree. The root node is indexed by $s = 0$. The parent of node s is denoted $s\overline{\gamma}$. The children of node s are labeled from left to right by $s\alpha_1$, $s\alpha_2$.

information. Although we restrict our attention to one-dimensional processes, the MAR framework is applicable in the multidimensional case as well. MAR models, first introduced in [4], [5], provide multiscale representations of a random phenomenon. Each such model is a collection of random vectors $\{x(s)\}$, called *states*, indexed by the nodes of a tree (see Fig. 2) with dynamics given by

$$x(s) = A(s)x(s\overline{\gamma}) + w(s) \tag{1}$$

where s is a node of a tree, $s\overline{\gamma}$ is its parent, $A(s)$ is a matrix of appropriate dimension, and $w(s)$ has covariance $Q(s)$, is uncorrelated from node to node, and is uncorrelated with $x(0)$, the root-node state which initializes the dynamics. For future reference, we also note that $s\alpha_1$ and $s\alpha_2$ are the left and right children, respectively, of node s, as indicated in Fig. 2.

A MAR process can be used to model a finite-length, one-dimensional random signal in the following way. We start with a vector-valued, zero-mean,[3] one-dimensional random signal $f = \{f(i); 1 \leq i \leq N\}$ of length N, with covariance $P = E\{ff^T\}$. We then construct a dyadic tree in which the random variables $\{f(i)\}$ are represented at the leaf nodes of the tree so that

$$f = \{x(s) | s \text{ is a leaf node}\}. \tag{2}$$

We then construct a MAR model for f by defining coarse-scale states. These coarse-scale states are "hidden" in the sense that, unlike the leaf-node states, their statistics are not known *a priori*.

Determining the coarse-scale state statistics and, using these, determining the MAR model dynamics is vastly simplified by considering so-called *internal* models [18], [19]. Internal MAR models have the additional property that each coarse scale state $x(s)$ is a linear function of the fine-scale process f. That is, $x(s) = L_s f$ for all nodes s and some set of matrices $\{L_s\}$.

Internal MAR models are useful for theoretical and practical reasons. As explained in great depth in [18], [19], internal models guarantee model consistency. Internal models are also useful in data fusion applications [8] because they permit the consistent inclusion of nonlocal linear functions of f at coarser scale nodes. This allows the optimal fusion of multiresolution measurements using the efficient MAR estimation algorithm [4].

[3]We assume, without loss of generality, that all processes considered are zero-mean.

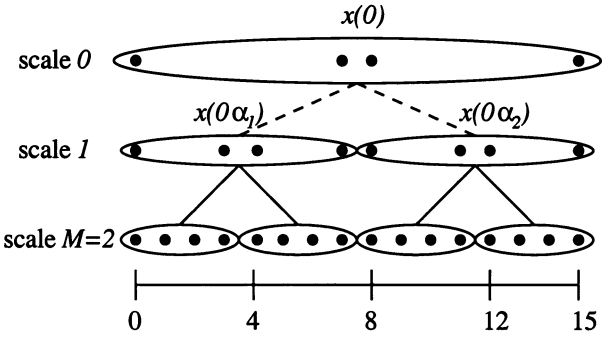

Fig. 3. MAR model for 16 samples of a first-order Markov process. Each ellipse includes the samples for a MAR state.

For internal models, the parameters[4] $A(\cdot)$, $Q(\cdot)$, and $P_{x(0)}$ are easy to determine from $\{L_s\}$ and P. To see this, notice that the MAR dynamics of (1) imply that the linear least squares (LLS) estimate of $x(s)$ from $x(s\overline{\gamma})$ is $A(s)x(s\overline{\gamma})$, and that $w(s) = x(s) - A(s)x(s\overline{\gamma})$ is the corresponding estimation error. Therefore,

$$A(s) = P_{x(s)x(s\overline{\gamma})}P^{-1}_{x(s\overline{\gamma})} \tag{3a}$$

$$Q(s) = P_{x(s)} - P_{x(s)x(s\overline{\gamma})}P^{-1}_{x(s\overline{\gamma})}P^{T}_{x(s)x(s\overline{\gamma})} \tag{3b}$$

where $P_{x(s)}$ is the covariance matrix for the state $x(s)$, and $P_{x(s)x(s\overline{\gamma})}$ is the cross-covariance matrix between the child state $x(s)$ and the parent state $x(s\overline{\gamma})$. For an internal MAR model, the state covariances and child–parent cross covariances are given by

$$P_{x(s)} = L_s P L_s^T \tag{4a}$$

$$P_{x(s)x(s\overline{\gamma})} = L_s P L_{s\overline{\gamma}}^T. \tag{4b}$$

We now turn to the MAR models for Markov processes discussed in [43]. As an illustration, consider a MAR model for a first-order Markov process f. The MAR model depicted in Fig. 3 is an *endpoint* model for this Markov process (the tree branches from $x(0)$ to $x(0\alpha_1)$ and $x(0\alpha_2)$ are indicated as dashed lines for the purposes of a subsequent discussion). Each state consists of four endpoints corresponding to two intervals. For instance,

$$x(0) \triangleq [\, f(0) \quad f(7) \quad f(8) \quad f(15)\,]^T \tag{5a}$$

$$x(0\alpha_1) \triangleq [\, f(0) \quad f(3) \quad f(4) \quad f(7)\,]^T \tag{5b}$$

$$x(0\alpha_2) \triangleq [\, f(8) \quad f(11) \quad f(12) \quad f(15)\,]^T \tag{5c}$$

and the other states are similarly defined as indicated in Fig. 3. Notice that $x(0)$ consists of endpoints of the intervals $[0, 7]$ and $[8, 15]$. Therefore, conditioned on $x(0)$, the sets of samples $\{f(i)\}_{i\in[0,\,7]}$ and $\{f(i)\}_{i\in[8,\,15]}$ are uncorrelated. It follows that $x(0\alpha_1)$ and $x(0\alpha_2)$ are uncorrelated when conditioned on $x(0)$. Hence, the statistical relationships among $x(0)$, $x(0\alpha_1)$, and $x(0\alpha_2)$ can be captured *exactly* by a MAR process. A similar argument applies to the other states of Fig. 3 and to the states of any MAR model for a first-order Markov process of length N.

[4]We use the notation P_x to denote the covariance matrix for random vector x and P_{xy} to denote the cross-covariance matrix for random vectors x and y.

Since the MAR states for a first-order Markov process consist of four endpoints, the L_s matrices are sparse, $4 \times N$ *selection matrices*.[5] Consequently, the covariances $P_{x(s)}$ and $P_{x(s)x(s\overline{\gamma})}$ are both 4×4, and each contains only a small number (independent of N) of correlations between fine-scale variables. Note that the MAR model based on these L_s matrices is internal by construction (further elaboration on this point and a formal description of the necessary and sufficient conditions for internality are found in [18], [19]).

The MAR model just discussed can be generalized to any length-N, order-k Markov process. Roughly speaking, to adapt the preceding approach to a general k, simply replace every sample (dot, •) in Fig. 3 with k consecutive samples corresponding to selection matrices $\{L_s\}$ that are $4k \times N$. Let us denote by $\eta(s)$ the index set characterizing the elements selected by L_s and stored in the MAR state $x(s)$. That is, for all $i \in \eta(s)$, there is some row of L_s whose ith entry is one (all other entries are zero). Also, let

$$\mathcal{C}_s = \eta(s) \cup \eta(s\overline{\gamma}). \tag{6}$$

A more explicit description of $\eta(s)$ and $\mathcal{C}_s$ is provided in the Appendix, as is a proof of the following proposition which characterizes exactly which elements of P are required for MAR model identification.

Proposition 1: Let $\mathcal{C}_s$ be as defined in (6). Then, there are $O(kN)$ elements of P that are needed to compute the model parameters and these elements are precisely those found in the principal submatrix of P whose rows and columns are indexed by $\bigcup_{s\neq 0}\mathcal{C}_s$. □

We introduce in Section IV a generalized Levinson algorithm, which can be used to determine these required $O(kN)$ covariance elements at a computational cost of $O(k^2N)$. Then, in Section V, we apply this generalized Levinson algorithm to the following problem. Given only knowledge of the diagonal bands of a covariance matrix, we compute a MAR completion for it of the form described in this section and illustrated in Fig. 3, such that the restriction of this MAR process to the leaves of the tree is an ME completion of the partially known covariance matrix. Note that, due to the special construction of this MAR process (as illustrated in Fig. 3 and discussed previously), the resulting MAR process is internal.

III. Background

The covariance extension problem addressed in this paper can be conveniently described in graph-theoretic terms. In this section, we first review some relevant aspects of graph theory and relate them to covariance extension. Following that, we review some known results pertaining to the characterization and computation of extensions.

A. Graph Theory and Existence of Covariance Extensions

Let $\{z(t)|t \in V\}$ be a zero-mean random process indexed by the set V where $|V| = N$. Let z_V denote the column vector of random variables indexed by $t \in V$, and let $P = \mathrm{E}[z_V z_V^T]$ be the covariance matrix for this random vector. Then P is a

[5]Each row of a selection matrix has a single nonzero entry, which equals one.

symmetric and positive-definite matrix (here and throughout we exclude the singular case).

Definition 1 (Partial Covariance Matrix): A *partial covariance matrix* P_E is a symmetric subset of P, i.e.,

$$P_E \triangleq \{p_{i,j} \,|\, (i, j) \in E\} \tag{7}$$

where $E \subset V \times V$. Symmetry means that $(i, j) \in E$ if and only if $(j, i) \in E$. □

We shall refer to the set E as the *support* of the partial covariance P_E. In order to avoid ill-posed extension problems (especially in the context of the ME extension), we shall assume that $(i, i) \in E$ for all i, namely, that P_E contains all diagonal elements.

We can capture the structure and sparsity of P_E by a *support graph* $G = (V, E)$, where we assign a vertex to every $i \in V$ and an edge to every $(i, j) \in E$. Since E is symmetric, all edges are bidirectional and the graph is undirected. This is equivalent to assuming (as we will) that (i, j) is an unordered pair, i.e., $(i, j) = (j, i)$. Also, since $(i, i) \in E$, every vertex has a self-loop. A completely specified matrix gives rise to a complete graph, i.e., $E = V \times V$.

Every principal submatrix of a (nonsingular) covariance matrix is positive definite. Therefore, the same condition is inherited by a partial covariance. The condition that every principal submatrix contained in P_E is positive definite will be denoted by $P_E > 0$.

Given a partial covariance matrix P_E and a graph (V, F) with $E \subset F$, the *covariance extension* problem is to find another partial covariance matrix P_F so that P_F agrees with P_E on the support set E. A partial covariance matrix P_F that satisfies these criteria is called an *extension* of P_E. A *covariance completion* is a covariance extension with $F = V \times V$, i.e., a fully specified valid covariance matrix $P > 0$ that agrees with P_E on the support set E. Note that $P_E > 0$ is a necessary condition for the existence of (nondegenerate) extensions and completions. However, as we discuss (and as is well known), it is not a sufficient condition.

A key concept in examining questions concerning extensions and completions is the notion of a chordal graph, which we introduce next.

Definition 2 (Cycle and Chord): A *cycle* of a graph $G = (V, E)$ is a sequence of distinct vertices $[v_0, v_1, \ldots, v_n]$ where $(v_{i-1}, v_i) \in E$ for all $i \in \{1, 2, 3, \ldots, n\}$ and, additionally, $(v_n, v_0) \in E$. The length of a cycle is the number of its vertices (i.e., $n + 1$). A cycle $[v_0, v_1, \ldots, v_n]$ is said to have a *chord* if $(v_k, v_j) \in E$ for $1 < |k - j| < n$, where $k, j \in \{0, 1, 2, \ldots, n\}$. □

Definition 3 (Chordal Graph): A graph $G = (V, E)$ is called *chordal* if every cycle of length greater than three has a chord. □

We have the following well-known result.

Proposition 2 [25]: Given a graph $G = (V, E)$, completions exist for all valid partial covariance matrices $P_E > 0$ if and only if G is a chordal graph. □

It is worth emphasizing that when G is not chordal, there may exist completions for partial covariance matrices $P_E > 0$ for *specific* choices of the entries $P_E(i, j)$ but not for other valid choices of these entries. However, Proposition 2 tells us that completions exist for *all* valid choices of the entries exactly when G is chordal.

As is well known, when a completion exists, it is not unique [25], [41]. One way to select a particular completion is by requiring that the determinant of the completed covariance matrix is maximal. This yields the ME completion, denoted here by P^{ME}, which maximizes the entropy of the probability density $p(z_1, \ldots, z_M)$, subject to the constraints of zero mean and a partially specified covariance P_E [25], [41].[6] The ME completion of P_E can be characterized by the pattern of zeros in its inverse. In particular, if P^{ME} is the ME completion of P_E then $(P^{\mathrm{ME}})^{-1}(i, j) = 0$ for $(i, j) \notin E$. This fact follows from the well-known solution to the problem of finding the entropy-maximizing probability density function with given moments [7], [41]. In the case of a partially specified covariance, the entropy-maximizing density has the functional form

$$p_{\mathrm{ME}}(z_1, \ldots, z_N) = C \exp\left\{ \sum_{(i,j)\in E} \lambda_{i,j} z_i z_j \right\} \tag{8}$$

where C is a normalization constant, $P_E(i, j) = \mathrm{cov}(z_i, z_j)$ are the given moments, and the Lagrange multipliers $\lambda_{i,j}$ are selected so that $p_{\mathrm{ME}}(\cdot)$ matches the given partial covariance. The density of (8) is Gaussian and clearly indicates that $(P^{\mathrm{ME}})^{-1}(i, j) = 0$ for $(i, j) \notin E$. The pattern of zeros in P^{ME} is equivalently characterized by the statement that a random process on the graph $G = (V, E)$ with covariance matrix P^{ME} is a Markov random field on G [11], [50]. This is probably the most important property of the ME completion.

In view of Proposition 2, we restrict our attention to partial covariances associated with chordal graphs. It turns out that chordality also ensures the existence of ME completions.[7]

Proposition 3 [25]: If $G = (V, E)$ is chordal, then every valid partial covariance P_E has a (unique) ME completion. □

B. One-Element Extensions and Chordal Sequences

Propositions 2 and 3 tell us when completions exist, but do not characterize them or indicate how to compute their elements. Also, nothing has been said yet about the extension of a partial covariance P_E to a larger partial covariance P_F, as opposed to a completion. Since the construction of a MAR model for a (kth-order) Markov process involves an extension in which F is sparse (in particular $|F| = O(k|V|)$), we have a particular interest in extension problems in which both $|E|$ and $|F|$ are small compared to $|V \times V| = |V|^2$. For this reason, we shall focus in the sequel on extensions (and completions) of partial covariances with $|E| \ll |V|^2$.

[6]Note that we have assumed the individual variances, i.e., the diagonal elements of P_E are specified, which is a necessary condition for entropy to be bounded and P^{ME} to exist.

[7]Here we assume that E contains the diagonal, as stated earlier.

The only characterization of P^{ME} we have so far is

$$\begin{aligned} P^{\mathrm{ME}}(i,\,j) &= P_E(i,\,j), \qquad \text{for } (i,\,j) \in E \\ \left(P^{\mathrm{ME}}\right)^{-1}(i,\,j) &= 0, \qquad \text{for } (i,\,j) \notin E. \end{aligned} \tag{9}$$

This implicit characterization was used [41] to construct a generalized AR (Markov) model, which makes it possible to determine the missing covariance elements one by one.[8] Since this procedure involves solving a set of $|E|$ generalized Yule–Walker equations, its overall cost is $O(|E|^3)$ computations. This should be compared with the cost of the "MAX DET" procedure of [21], which determines ME completions *regardless of chordality*. The MAX DET procedure involves an (infinite) sequence of iterations, with a cost of $O(|V|^4)$ computations *per iteration*. This is much higher than the cost of the procedure in [41], especially when $|E| \ll |V|^2$.

A more efficient alternative was described in [6], [33]. It uses the sparsity of the generalized AR model to decompose the set of generalized Yule–Walker equations into a sequence of smaller sets of equations. This results in a reduction of the overall computational cost by (approximately) a factor of $|V|$, namely, from $O(|E|^3)$ to $O(\frac{|E|^3}{|V|})$. The resulting decomposed procedure consists of a sequence of one-element extensions: in each step, one solves a small set of equations in order to determine a single unknown covariance element.

Definition 4 (One-Element Extension): Let $G = (V,\, E)$ and $H = (V,\, F)$ be graphs such that $F = E \cup e$ with $e \notin E$, that is, the graph H is the graph G with one additional edge e. Let P_F be the extension of the partial covariance P_E, that is, P_F is P_E with one additional element, corresponding to the new edge e. Then, P_F is called a *one-element extension* of P_F. □

The key result of [6], [33], which we describe below in some detail, is that every extension (and, therefore, also every completion) of a given partial covariance can be obtained via an appropriately selected sequence of one-element extensions. Moreover, the global ME completion is obtained by selecting the local ME extension in each step of this sequence.

In Section IV, we present a generalized Levinson algorithm which results in a further reduction in cost. Our derivation is motivated by the classical Levinson algorithm [1], [26], [51], which provides a very efficient solution to the problem of extending a banded partial Toeplitz covariance matrix. In this very special case, $P_E(i,\, j)$ is a function of $i - j$, and so

$$E = \{(i,\, j);\ |i - j| \le k\} \tag{10}$$

where the parameter k is known as the bandwidth of the (banded) partial covariance.

The Levinson algorithm is often presented as computing an AR model corresponding to the ME extension in a very efficient manner. However, as is perhaps less commonly known, the Levinson algorithm also provides a characterization of *all* completions of a banded, partial covariance matrix. In particular, there is a one-to-one correspondence between each element of *any* completion (not just the ME one) and a so-called *reflection coefficient*. By choosing values for reflection coefficients and applying the Levinson algorithm the elements of any completion (or extension) can be obtained [51].

It is even less well known that a modified version of the Levinson algorithm still applies to non-Toeplitz covariances [40]. In this more general case, the associated AR model is time-variant, resulting in $O(|V|)$ sets of generalized Yule–Walker equations. In order to extend a banded partial covariance P_E, with E as in (10), one has first to determine the coefficients of the corresponding time-variant AR model (of order k), at a cost of $O(k^2|V|)$ computations, and then find the missing covariance elements one by one, at a cost of k computations per element. Since there are $O(|V|^2)$ missing covariance elements,[9] the overall cost of a completion is dominated by the cost of finding the missing covariance elements, which is $O(k|V|^2)$. This is significantly less than the cost of the procedure in [6], which requires $O(\frac{|E|^3}{|V|}) = O(k^3|V|^2)$ computations, or the procedure in [41], which has a cost of $O(|E|^3) = O(k^3|V|^3)$. We use here the fact that for a banded partial covariance

$$|E| = (2k+1)|V| - k(k+1) = O(k|V|)$$

so that the assumption $|E| \ll |V|^2$ translates into $k \ll |V|$ and, consequently, $|E| \approx (2k+1)|V|$.

There are several ways to consider generalizing the classical Levinson results, each leading to important questions. First, consider the case in which the partial covariance matrix is *not* banded but has entries corresponding to an arbitrary chordal graph. In this case, in what order may the elements of a completion be computed? Second, whether the given partial covariance is banded or not, what if we are interested in an extension rather than a completion: under what conditions can we compute just the entries of interest? Third, for either of these cases, is there a convenient parameterization of all possible extensions? Finally, do efficient algorithms exist for these more general extension problems? Answers to the first two questions are provided by other authors and reviewed in this section. Answers to the last two are provided in Section IV.

In particular, we present in Section IV a generalized Levinson algorithm that can be used to extend arbitrary (chordal) partial covariances. Our goal is a covariance extension procedure with a cost that is comparable to the banded case, i.e., $O(\beta|V|^2)$ where, in general, we define the bandwith β as a measure of sparsity, *viz.*,

$$\beta = \frac{1}{2}\left(\frac{|E|}{|V|} - 1\right). \tag{11}$$

Again, our earlier assumption $|E| \ll |V|^2$ implies that $\beta \ll |V|$. With this assumption, our definition of β coincides, for banded partial covariances, with the notion of bandwidth k from (10).

In the remainder of this section we introduce graph-theoretic concepts that are needed both for the discussion of the results of [6], [33] and for the derivation of our generalized Levinson algorithm.

[8] As stated earlier, we assume here and in the sequel that the graph $G = (V,\, E)$ associated with the partial covariance P_E is chordal.

[9] As stated earlier, we assume here and in the sequel that $|E| \ll |V|^2$.

Definition 5 [(Complete) Chordal Sequence]: Let $G^i = (V, E^i)$ be a chordal graph for $i \in \{0, 1, \ldots, n\}$. Then $[G^0, G^1, \ldots G^n]$ is a *chordal sequence* if $E^i = E^{i-1} \cup \{e^i\}$ with $e^i \notin E^{i-1}$, i.e., if E^i includes only one additional edge beyond those in E^{i-1}. If, in addition, G^n is the complete graph then $[G^0, G^1, \ldots G^n]$ is a *complete chordal sequence*. □

Definition 6 (Induced Subgraph): Let $G = (V, E)$ be a graph. If $U \subseteq V$ then by $G_U = (U, E_U)$ we denote the subgraph of G induced by U where

$$E_U = \{(u, v) \in E \mid u, v \in U\} = E \cap (U \times U). \qquad (12)$$

□

Definition 7 [(Maximal) Clique]: A *clique* is a set of vertices U that induces a complete subgraph, G_U. A clique is *maximal* if it is not a proper subset of another clique. □

As mentioned earlier, the key result [6, Theorem 7.2.7], [33, discussion in proof of Theorem 2], [33, Theorem 6] states that elements of completions (and extensions) can be computed one by one if the order of one-element extensions corresponds to a chordal sequence, and that local ME choices results in the unique global ME completion.

Proposition 4 [6], [33]: Let $[G^0, G^1, \ldots, G^n]$ be a complete chordal sequence where $G^i = (V, E^i)$ and $E^n = V \times V$. Then the elements of any completion of P_{E^0} can be obtained via a sequence of one-element extensions $[P_{E^1}, P_{E^2}, \ldots, P_{E^n}]$. The edge added in the step from G^{i-1} to G^i completes exactly one maximal clique in G^i, corresponding to a newly completed submatrix of P_{E^i} which we denote by Φ_i. The global ME completion P^{ME} is obtained by maximizing the determinant of Φ_i in every step of the sequence of one-element extensions. □

Proposition 4 means that there is a maximal clique in G^i that is not a clique in G^{i-1}. The maximal cliques of graphs in a chordal sequence have a special property described by the following lemma.

Lemma 1 [25]: Let $G^i = (V, E^i)$ be a chordal sequence. Denote by e^i the edge that is added in the step $G^{i-1} \to G^i$, namely, $E^i = E^{i-1} \cup e^i$. If $e^i = (a, b)$, then the unique maximal clique of G^i containing a and b is of the form $Q = \{a, b\} \cup U$ where U is a clique of G^{i-1}. □

We will call the endpoints a and b of e^i the *active vertices* of the maximal clique Q. An illustration of the relationships among a, b, U, and Q is provided in Fig. 4, along with some other notation to be defined in the sequel.

Terminating a complete chordal sequence prior to arriving at the complete graph results in an extension rather than a completion. It turns out that one can always associate a chordal sequence with every extension (or completion) problem that is specified in terms of chordal graphs.

Proposition 5 [6]: If (V, E) and (V, F) are chordal graphs with $E \subset F$ then there exists a chordal sequence $[G^0, G^1, \ldots, G^n]$ such that $G^0 = (V, E)$ and $G^n = (V, F)$. □

Therefore, the elements of an extension P_{E^n} of a partial covariance matrix P_{E^0} may be determined without computing any elements not in the extension if 1) $G^0 = (V, E^0)$ is chordal, 2) $G^n = (V, E^n)$ is chordal, and 3) $E^0 \subset E^n$. Proposition 5 does not address the problem of finding a chordal sequence between G^0 and G^n. However, for the special case in which G^n is the complete graph, there exists an algorithm with complexity $O(|V| + |E|)$ for finding one [25], [47]. This is significantly less than the complexity of finding the missing covariance elements, which cannot be less than $O(|V|\,|E|) = O(k|V|^2)$, as discussed in Section III-B. More generally, one may always construct a chordal sequence recursively: given G^{i-1}, we form G^i by searching over $E^n - E^{i-1}$ for an edge that preserves chordality. For each candidate edge in $E^n - E^{i-1}$, the complexity of checking chordality is no larger than $O(|V| + |E^i|)$ [52].

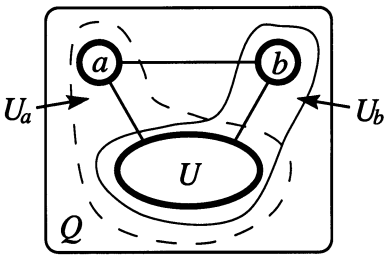

Fig. 4. Maximal clique Q is the union of clique U and $\{a, b\}$. U_a is the union of U with $\{a\}$ and U_b is the union of U with $\{b\}$. For visual clarity, we have drawn one line between a and U and also between b and U although a and b are connected by edges to all of the elements of U.

Proposition 4 states that the elements of every extension (or completion) may be computed sequentially by following a complete chordal sequence but it does not provide an algorithm to do so. In the next section, we *do* provide a Levinson-type algorithm, based on the notion of order-recursive linear prediction and generalized reflection (or partial correlation (PARCOR)) coefficients. In particular, the ME extension is obtained by setting these coefficients to zero along the chordal sequence.

IV. Generalized Reflection Coefficients and a Generalized Levinson Algorithm

In this section, we consider the extension of a given partial covariance matrix P_E with chordal support to a larger chordal support $F \supset E$, such that P_F agrees with P_E on the support set E. According to Propositions 4 and 5, this can be accomplished by a sequence of one-element extensions, following a chordal sequence from E to F. Our approach to obtaining a computationally efficient extension procedure is based on an order-recursive linear prediction interpretation of the one-element extension step. This leads us to introduce a generalized Levinson algorithm, which we use to compute the linear prediction parameters needed to determine a single unspecified covariance element. We show that the flexibility in selecting the values for the unspecified elements is completely characterized in terms of a sequence of generalized reflection coefficients, one for each single-element extension step. These coefficients, as their name suggests, are generalizations of the reflection coefficients associated with the classical Levinson algorithm, with values bounded by unity. In particular, the maximum-entropy choice corresponds to selecting all unspecified reflection coefficients equal to zero. Under certain graph-theoretic conditions, which we discuss below, our generalized Levinson recursion becomes particularly efficient.

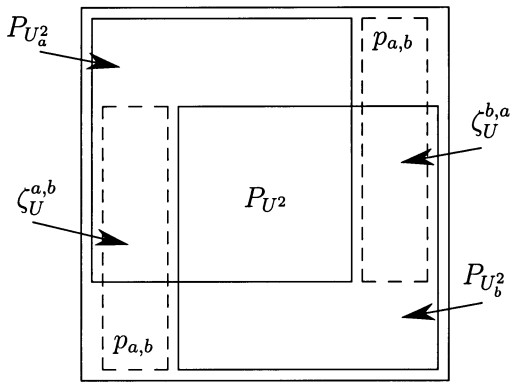

Fig. 5. The maximal principal submatrix P_{Q^2} contains principal submatrices $P_{U_a^2}$ (upper left), $P_{U_b^2}$ (lower right), and P_{U^2} (center). The vectors $\zeta_U^{a,b}$ and $\zeta_U^{b,a}$ and the element $p_{a,b}$ are also indicated.

A. One-Element Extensions, Linear Prediction, and Normal Equations

Since covariance extensions can be obtained via a sequence of one-element extensions, we need only consider a single step in such a sequence. In each step, we determine a single new covariance element, using other previously determined covariance elements. In the ith step, the partial covariance $P_{E^{i-1}}$ is extended to P_{E^i}, with the new covariance element being represented by the new edge e^i, so that $E^i = E^{i-1} \cup e^i$. Proposition 4 and Lemma 1 tell us that $G^i = (V, E^i)$ contains exactly one maximal clique that was not in $G^{i-1} = (V, E^{i-1})$. Because all other cliques were already included in G^{i-1}, and because maximal principal submatrices of P_{E^i} correspond to maximal cliques of G^i, it follows that we need only be concerned with the positivity of the new submatrix. In other words, assuming that we had $P_{E^{i-1}} > 0$, then $P_{E^i} > 0$ if and only if the maximal principal submatrix P_{Q^2} is positive definite, where Q is as defined in Lemma 1 and $Q^2 = Q \times Q$.

To be more specific, recall that the new covariance element corresponds to an edge $e^i = (a, b)$, and that the "new submatrix" corresponds to a subset of vertices $U \cup \{a, b\}$. The (known) covariance elements that determine the range of values of the new element $p_{a,b}$ are described in Fig. 5. The new submatrix is the covariance matrix of the random vector $[z_a^T \quad z_U^T \quad z_b^T]^T$ where we use the notation z_D to denote the column vector of random variables (i.e., random vector) indexed by an arbitrary subset D of Q. If D is a singleton set, e.g., $D = \{a\}$, we will typically write this as z_a rather than $z_{\{a\}}$. The bottom left (as well as the top right) element of this matrix is the yet to be determined $p_{a,b}$. The rest of the elements in this matrix are known from $P_{E^{i-1}}$.

In order to facilitate our discussion we need to introduce several notational shortcuts as follows:

$$U_a \triangleq \{a\} \cup U, \qquad U_b \triangleq \{b\} \cup U \tag{13a}$$

$$Q = \{a, b\} \cup U = U_a \cup U_b \tag{13b}$$

$$P_{U^2} = \{p_{i,j} \,|\, (i,j) \in U \times U = U^2\} \tag{13c}$$

$$P_{U_a^2} = \{p_{i,j} \,|\, (i,j) \in U_a \times U_a = U_a^2\} \tag{13d}$$

$$P_{U_b^2} = \{p_{i,j} \,|\, (i,j) \in U_b \times U_b = U_b^2\}. \tag{13e}$$

These principal submatrices, as well as some of the other notation to be used in this section, are illustrated in Fig. 5. The matrix $P_{U_a^2}$ contains all but the last row and column of P_{Q^2} and is indicated in the upper left of Fig. 5; $P_{U_b^2}$ contains all but the first row and column of P_{Q^2} and is indicated in the lower right of Fig. 5; and P_{U^2} contains all but the first and last row and column of P_{Q^2} (it is the intersection of $P_{U_a^2}$ and $P_{U_b^2}$), and is indicated in the center of Fig. 5. The only unknown covariance element $p_{a,b}$ occupies the upper right and lower left corners of P_{Q^2}.

The range of valid values of $p_{a,b}$ is constrained only by the fact that P_{Q^2} must be positive definite, a property that can be captured in several different ways, e.g., by checking the leading principal submatrices. Instead, we choose a characterization that relies on the notions of linear prediction and PARCOR coefficients. The relation between one-element extensions and linear prediction motivates us to develop an order-recursive algorithm, akin to the classical Levinson algorithm, that allows efficient calculation of the unknown covariance elements.

The concepts of linear prediction and generalized reflection coefficients are both defined in terms of certain random variables associated with the submatrix

$$P_{Q^2} = \mathrm{E}\left\{ \begin{bmatrix} z_a \\ z_U \\ z_b \end{bmatrix} \begin{bmatrix} z_a \\ z_U \\ z_b \end{bmatrix}^T \right\}.$$

Referring to Figs. 4 and 5, of particular interest to us are estimates and associated estimation errors corresponding to estimating z_a and z_b based on z_U. Let

$$\tilde{z}_a \triangleq z_a - \hat{\mathrm{E}}[z_a \,|\, z_U] \triangleq z_a - \hat{z}_a \tag{14a}$$

$$\tilde{z}_b \triangleq z_b - \hat{\mathrm{E}}[z_b \,|\, z_U] \triangleq z_b - \hat{z}_b \tag{14b}$$

where $\hat{\mathrm{E}}[x \,|\, y]$ is the LLS estimate of x based on y. Define the *generalized reflection coefficient* $\rho_U^{a,b}$ as

$$\rho_U^{a,b} \triangleq \frac{\mathrm{cov}(\tilde{z}_a, \tilde{z}_b)}{\sqrt{\mathrm{var}(\tilde{z}_a)\,\mathrm{var}(\tilde{z}_b)}} \tag{15}$$

namely, the correlation coefficient of $\tilde{z}_a$ and $\tilde{z}_b$. It is also known as the PARCOR coefficient of z_a and z_b given z_U. The coefficient $\rho_U^{a,b}$ provides a simple and efficient characterization both of the ME choice for $p_{a,b}$ and the range of all valid choices for $p_{a,b}$.

Proposition 6 (PARCOR Characterization of One-Element Extensions): Using the notation defined previously, there is a one-to-one correspondence between the PARCOR coefficient $\rho_U^{a,b}$ and the covariance element $p_{a,b}$. Moreover, $P_{Q^2} > 0$ if and only if $|\rho_U^{a,b}| < 1$. The choice of $\rho_U^{a,b} = 0$ maximizes the determinant of P_{Q^2} and, thereby, provides the global ME value of $p_{a,b}$. □

The proof of Proposition 6, provided later in this section, relies on several results derived from the orthogonality principle of LLS estimation.

To relate the generalized reflection coefficient $\rho_U^{a,b}$ to the unknown covariance element $p_{a,b}$ and, thereby, perform a one-element covariance extension, requires solving sets of normal equations associated with the residuals of (14). These equations

follow from the orthogonality principle, namely, $z_a - \hat{z}_a$ and $z_b - \hat{z}_b$ are both orthogonal to z_U. Since $\hat{z}_b$ is a linear combination of the elements of the random vector z_U, we can write it in the form

$$\hat{z}_b = L_U^b z_U \tag{16a}$$

where L_U^b is a row vector of deterministic coefficients. Consequently

$$\tilde{z}_b = [-L_U^b \quad 1] \begin{bmatrix} z_U \\ z_b \end{bmatrix}$$

and cross correlating $\tilde{z}_b$ with $\left[\begin{smallmatrix} z_U \\ z_b \end{smallmatrix}\right]$ we obtain the normal equations

$$[0 \quad \cdots \quad 0 \quad \varepsilon_U^b] = \underbrace{[-L_U^b \quad 1]}_{B_U^b} P_{U_b^2} \tag{16b}$$

where (cf., (14)) $\varepsilon_U^b \triangleq \mathrm{var}(\tilde{z}_b)$ is the LLS estimation error variance and B_U^b is as defined in (16b).

Our next step is to extend the normal equations (16b) in such a way that $p_{a,b}$ is introduced. By augmenting B_U^b with a zero and expanding $P_{U_b^2}$ by one row and column we have

$$[0 \quad B_U^b] P_{Q^2} = [\delta_U^{a,b} \quad 0 \quad \cdots \quad 0 \quad \varepsilon_U^b] \tag{17a}$$

where

$$\delta_U^{a,b} \triangleq B_U^b \zeta_U^{a,b} \tag{17b}$$

and where $\zeta_U^{a,b}$ denotes a particular subcolumn of elements of P_{Q^2} that includes $p_{a,b}$ (see Fig. 5). Specifically

$$\zeta_U^{a,b} \triangleq \mathrm{E}\left\{ \begin{bmatrix} z_U \\ z_b \end{bmatrix} z_a \right\} = \begin{bmatrix} \mathrm{E}[z_U z_a] \\ p_{a,b} \end{bmatrix}. \tag{17c}$$

The augmented normal equations (17a) serve three purposes: i) they introduce $\delta_U^{a,b}$, which is central to the solution of the one-element extension problem; ii) they are used in proving Proposition 6; and iii) they are key to the construction of a generalized Levinson algorithm.

The linear estimation of a random signal sample (say z_b) from other samples of the same random signal (say z_U) is usually known as linear prediction. For this reason, the residual $\tilde{z}_b$ is known as a linear prediction residual, the coefficient vector B_U^b is known as a linear prediction vector, and ε_U^b is known as a linear prediction error variance.

Analogous to (16) and (17), we also need to consider the LLS estimate of the scalar random variable z_a based on the random vector z_U which we can write as

$$\hat{z}_a = L_U^a z_U. \tag{18a}$$

The row vector of linear prediction coefficients L_U^a must satisfy the following normal equations:

$$\underbrace{[1 \quad -L_U^a]}_{A_U^a} P_{U_a^2} = [\varepsilon_U^a \quad 0 \quad \cdots \quad 0] \tag{18b}$$

where (cf., (14)) $\varepsilon_U^a \triangleq \mathrm{var}(\tilde{z}_a)$ is the estimation error variance, and A_U^a is as defined in (18b). Notice that $A_U^a = [1 \quad -L_U^a]$ has the unity element on the left, while $B_U^b = [-L_U^b \quad 1]$ has the unity element on the right. By augmenting A_U^a with a zero element and expanding $P_{U_a^2}$ by one row and column, we have

$$[A_U^a \quad 0] P_{Q^2} = [\varepsilon_U^a \quad 0 \quad \cdots \quad 0 \quad \delta_U^{b,a}] \tag{19a}$$

where

$$\delta_U^{b,a} \triangleq A_U^a \zeta_U^{b,a} \tag{19b}$$

and where $\zeta_U^{b,a}$ denotes another subcolumn of P_{Q^2}, again including $p_{a,b}$ (see Fig. 5). Specifically

$$\zeta_U^{b,a} \triangleq \mathrm{E}\left\{ \begin{bmatrix} z_a \\ z_U \end{bmatrix} z_b \right\} = \begin{bmatrix} p_{a,b} \\ \mathrm{E}[z_U z_b] \end{bmatrix}. \tag{19c}$$

Our proof of Proposition 6 relies on three lemmas, which relate the generalized reflection coefficient $\rho_U^{a,b}$ to the inner product $\delta_U^{a,b}$ (Lemma 4) and through it to $p_{a,b}$ (Lemma 3).

Lemma 2: $\delta_U^{a,b}$ and $\delta_U^{b,a}$ of (17b) and (19b), respectively, are equal.

Proof: Using (19a) and the fact that

$$[0 \quad B_U^b] = [0 \quad -L_U^b \quad 1]$$

we have

$$[A_U^a \quad 0] P_{Q^2} [0 \quad B_U^b]^T = \delta_U^{b,a}. \tag{20a}$$

On the other hand, using (17a) and the fact that

$$[A_U^a \quad 0] = [1 \quad -L_U^a \quad 0]$$

we have

$$[0 \quad B_U^b] P_{Q^2} [A_U^a \quad 0]^T = \delta_U^{a,b}. \tag{20b}$$

From (20a) and (20b) it follows that $\delta_U^{a,b} = \delta_U^{b,a}$. □

Lemma 3: Using the notation defined previously

$$\delta_U^{a,b} = -L_U^b \mathrm{E}[z_U z_a] + p_{a,b} \tag{21a}$$

$$= -L_U^a \mathrm{E}[z_U z_b] + p_{a,b}. \tag{21b}$$

Proof: Equation (21a) follows by substituting the definition of $\zeta_U^{a,b}$ given in (17c) and the definition of B_U^b given in (16b) into (17b). Equation (21b) follows from Lemma 2 and by substituting the definition of $\zeta_U^{b,a}$ given in (19c) and the definition of A_U^a given in (18b) into (19b). □

Lemma 4: Using the notation defined previously

$$\rho_U^{a,b} = \frac{\delta_U^{a,b}}{\sqrt{\varepsilon_U^a \varepsilon_U^b}}. \tag{22}$$

Proof: Since $\varepsilon_U^a = \mathrm{var}(\tilde{z}_a)$, $\varepsilon_U^b = \mathrm{var}(\tilde{z}_b)$, and using (15), it suffices to show that

$$\delta_U^{a,b} = \mathrm{cov}(\tilde{z}_a, \tilde{z}_b) \triangleq \mathrm{E}[(z_a - \hat{z}_a)(z_b - \hat{z}_b)].$$

This is easily done as follows:

$$\mathrm{E}[(z_a - \hat{z}_a)(z_b - \hat{z}_b)] = \mathrm{E}[(z_a - \hat{z}_a) z_b] \tag{23a}$$

$$= p_{a,b} - L_U^a \mathrm{E}[z_U z_b] \tag{23b}$$

$$= \delta_U^{a,b} \tag{23c}$$

where in (23a) we have used the fact that $z_a - \hat{z}_a$ is orthogonal to z_U and, hence, to $\hat{z}_b$; in (23b) we have taken expectations; in (23c) we have used Lemma 3. □

We now provide the proof of Proposition 6.

Proof (Proposition 6): Using Lemmas 3 and 4 we have

$$p_{a,b} = L_U^b \,\mathrm{E}[z_U z_a] + \rho_U^{a,b}\sqrt{\varepsilon_U^a \varepsilon_U^b} \tag{24a}$$

$$= L_U^a \,\mathrm{E}[z_U z_b] + \rho_U^{a,b}\sqrt{\varepsilon_U^a \varepsilon_U^b}. \tag{24b}$$

These relations establish a one-to-one correspondence between $p_{a,b}$ and $\rho_U^{a,b}$. Thus, we only need to show that: i) $|\rho_U^{a,b}| < 1$ if and only if $P_{Q^2} > 0$, and ii) $\rho_U^{a,b} = 0$ is the ME value. Using (17a) we have

$$\left(\begin{array}{c|c} & 0 \\ I & \vdots \\ & 0 \\ \hline 0 & B_U^b \end{array}\right) P_{Q^2} \left(\begin{array}{c|c} & 0 \\ I & \vdots \\ & 0 \\ \hline 0 & B_U^b \end{array}\right)^T = \underbrace{\left(\begin{array}{ccc|c} & & & \delta_U^{a,b} \\ & P_{U_a^2} & & 0 \\ & & & \vdots \\ & & & 0 \\ \hline \delta_U^{a,b} & 0 \;\cdots\; & 0 & \varepsilon_U^b \end{array}\right)}_{K}. \tag{25}$$

Observe that the first (leftmost) matrix in (25) is lower triangular with unity diagonal elements, because the last element of the row vector B_U^b is one. Thus, by Sylvester's law of inertia [22], we conclude that $P_{Q^2} > 0$ if and only if the matrix on the right-hand side of (25), which we have denoted by K, is positive definite. Since a matrix $\left(\begin{smallmatrix} F_1 & F_2 \\ F_2^T & F_3 \end{smallmatrix}\right) > 0$ if and only if $F_1 > 0$ and $F_3 - F_2^T F_1^{-1} F_2 > 0$, and since[10] $P_{U_a^2} > 0$, we conclude from the form of the matrix K (in the right-hand side of (25)) that $P_{Q^2} > 0$ if and only if

$$\varepsilon_U^b - \left(\delta_U^{a,b}\right)^2 [1 \quad 0 \quad \cdots \quad 0] P_{U_a^2}^{-1} [1 \quad 0 \quad \cdots \quad 0]^T > 0. \tag{26a}$$

This positivity condition can be simplified by using (18b). Doing so, we have

$$1 = A_U^a [1 \quad 0 \quad \cdots \quad 0]^T = \varepsilon_U^a [1 \quad 0 \quad \cdots \quad 0] P_{U_a^2}^{-1} [1 \quad 0 \quad \cdots \quad 0]^T$$

so that (26a) simplifies to

$$\varepsilon_U^b - \left(\delta_U^{a,b}\right)^2 [1 \quad 0 \quad \cdots \quad 0] P_{U_a^2}^{-1} [1 \quad 0 \quad \cdots \quad 0]^T = \varepsilon_U^b - \frac{\left(\delta_U^{a,b}\right)^2}{\varepsilon_U^a} > 0 \tag{26b}$$

or, using the fact that ε_U^a is positive

$$\left(\delta_U^{a,b}\right)^2 < \varepsilon_U^a \varepsilon_U^b.$$

[10]The matrix $F_3 - F_2^T F_1^{-1} F_2$ is known as a Schur complement.

Thus, by Lemma 4, the positivity condition (26a) is equivalent to

$$\left|\rho_U^{a,b}\right| = \left|\frac{\delta_U^{a,b}}{\sqrt{\varepsilon_U^a \varepsilon_U^b}}\right| < 1$$

and we have now established that $P_{Q^2} > 0$ if and only if $|\rho_U^{a,b}| < 1$.

That the choice of $\rho_U^{a,b} = 0$ maximizes the determinant of P_{Q^2} can be seen as follows.[11] Using (25), (26a), and the fact that $B_U^b = [-L_U^b \;\; 1]$ we have that

$$\det(P_{Q^2}) = \det\left(P_{U_a^2}\right)\left(\varepsilon_U^b - \left(\delta_U^{a,b}\right)^2 [1 \quad 0 \quad \cdots \quad 0] \cdot P_{U_a^2}^{-1} [1 \quad 0 \quad \cdots \quad 0]^T\right).$$

Applying (26b) and using Lemma 4, we have

$$\det(P_{Q^2}) = \det\left(P_{U_a^2}\right)\varepsilon_U^b \left[1 - \left(\rho_U^{a,b}\right)^2\right]. \tag{27}$$

Hence, the choice of $\rho_U^{a,b} = 0$ maximizes the determinant of P_{Q^2}. □

In view of Proposition 4, the corresponding value of $p_{a,b}$ coincides with its value in the global ME completion P^{ME}. Proposition 6 establishes, among other things, that the choice of $\rho_U^{a,b} = 0$ provides the determinant-maximizing value of $p_{a,b}$. Since, as stated previously, determinant maximization and entropy maximization coincide

$$p_{a,b}^{\mathrm{ME}} \triangleq L_U^b \,\mathrm{E}[z_U z_a] = L_U^a \,\mathrm{E}[z_U z_b] \tag{28a}$$

is the entropy-maximizing value of $p_{a,b}$ where we have used (24). Our final expression for $p_{a,b}$ is

$$p_{a,b} = p_{a,b}^{\mathrm{ME}} + \rho_U^{a,b}\sqrt{\varepsilon_U^a \varepsilon_U^b}. \tag{28b}$$

The quantities A_U^a, B_U^b, ε_U^a, and ε_U^b are computed using only the already-known elements of P_{Q^2}. The new element $p_{a,b}$ is then calculated via (28b) using these known elements together with the selected value of the remaining degree of freedom, namely, $\rho_U^{a,b}$.

The complete solution to a covariance extension problem consists of a sequence of one-element extensions, ordered according to a chordal sequence. Each step along the chordal sequence involves a clique associated with a set of vertices $U^i \subset V$ and requires the solution of two sets of normal equations. The size of each of these systems of equations is given by the cardinality of the set U^i. Hence, the complexity of direct solution of the normal equations is $O(|U^i|^3)$ (see also [33, pp. 174–175]). If the extension problem is such that $|U^i|$ is small and does not grow with i (or is bounded in size with the bound being independent of $|V|$), then direct solution of the normal equations is computationally feasible.

Indeed, if for each step in the chordal sequence $|U^i| \leq \mathcal{K} \ll |V|$, then the average cost per each missing covariance element

[11]Alternatively, the Fischer inequality [28] can also be used to establish this fact.

is $O(\mathcal{K}^3)$. This is already a major improvement over the (nonrecursive) method described in [41], which involves solving a set of linear equations in $|E^0|$ unknowns and using the results to extend the partial covariance. Since $|E^0| = O(\mathcal{K}|V|)$, the technique in [41] requires $\frac{O(\mathcal{K}^3|V|^3)}{O(|V|^2)} = O(\mathcal{K}^3|V|)$ computations per missing covariance element, which is much more than the cost of a direct solution implementation of our technique.

On the other hand, if $|U^i|$ does grow with i, then the complexity of direct solution of the normal equations may become prohibitive. Under certain graph-theoretic conditions, however, there is a way to solve the normal equations *efficiently*, namely, with a complexity that is *linear* in $|U^i|$. This efficient procedure is discussed in the next subsection.

B. Efficient Generalized Levinson Recursion

We have shown that all valid choices for the missing covariance element $p_{a,b}$ are characterized via (28) in terms of the linear prediction coefficient vectors L_U^a, L_U^b and the associated error variances ε_U^a, ε_U^b. The direct evaluation of these quantities (via the normal equations (16b) and (18b)) can be avoided by using the generalized Levinson recursion, which uses the linear prediction coefficients L_U^a, ε_U^a, L_U^b, ε_U^b to determine the higher order coefficients $L_{U_b}^a$, $\varepsilon_{U_b}^a$, $L_{U_a}^b$, $\varepsilon_{U_a}^b$. Now the higher order linear prediction coefficients $(L_{U_a}^b, \varepsilon_{U_a}^b)$ must satisfy the higher order normal equations associated with the LLS estimate of z_b given $z_{U_a} = [z_a \quad z_U^T]^T$: this estimator has the form $L_{U_a}^b z_{U_a}$ with the row vector $L_{U_a}^b$ satisfying the normal equations (recall (16b))

$$\underbrace{[-L_{U_a}^b \quad 1]}_{B_{U_a}^b} P_{Q^2} = [0 \quad \cdots \quad 0 \quad \varepsilon_{U_a}^b] \tag{29a}$$

where $\varepsilon_{U_a}^b$ is the estimation error variance, and $B_{U_a}^b$ is as defined in (29a). Analogously, the higher order linear prediction coefficients $(L_{U_b}^a, \varepsilon_{U_b}^a)$ must satisfy the normal equations associated with LLS estimate of z_a given $z_{U_b} = [z_U^T \quad z_b]^T$: this estimator has the form $L_{U_b}^a z_{U_b}$ with the row vector $L_{U_b}^a$ satisfying the normal equations (recall (18b))

$$\underbrace{[1 \quad -L_{U_b}^a]}_{A_{U_b}^a} P_{Q^2} = [\varepsilon_{U_b}^a \quad 0 \quad \cdots \quad 0] \tag{29b}$$

where $\varepsilon_{U_b}^a$ is the estimation error variance, and $A_{U_b}^a$ is as defined in (29b).

Proposition 7: Given (L_U^a, ε_U^a) and (L_U^b, ε_U^b) we can determine $(L_{U_b}^a, \varepsilon_{U_b}^a)$ and $(L_{U_a}^b, \varepsilon_{U_a}^b)$ via the expressions

$$\underbrace{[-L_{U_a}^b \quad 1]}_{B_{U_a}^b} = [0 \quad B_U^b] - \rho_U^{a,b}\sqrt{\frac{\varepsilon_U^b}{\varepsilon_U^a}}\,[A_U^a \quad 0],$$

$$\varepsilon_{U_a}^b = \varepsilon_U^b\left(1 - \left(\rho_U^{a,b}\right)^2\right) \tag{30a}$$

$$\underbrace{[1 \quad -L_{U_b}^a]}_{A_{U_b}^a} = [A_U^a \quad 0] - \rho_U^{a,b}\sqrt{\frac{\varepsilon_U^a}{\varepsilon_U^b}}\,[0 \quad B_U^b],$$

$$\varepsilon_{U_b}^a = \varepsilon_U^a\left(1 - \left(\rho_U^{a,b}\right)^2\right). \tag{30b}$$

Proof: Using the augmented normal equations (17a) and (19a) we have

$$\left\{[0 \quad B_U^b] - \frac{\delta_U^{a,b}}{\varepsilon_U^a}[A_U^a \quad 0]\right\} P_{Q^2}$$
$$= \left[0 \quad \cdots \quad 0 \quad \varepsilon_U^b - \frac{\left(\delta_U^{a,b}\right)^2}{\varepsilon_U^a}\right]$$
$$= \left[0 \quad \cdots \quad 0 \quad \varepsilon_U^b\left(1 - \left(\rho_U^{a,b}\right)^2\right)\right].$$

By comparing this expression to (29a), we conclude that

$$\underbrace{[-L_{U_a}^b \quad 1]}_{B_{U_a}^b} = [0 \quad B_U^b] - \frac{\delta_U^{a,b}}{\varepsilon_U^a}[A_U^a \quad 0]$$
$$= [0 \quad B_U^b] - \rho_U^{a,b}\sqrt{\frac{\varepsilon_U^b}{\varepsilon_U^a}}\,[A_U^a \quad 0]$$

and

$$\varepsilon_{U_a}^b = \varepsilon_U^b\left(1 - \left(\rho_U^{a,b}\right)^2\right).$$

Using the augmented normal equations (17a) and (19a) we have

$$\left\{[A_U^a \quad 0] - \frac{\delta_U^{a,b}}{\varepsilon_U^b}[0 \quad B_U^b]\right\} P_{Q^2}$$
$$= \left[\varepsilon_U^a - \frac{\left(\delta_U^{a,b}\right)^2}{\varepsilon_U^b} \quad 0 \quad \cdots \quad 0\right]$$
$$= \left[\varepsilon_U^a\left(1 - \left(\rho_U^{a,b}\right)^2\right) \quad 0 \quad \cdots \quad 0\right].$$

By comparing this expression with (29b), we conclude that

$$\underbrace{[1 \quad -L_{U_b}^a]}_{A_{U_b}^a} = [A_U^a \quad 0] - \frac{\delta_U^{a,b}}{\varepsilon_U^b}[0 \quad B_U^b]$$
$$= [A_U^a \quad 0] - \rho_U^{a,b}\sqrt{\frac{\varepsilon_U^a}{\varepsilon_U^b}}\,[0 \quad B_U^b]$$

and

$$\varepsilon_{U_b}^a = \varepsilon_U^a\left(1 - \left(\rho_U^{a,b}\right)^2\right). \qquad \square$$

Comment 1: An examination of (30a) shows that the complexity of solving for $B_{U_a}^b$ (or, equivalently, for $L_{U_a}^b$) is linear in the size of the vectors A_U^a and B_U^b and, thus, is linear in $|U|$. The same conclusion holds for the calculation of $(L_{U_b}^a, \varepsilon_{U_b}^a)$ via (30b). Thus, the computational cost per each missing covariance element is linear in $|U^i| \leq \mathcal{K}$, as compared to $O(\mathcal{K}^3)$, the cost of the direct method described in the first part of this section. Moreover, in the ME case, when $\rho_U^{a,b} = 0$, these calculations are essentially trivial. Indeed, in this case we have

$$B_{U_a}^b = [0 \quad B_U^b], \qquad \varepsilon_{U_a}^b = \varepsilon_U^b \tag{31a}$$
$$A_{U_b}^a = [A_U^a \quad 0], \qquad \varepsilon_{U_b}^a = \varepsilon_U^a \tag{31b}$$

so that the only computation left is the determination of $p_{a,b}^{\mathrm{ME}}$ via (28), again at a cost of $O(|U^i|)$.

Combining the global graph-theoretic perspective of Proposition 4 with the local algebraic detail provided by the generalized Levinson recursion of Proposition 7, we conclude that a complete solution of a covariance extension problem consists of a sequence of one-element extensions, each one using the input linear prediction coefficients (L_U^a, ε_U^a) and (L_U^b, ε_U^b) to determine a single covariance element $p_{a,b}$ via (28). Each step of this recursion also determines the output linear prediction coefficients $(L_{U_b}^a, \varepsilon_{U_b}^a)$ and $(L_{U_a}^b, \varepsilon_{U_a}^b)$, which are then used as the starting point for subsequent steps. The ordering of recursive steps is specified by a choice of a chordal sequence.

Some of the input (lower order) linear prediction coefficients that are required in the generalized Levinson recursion may be directly computable from the known partial covariance data (via appropriate normal equations). From now on, we assume that this initial set of linear prediction coefficients is available prior to carrying out the generalized Levinson recursion.

We consider the generalized Levinson recursion *efficient* if the input linear prediction information required at each step is either available as output information from previous steps of the recursion or is included in the initial set of linear prediction coefficients. Efficiency is clearly a property of the chordal sequence associated with the recursion. That not every chordal sequence is efficient is easily shown by a counterexample [18]. The following graph-theoretic property provides a formal characterization of efficient chordal sequences. (The term *active vertex* used in the following definition is defined in the discussion immediately following Lemma 1.)

Definition 8 (Efficient Chordal Sequence): Let $[G^0, \ldots, G^n]$ be a sequence of chordal graphs with

$$G^i = (V, E^i) \quad \text{and} \quad E^i = E^{i-1} \cup \{e^i\}$$

where $e^i = (a^i, b^i) \notin E^{i-1}$. Let U^i be a clique of G^{i-1} such that the unique maximal clique of G^i containing a^i, b^i is $\{a^i, b^i\} \cup U^i$. Then this sequence of graphs is said to be an *efficient chordal sequence* if, for some $j < i$, and $k < i$,

i) G^j has a maximal clique $\{a^i\} \cup U^i$ with a^i an active vertex, and
ii) G^k has a maximal clique $\{b^i\} \cup U^i$ with b^i an active vertex. □

If our chordal sequence is not efficient this does not mean that we cannot perform an extension. It simply means that for some steps we must solve the normal equations directly and with cubic rather than linear complexity. Given an arbitrary choice of two chordal graphs $G^0 = (V, E^0)$ and $G^n = (V, E^n)$ such that $E^0 \subset E^n$, it is unclear (as of this writing) whether there exists an *efficient* chordal sequence between them.

Proposition 8: Let $G^0 = (V, E^0)$ be the chordal graph associated with partial covariance matrix P_{E^0}. Let $|E^0| = |V| + m$ (i.e., there are m edges in addition to all self-loops). Let

$$\Gamma \triangleq \left[G^{-m}, G^{-m+1}, \ldots, G^0, \ldots, G^{n-1}, G^n\right] \tag{32}$$

be a chordal sequence where G^{-m} has no edges other than self-loops and where $G^i = (V, E^i)$ with $E^i = E^{i-1} \cup \{(a^i, b^i)\}$. Associated with Γ is a sequence of new maximal cliques $[Q^{-m+1}, \ldots, Q^n]$ where $Q^i = \{a^i, b^i\} \cup U^i$ and U^i is a clique of graph G^{i-1}. Consider the sequence of one-element extensions $[P_{E^{-m+1}}, \ldots, P_{E^n}]$ based on Γ where each new covariance element added with each step in the subsequence $[P_{E^{-m+1}}, \ldots, P_{E^0}]$ is predetermined (because P_{E^0} is given). If Γ is an efficient sequence then the complexity of the ith extension from $P_{E^{i-1}}$ to P_{E^i} is no larger than $O(|U^i|)$. In the ME case, the complexity is a constant size, independent of $|U^i|$. □

The purpose of introducing the subsequence of one-element extensions $[P_{E^{-m+1}}, \ldots, P_{E^0}]$ is to ensure that the normal equations associated with these one-element extensions have been solved and are, therefore, available to be used in the efficient computation of subsequent one-element extension steps. For example, the first such step from P_{E^0} to P_{E^1} requires normal equations associated with $[P_{E^{-m+1}}, \ldots, P_{E^0}]$ (in the same way that (17) and (19) can be used to efficiently solve higher order normal equations, as shown in Proposition 7).

Proof: If our chordal sequence is efficient then the normal equations that arise in each one-element extension can be solved using Proposition 7. As discussed in Comment 1, complexity of solution of the normal equations that arise in the ith such step is $O(|U^i|)$. Also, as has been discussed, in the ME case all that is required is zero padding so that the complexity of each one-element extension is a constant, independent of the cardinality of U^i. □

V. Application to MAR Modeling

We now return to the MAR model construction problem, described in Section II, which corresponds to the following covariance extension problem. Let $G^0 = (V, E^0)$ and $G^n = (V, E^n)$ be graphs such that

$$V \triangleq \{0, 1, \ldots, N-1\} \tag{33a}$$

$$E^0 \triangleq \{(m, n) \in V \times V \,|\, |m - n| \le k\} \tag{33b}$$

$$E^n \triangleq \bigcup_{s \neq 0} \mathcal{C}_s \times \mathcal{C}_s \tag{33c}$$

where $\mathcal{C}_s$ is as defined in (6). To build an MAR model for the ME *completion* of P_{E^0}, we require the *extension* of P_{E^0} to P_{E^n}.

We now discuss the application of our generalized Levinson algorithm to the computation of P_{E^n}. For the moment, assume that 1) $E^0 \subset E^n$, and 2) $G^n = (V, E^n)$ is chordal. We will prove these two facts shortly. Assuming these, there exists a chordal sequence from G^0 to G^n. We may, therefore, find the elements of P_{E^n} using the generalized Levinson algorithm of Section IV by setting the unspecified generalized reflection coefficients to zero. We now show that doing so results in $O(N)$ computational complexity even when the chordal sequence upon which we base our one-element extensions is not efficient. Recall that if our chordal sequence is not efficient we must solve some (in the worst case, all) of the normal equations that arise in the generalized Levinson algorithm explicitly, rather than using Proposition 7. However, the largest maximal clique of G^n has cardinality $|\mathcal{C}_s| < 8k$. Hence, at worst, each one-element extension requires $O(k^3)$ computations. This leads to an overall complexity of $O(N)$ for computing P_{E^n}. The complexity of computing P_{E^n} using an efficient chordal sequence is also

$O(N)$ so there is no computational advantage (asymptotically) in using an efficient sequence. In the sequel, we do not assume that our chordal sequence is efficient. Note that the preceding discussion shows that the complexity of computing an MAR model for P^{ME} is of the same order as computing an AR model using the classical Levinson algorithm (in the nonstationary case).

The theory developed in previous sections can be applied to this covariance extension problem if two conditions are satisfied: 1) $E^0 \subset E^n$ and 2) the final graph G^n is chordal. That the original graph G^0 is chordal is both well known and clear. We now establish that $E^0 \subset E^n$.

Proposition 9: $E^0 \subset E^n$ where these are defined by (33).

Proof: The proof follows from the fact that, by construction, every interval of length $k+1$ is contained in one of the sets $\mathcal{C}_s$ (cf., the Appendix). This is most easily seen by considering Fig. 3 which represents an MAR model for a length-16, first-order Markov process. It is clear from this figure that every interval of length 2 can be found in some $\mathcal{C}_s$. The algebraic details proving this fact for a general Markov process are found in [18]. □

We now show that G^n is chordal.

Proposition 10: $G^n = (V, E^n)$ is chordal where V and E^n are defined in (33). □

To prove Proposition 10, we will rely on the following.

Definition 9 (Junction Tree): A *junction tree* for a graph $G = (V, E)$ is a tree $T = (\mathcal{K}, \mathcal{E})$ whose vertex set $\mathcal{K}$ is the set of maximal cliques of G and where for any $v \in V$, each induced subgraph $T_{\mathcal{K}_v}$ is connected, where $\mathcal{K}_v$ consists of those maximal cliques of G that contain v. □

The following provides a connection between chordal graphs and junction trees.

Proposition 11 [23]: G is chordal if and only if there exists a junction tree for G. □

Proof (Proposition 10): By Proposition 11 it suffices to exhibit a junction tree for G^n. To this end, let $T = (\mathcal{K}, \mathcal{E})$ be a graph where $\mathcal{K}$ is the set of maximal cliques of G^n. That is,

$$\mathcal{K} = \{\mathcal{C}_s\}_{s \neq 0}. \tag{34}$$

Let $\mathcal{E}$ be the following set of edges between elements of $\mathcal{K}$:

$$\bigcup_{t\text{: a tree node}} \left(\{(\mathcal{C}_{t\alpha_1}, \mathcal{C}_{t\alpha_2})\} \cup \{(\mathcal{C}_{t\alpha_1}, \mathcal{C}_{t\alpha_1\alpha_2})\} \cup \{(\mathcal{C}_{t\alpha_2}, \mathcal{C}_{t\alpha_2\alpha_1})\}\right). \tag{35}$$

An illustration of T is provided in Fig. 6. The junction tree associated with the specific MAR model illustrated in Fig. 3 is shown in Fig. 7. It is straightforward but notationally tedious to verify that T is a junction tree for G^n. The formal details are omitted and may be found in [18]. □

While we have completed the formal description of how to build an MAR model for the ME extension of a banded, partial covariance matrix, there is one remaining algorithmic detail—that of finding a chordal sequence between G^0 and G^n. For the problem at hand, there is a particular chordal sequence with appealing scale-recursive structure.

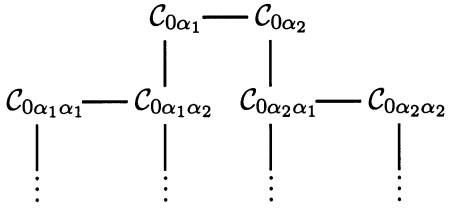

Fig. 6. The junction tree T described in the Proof of Proposition 10.

Proposition 12: Using the notation previously defined, there exists a chordal sequence from G^0 to G^n such that the edges of $\mathcal{C}_s$ are added prior to those of $\mathcal{C}_t$ for every node s and t such that s is at a finer scale than t. □

Proposition 12 permits us to find the joint child–parent statistics (and, hence, the parameters) of our MAR model scale-recursively, beginning at the finest scale and proceeding to each successive coarser scale. For example, referring to Fig. 3, the child–parent joint statistics

$$\begin{pmatrix} P_{x(s)} & P_{x(s)x(s\overline{\gamma})} \\ P^T_{x(s)x(s\overline{\gamma})} & P_{x(s\overline{\gamma})} \end{pmatrix} \tag{36}$$

are first computed for all child–parent pairs linked by solid lines. Then, proceeding to the next coarser scale, the joint child–parent statistics are computed for the pairs linked by dashed lines. We refer the reader to [18] for a proof of Proposition 12 and additional details regarding this particular chordal sequence. Once the joint child–parent statistics have been found, they are used in (3) to compute the MAR parameters for the ME extension. This MAR model is guaranteed to be internal by construction (due to properties of the matrices L_s described in Section II and in [18], [19]).

VI. Conclusion

This paper provides two important contributions in the area of covariance extension. First, we have generalized the classical Levinson algorithm to accommodate a broader range of extension problems. In particular, our generalized Levinson algorithm can address any extension problem for which the known values in the initial partial covariance matrix and the computed values in the final partial covariance matrices correspond to chordal graphs. We have also characterized the conditions under which our generalized Levinson algorithm is computationally efficient.

Our second contribution was to use our generalized Levinson algorithm to solve efficiently (with a complexity linear in problem size) the nonstandard extension problem that arises in the identification of an MAR model for the ME completion of a banded, partial covariance matrix. Like AR models for standard time series, the implicit covariance characterization that MAR models provide can be exploited to achieve efficiencies in computation and storage. The connection between AR processes and ME completions is well known and is provided by the classical Levinson algorithm. The connection to MAR processes has, until our work, been unknown and, as we have shown, relies on a generalization of the classical Levinson algorithm.

MAR models represent signals at multiple resolutions and are capable of optimally fusing multiresolution data. Therefore,

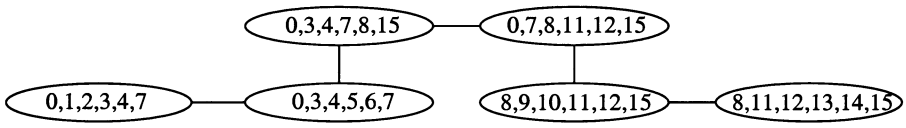

Fig. 7. The junction tree associated with the MAR model illustrated in Fig. 3. Numbers in each ellipse indicate the sample indexes associated with each maximal clique.

they is a natural framework in which to consider *multiresolution covariance extension*. The multiresolution covariance extension problem is one for which in addition to the local covariance information, some coarse-scale information is also provided. For instance, consider a problem in which diagonal bands of a covariance matrix are provided and, in addition, covariances between coarse-scale averages are provided. The incorporation of this coarse-scale information leads to an extension problem with linear constraints on the unknown covariance elements. While some preliminary work has been done on this topic [45], there are a number of challenges associated with building an MAR model for such an extension. For example, referring to Fig. 2 and (1), suppose that a component of a coarser scale state variable $x(s)$ represents a weighted average of a window of the finest scale samples of the process being modeled. Then, there is an implied constraint on the coarse-to-fine dynamics of (1) in order to guarantee that the coarser scale variable does indeed equal the weighted average of fine-scale samples with probability one. While this consistency problem has been addressed in another context [18], [19], its consideration in the context of the covariance extension problem is an open research topic.

APPENDIX
DESCRIPTION OF MAR ENDPOINT MODEL AND PROOF OF PROPOSITION 1

For convenience,[12] we assume that $N = 4k2^M$ for positive integer M. A state $x(s)$ where s resides at the mth scale and ith shift[13] in an $(M+1)$-scale MAR model for a kth-order Markov process consists of the samples of the process indexed by $\eta(s) = \eta_1(s) \cup \eta_2(s) \cup \eta_3(s)$ where

$$\eta_1(s) = i4k2^{M-m} + [0,\, k-1] \tag{37a}$$

$$\eta_2(s) = i4k2^{M-m} + 4k2^{M-m-1} + [-k,\, k-1] \tag{37b}$$

$$\eta_3(s) = i4k2^{M-m} + 4k2^{M-m} + [-k,\, -1] \tag{37c}$$

and $c + [a,\, b] \triangleq [a+c,\, b+c]$.

Proof (Proposition 1): The elements of P required to compute the parameters for an MAR model just described are characterized as follows. To determine the parameters, we require the joint child–parent statistics (cf., (3)). A given child–parent pair includes the samples of f indexed by $\mathcal{C}_s \triangleq \eta(s) \cup \eta(s\overline{\gamma})$. Hence, for this child–parent pair, we require knowledge of the principal submatrix of P whose rows and columns are indexed by $\mathcal{C}_s$. So, to build the entire model we require knowledge of the principal submatrix of P whose rows and columns are indexed by $\bigcup_{s\neq 0} \mathcal{C}_s$. Since $|\eta(s)| = 4k$ we have that $|\mathcal{C}_s| \leq 8k$ and $|\mathcal{C}_s \times \mathcal{C}_s| \leq 64k^2$. These are upper bounds because the sets $\eta(s)$ as well as the sets $\mathcal{C}_s \times \mathcal{C}_s$ are not mutually exclusive, which can be seen by inspection of Fig. 3. Therefore, the total number of elements of P required is bounded above by $64k^2(2^{M+1}-1)$. Finally, since $N = 4k2^M$, we have that the total number of elements required is bounded above by $32kN$. While this is an upper bound, it is of the right order and it indicates that we require only $O(kN)$ elements of P. □

[12]It is a simple matter of bookkeeping to accommodate a process with length $N \neq 4k2^M$. In particular, the mapping of elements to nodes will be less regular and some nodes will have more or fewer than $4k$ elements [43].

[13]The root node resides at scale 0, the next finer scale is scale 1, etc. For a given scale m, the leftmost node has shift $i = 0$, the second leftmost node has shift $i = 1$, etc.

REFERENCES

[1] P. Brockwell and R. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1987.
[2] J. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Annu. Int. Meeting of The Society of Exploratory Geophysics*, Oklahoma City, OK, Oct. 1967.
[3] ——, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975.
[4] K. Chou, A. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 464–478, Mar. 1994.
[5] K. Chou, A. Willsky, and R. Nikoukhah, "Multiscale systems, Kalman filters, and Riccati equations," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 479–492, Mar. 1994.
[6] T. Constantinescu, *Schur Parameters, Factorization, and Dilation Problems*. Basel, Switzerland: Birkhauser Verlag, 1991.
[7] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[8] M. Daniel and A. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proc. IEEE*, vol. 85, pp. 164–180, Jan. 1997.
[9] M. Daniel, A. Willsky, and D. McLaughlin, "Travel time estimation using a multiscale stochastic framework," *Adv. Water Resources*, vol. 23, no. 6, pp. 571–665, May 2000.
[10] K. Daoudi, A. Frakt, and A. Willsky, "Multiscale autoregressive models and wavelets," *IEEE Trans. Inform. Theory*, vol. 45, pp. 828–845, Apr. 1999.
[11] J. Darroch, S. Lauritzen, and T. Speed, "Markov fields and log-linear interaction models for contingency tables," *Ann. Statist.*, vol. 8, no. 3, pp. 522–539, 1980.
[12] P. Dewilde and E. Deprettere, "Modeling VLSI interconnections as an inverse scattering problem," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1987, pp. 147–153.
[13] P. Fieguth, W. Karl, A. Willsky, and C. Wunsch, "Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 280–292, Mar. 1995.
[14] P. Fieguth, D. Menemenlis, T. Ho, A. Willsky, and C. Wunsch, "Mapping Mediterranean altimeter data with a multiresolution optimal interpolation algorithm," *J. Atmosph. Ocean. Technol.*, vol. 15, pp. 535–546, Apr. 1998.
[15] P. Fieguth and A. Willsky, "Fractal estimation using models on multiscale trees," *IEEE Trans. Signal Processing*, vol. 44, pp. 1297–1300, May 1996.
[16] P. Fieguth, A. Willsky, and W. Karl, "Efficient multiresolution counterparts to variational methods for surface reconstruction," *Comput. Vision Image Understand.*, vol. 70, no. 2, pp. 157–176, May 1998.
[17] C. Fosgate, H. Krim, W. Irving, and A. Willsky, "Multiscale segmentation and anomaly enhancement of SAR imagery," *IEEE Trans. Image Processing*, vol. 6, pp. 7–20, Jan. 1997.
[18] A. Frakt, "Internal multiscale autoregressive processes, stochastic realization, and covariance extension," Ph.D. dissertation, MIT, Cambridge, MA, Aug. 1999.

[19] A. Frakt and A. Willsky, "Computationally efficient stochastic realization for internal multiscale autoregressive models," *Multidim. Syst. Signal Processing*, vol. 12, pp. 109–142, 2001.
[20] B. Friedlander, "Lattice methods for spectral estimation," *Proc. IEEE*, vol. 70, pp. 990–1017, Sept. 1982.
[21] W. Glunt, T. Hayden, C. Johnson, and P. Tarazaga, "Positive definite completions and determinant maximization," *Linear Alg. Its Applic.*, vol. 288, pp. 1–10, 1999.
[22] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. London, U.K.: Johns Hopkins Univ. Press, 1996.
[23] M. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*. New York: Academic, 1980.
[24] J. Grandell, M. Hamrud, and P. Toll, "A remark on the correspondence between the maximum-entropy method and the autoregressive model," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 750–751, Nov. 1980.
[25] R. Grone, C. Johnson, E. Sa, and H. Wolkowicz, "Positive definite completions of partial Hermitian matrices," *Linear Alg. Its Applic.*, vol. 58, pp. 109–124, 1984.
[26] M. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: Wiley, 1996.
[27] T. Ho, P. Fieguth, and A. Willsky, "Multiresolution stochastic models for the efficient solution of large-scale space–time estimation problems," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 6, Atlanta, GA, May 1996, pp. 3097–3100.
[28] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[29] S. Ihara, "Maximum entropy spectral analysis and ARMA processes," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 377–380, Mar. 1984.
[30] W. Irving, L. Novak, and A. Willsky, "A multiresolution approach to discriminating targets from clutter in SAR imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 1157–1169, Oct. 1997.
[31] W. Irving and A. Willsky, "A canonical correlations approach to multiscale stochastic realization," *IEEE Trans. Automat. Contr.*, vol. 46, pp. 1514–1528, Oct. 2001.
[32] E. Jaynes, "On the rationale of maximum-entropy methods," *Proc. IEEE*, vol. 70, pp. 939–952, Sept. 1982.
[33] C. Johnson, "Matrix completion problems: A survey," in *Matrix Theory and Applications, Proceedings of Symposia in Applied Mathematics*. Providence, RI, 1990.
[34] A. Kannan, "Adaptation of spectral trajectory models for LVCSR," Ph.D. dissertation, Boston Univ., Boston, MA, 1997.
[35] A. Kannan and S. Khudanpur, "Tree-structured models of parameter dependence for rapid adaptation in large vocabulary conversational speech recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Phoenix, AZ, 1999.
[36] A. Kannan and M. Ostendorf, "Modeling dependence in adaptation of acoustic models using multiscale tree processes," in *Proc. EUROSPEECH*, 1997, pp. 1863–1866.
[37] A. Kim and H. Krim, "Hierarchical stochastic modeling of SAR imagery for segmentation/compression," *IEEE Trans. Signal Processing*, vol. 47, pp. 458–468, Feb. 1999.
[38] P. Kumar, "A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 182–197, Jan. 1999.
[39] H. Lev-Ari, "Nonstationary lattice-filter modeling," Ph.D. dissertation, Stanford Univ., Stanford, CA, Dec. 1983.
[40] H. Lev-Ari and T. Kailath, "Schur and Levinson algorithms for nonstationary processes," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Atlanta, 1981, pp. 860–864.
[41] H. Lev-Ari, S. Parker, and T. Kailath, "Multidimensional maximum-entropy covariance extension," *IEEE Trans. Inform. Theory*, vol. 35, pp. 497–508, May 1989.
[42] M. Luettgen, W. Karl, and A. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Processing*, vol. 3, pp. 41–64, Jan. 1994.
[43] M. Luettgen, W. Karl, A. Willsky, and R. Tenney, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, vol. 41, pp. 3377–3396, Dec. 1993.
[44] M. Luettgen and A. Willsky, "Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination," *IEEE Trans. Image Processing*, vol. 4, pp. 194–207, Feb. 1995.
[45] M. Lundquist and C. Johnson, "Linearly constrained positive definite completions," *Linear Alg. Its Applic.*, vol. 150, pp. 195–208, 1991.
[46] A. Papoulis, "Maximum entropy and spectral estimation: A review," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1176–1186, Dec. 1981.
[47] D. Rose, R. Tarjan, and G. Lueker, "Algorithmic aspects of vertex elimination on graphs," *SIAM J. Comput.*, vol. 5, no. 2, pp. 266–283, June 1976.
[48] M. Schneider, P. Fieguth, W. Karl, and A. Willsky, "Multiscale methods for the segmentation of images," *IEEE Trans. Image Processing*, vol. 9, pp. 456–468, Mar. 2000.
[49] J. Schroeder and D. Howard, "Multiscale modeling for target detection in complex synthetic aperture radar," in *Proc. Asilomar Conf. Signals, Systems, and Signal Processing*, Nov. 1998.
[50] T. Speed and H. Kiiveri, "Gaussian Markov distributions of finite graphs," *Ann. Statist.*, vol. 14, no. 1, pp. 138–150, 1986.
[51] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[52] R. Tarjan and M. Yannakakis, "Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs," *SIAM J. Comput.*, vol. 13, no. 3, pp. 566–579, 1984.