Multiscale Systems, Kalman Filters, and Riccati Equations

Kenneth C. Chou, *Member, IEEE,* Alan S. Willsky, *Fellow, IEEE,* and Ramine Nikoukhah *Member, IEEE*

***Abstract*—In [1] we introduced a class of multiscale dynamic models described in terms of scale-recursive state space equations on a dyadic tree. An algorithm analogous to the Rauch–Tung–Striebel algorithm—consisting of a fine-to-coarse Kalman filter-like sweep followed by a coarse-to-fine smoothing step—was developed. In this paper we present a detailed system-theoretic analysis of this filter and of the new scale-recursive Riccati equation associated with it. While this analysis is similar in spirit to that for standard Kalman filters, the structure of the dyadic tree leads to several significant differences. In particular, the structure of the Kalman filter error dynamics leads to the formulation of an ML version of the filtering equation and to a corresponding smoothing algorithm based on triangularizing the Hamiltonian for the smoothing problem. In addition, the notion of stability for dynamics requires some care as do the concepts of reachability and observability. Using these system-theoretic constructs, we are then able to analyze the stability and steady-state behavior of the fine-to-coarse Kalman filter and its Riccati equation.**

I. Introduction

The use of pyramidal representations for signals and images has been and continues to be of considerable interest, both in research and in application. The reasons for this include the computational efficiencies that such representations may suggest (e.g., as in the use of multigrid methods for the solution of partial differential equations [16], [17]), the fact that many phenomena including those with fractal or self-similar features can be captured in natural and analytically useful ways in this setting [11], [12], and the development of the wavelet transform [13]–[15] which has sparked interest in developing multiresolution methods for a vast array of applications. As described in [1], [10], the interest in multiresolution representations and its apparently substantial promise provided motivation for the development of a framework for statistical modeling and optimal processing based on such pyramidal representations. In particular in [1] we introduced a class of multiscale state-space models evolving on dyadic trees (in which each level in the tree corresponds to a particular level of resolution in signal representation), we derived an efficient and highly parallelizable optimal estimation algorithm on the dyadic tree, and we illustrated the potential of this framework both for problems of optimal fusion of multiresolution data and for the efficient solution of computationally intensive problems of signal and image analysis through the use of "fractal regularization" techniques based on our models. In [18], the straightforward extension of our algorithm to quadtrees is used to achieve computational reductions of between one and two orders of magnitude for a typical image processing/computer vision problem, while in [19] we demonstrate that the classes of processes that can be captured in this setting are quite rich, including all Gauss–Markov processes and Gaussian–Markov random fields.

Manuscript received December 16, 1991; revised February 16, 1991. Recommended by Past Associate Editor W. S. Wong. The work of these authors was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002, by the National Science Foundation under Grants MIP-9015281 and INT-9002393, and by the Office of Naval Research under Grant N00014-91-J-1004.

K. C. Chou is with SRI International, Menlo Park, CA 94025.

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

R. Nikoukhah is with INRIA, Domaine de Voluceau, Rocquencourt, BP105, 78153 Le Chesnay, Cedex, France.

IEEE Log Number 9213719.

All of this, we feel, not only establishes the promise of this new framework but also identifies additional system-theoretic questions of some importance. In particular, the optimal estimation algorithm [1] is a direct generalization of Kalman filtering and state-space smoothing algorithms, introducing a new class of scale-recursive Riccati equations. This suggests, among other things, the development of a system theory for multiresolution modeling and realization as well as the detailed system-theoretic analysis of the filtering and Riccati equations introduced in [1]. The objective of this paper is to tackle this latter problem, while an initial investigation of multiscale deterministic realization theory is the subject of [2].

In the next section we briefly review the multiscale state-space model and optimal estimation algorithm of [1]. The objective of error and stability analysis for multiscale filtering leads directly to a variation on this algorithm which we develop in Section III. This "ML algorithm" also has a direct connection with the solution of the estimation problem via the triangularization of the smoothing Hamiltonian, which we describe in an appendix. In Section IV we then turn to the system-theoretic analysis of our models and, as we will see, the notions of reachability, observability, and, especially, stability have significant variations as compared to their counterpart for ordinary state-space models. These tools are then used in Section V where we analyze the properties of the error covariance for our optimal filter and the stability and asymptotic behavior of the filter error dynamics and our new Riccati equation.

II. State-Space Models and Multiscale Estimation on Dyadic Trees

As illustrated in Fig. 1, the basic data structure for multiresolution modeling is the dyadic tree. Here each node t in the tree

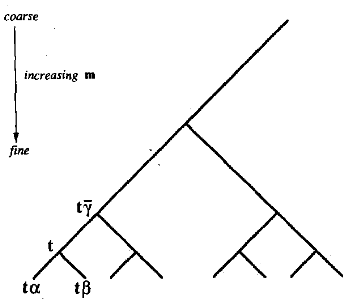


Fig. 1. The dyadic tree and some notation used in the paper.

T corresponds to a pair of integers (m, n), where m denotes the scale corresponding to node t and n its translational offset. Thus, if $z(t)$ denotes a signal defined on T, then the restriction of z to any particular level, i.e., the collection of values of $z(t)$ for $t = (m, n)$ with m fixed, corresponds to the representation of a signal (viewed as a function of n) at the mth scale. It is useful to visualize T as having horizontal levels corresponding to different scales, where increasing m corresponds to moving to finer resolutions. We will use the more compact notation t for nodes on T and will denote the scale of a particular node t by $m(t)$. Also, as illustrated in the figure, there are natural shift operators on T, namely the unique backward shift $\bar{\gamma}$ and two forward shifts α and β. In particular if $t = (m, n)$, then $t\alpha = (m{+}1, 2n), t\beta = (m{+}1, 2n{+}1)$, and $t\bar{\gamma} = (m-1, [n/2])$. The basic picture one should have is that finer scales introduce additional detail into the signal representation, while coarser scales involve successively decimated and lower resolution representations (see [1] for further discussion and references).

There are two alternate classes of scale-recursive linear dynamic models that are of interest. The first of these is the class of coarse-to-fine state space models on T

$$x(t) = A(t)x(t\bar{\gamma}) + B(t)w(t) \tag{2.1}$$

$$y(t) = C(t)x(t) + v(t). \tag{2.2}$$

The term $A(t)x(t\bar{\gamma})$ in (2.1) represents a coarse-to-fine interpolation, $B(t)w(t)$ represents the higher resolution detail added in going from one scale to the next, and $y(t)$ is the measured variable at the particular scale m and location n represented by t. This model is the basis for multiscale modeling of stochastic processes developed in [1]. In contrast, the fine-to-coarse Kalman filtering step of our estimation algorithm falls into the class of fine-to-coarse recursive models of the form

$$\begin{aligned} x(t) = F_1(t\alpha)x(t\alpha) + F_2(t\beta)x(t\beta) \\ +G(t\alpha)w(t\alpha) + G(t\beta)w(t\beta). \end{aligned} \tag{2.3}$$

An important special case of (2.1)–(2.3) is that the system parameters are constant at each scale but may vary from scale to scale, in which case we abuse notation by writing $A(t) = A(m(t))$, etc. Such a model is useful for capturing scale-dependent effects and fractal behavior [1], [11]. For simplicity we focus the detailed covariance analysis and stability results on this case. Also, if we wish to consider representations of signals of unbounded extent, we must deal with the full infinite tree T, i.e., $\{(m, n)| - \infty < m, n < \infty\}$. This will be of interest when we consider asymptotic properties such as stability and steady-state behavior. In any practical application, of course, we must deal with a compact interval of data. In this case, the index set of interest represents a finite version of the tree of Fig. 1, consisting of $M + 1$ levels beginning with the coarsest scale represented by a unique root node, denoted by 0, and M subsequent levels, the finest of which has 2^M nodes.

Suppose that $w(t)$ and $v(t)$ are independent, zero-mean white noise processes with covariances I and $R(t)$, respectively. The covariance $P_x(t) = E[x(t)x^T(t)]$ then evolves according to a Lyapunov equation on the tree:

$$P_x(t) = A(t)P_x(t\bar{\gamma})A^T(t) + B(t)B^T(t). \tag{2.4}$$

If the model parameters vary in scale only and if at some scale $P_x(t) = P_x(m(t))$, then this holds at each scale, and

$$P_x(m + 1) = A(m)P_x(m)A^T(m) + B(m)B^T(m). \tag{2.5}$$

If we further specialize our model to the case in which A and B are constant, and if A is stable, then (2.5) admits a steady-state solution, to which $P_x(m)$ converges, which is the unique solution of the usual algebraic Lyapunov equation.

In [1] we also encounter the reversal of (2.1), *i.e.*, a model representing $x(t\bar{\gamma})$ as a linear function of $x(t)$ and a noise that is uncorrelated with $x(t)$ is given by

$$x(t\bar{\gamma}) = F(t)x(t) - A^{-1}(t)B(t)\tilde{w}(t) \tag{2.6}$$

$$\begin{aligned} F(t) &= A^{-1}(t)[I - B(t)B^T(t)P_X^{-1}(t)] \\ &= P_x(t\bar{\gamma})A^T(t)P_x^{-1}(t) \end{aligned} \tag{2.7}$$

$$\tilde{w}(t) = w(t) - E[w(t)|x(t)] \tag{2.8}$$

$$E[\tilde{w}(t)\tilde{w}^T(t)] = I - B^T(t)P_x^{-1}(t)B(t) \triangleq \tilde{Q}(t). \tag{2.9}$$

In [1] we derive a generalization of the Rauch–Tung–Striebel (RTS) smoothing algorithm consisting of a fine-to-coarse Kalman filtering step followed by coarse-to-fine smoothing step. Let $\hat{x}(s|t)$ denote the optimal estimate of $x(s)$ based on data Y_t at or "below" node t (i.e., $y(\tau)$ for $\tau = t$ or τ a descendent of t), and let $\hat{x}(s|t+)$ denote the optimal estimate of $x(s)$ based on data strictly "below" t (i.e., $y(\tau)$ for τ a strict descendent of t). Let $P(s|t)$ and $P(s|t+)$ be the corresponding error covariances. Then the coarse-to-fine Kalman filter consists of a measurement update step

$$\hat{x}(t|t) = \hat{x}(t|t+) + K(t)[y(t) - C(t)\hat{x}(t|t+)] \tag{2.10}$$

$$K(t) = P(t|t+)C^T(t)V^{-1}(t) \tag{2.11}$$

$$V(t) = C(t)P(t|t+)C^T(t)R(t) \tag{2.12}$$

$$P(t|t) = [I - K(t)C(t)]P(t|t+). \tag{2.13}$$

a coarse-to-fine one-step prediction step

$$\hat{x}(t|t\alpha) = F(t\alpha)\hat{x}(t\alpha|t\alpha) \tag{2.14}$$

$$P(t|t\alpha) = F(t\alpha)P(t\alpha|t\alpha)F^T(t\alpha) + \mathcal{Q}(t\alpha) \tag{2.15}$$

$$\mathcal{Q}(t\alpha) = A^{-1}(t\alpha)B(t\alpha)\tilde{Q}(t\alpha)B^T(t\alpha)A^{-T}(t\alpha) \tag{2.16}$$

with analogous equations for $\hat{x}(t|t\beta)$ and $P(t|t\beta)$ obtained by replacing $t\alpha$ with $t\beta$ in (2.14)–(2.16), and a fusion step to merge these estimates to form $\hat{x}(t|t+)$:

$$\hat{x}(t|t+) = P(t|t+)[P^{-1}(t|t\alpha)\hat{x}(t|t\alpha) + P^{-1}(t|t\beta)\hat{x}(t|t\beta)] \tag{2.17}$$

$$P(t|t+) = [P^{-1}(t|t\alpha) + P^{-1}(t|t\beta) - P_x^{-1}(t)]^{-1}. \tag{2.18}$$

This algorithm has a pyramidal structure, allowing substantial parallelization. Also, while the update and prediction steps are analogous to corresponding steps in usual Kalman filtering,[1] the fusion step has no counterpart.

Let $\hat{x}_s(t)$ denote the estimate of $x(t)$ based on all data on a finite subtree with root node 0 and M scales below it. Once the Kalman filter reaches the root node, $\hat{x}_s(0) = \hat{x}(0|0)$ serves as the initial condition for the coarse-to-fine smoothing sweep:

$$\hat{x}_s(t) = \hat{x}(t|t) + J(t)[\hat{x}_s(t\bar{\gamma}) - \hat{x}(t\bar{\gamma}|t)] \tag{2.19}$$

$$J(t) \triangleq P(t|t)F^T(t)P^{-1}(t\bar{\gamma}|t) \tag{2.20}$$

where $P_s(t)$, the smoothing error covariance, satisfies

$$P_s(t) = P(t|t) + J(t)[P_s(t\bar{\gamma}) - P(t\bar{\gamma}|t)]J^T(t). \tag{2.21}$$

III. The ML Filter

The Riccati equation (2.11)–(2.13), (2.15), and (2.18) differs from standard Riccati equations in two respects: 1) the explicit presence of the prior state covariance $P_x(t)$ and 2) the fusion of two sources of information in (2.18). The latter of these is intrinsic to our Riccati equations and has important consequences in the stability analysis of fine-to-coarse filtering. The presence of $P_x(t)$, on the other hand, points to an apparent complication in analyzing our filter that motivates an alternate filtering algorithm in which it does not appear. Specifically, in standard Kalman filtering analysis, the error evolves as a state process itself without explicitly coupling to $x(t)$. This is not the case here because of the explicit presence of $P_x(t)$ in (2.18) and in the backward model parameters (2.6)–(2.9) that enter into the fine-to-coarse prediction step (2.15). On first examination, this might not appear to be a new problem, as backward models for standard temporal models also involve the state covariance. The present situation, however, is not as simple, thanks to the new fusion step. If we examine the backward model (2.6)–(2.9) and the Kalman filter (2.10), (2.14), (2.17), we find that the upward dynamics for the error $x(t) - \hat{x}(t|t)$ are *not* decoupled from $x(t)$ unless $P_x^{-1}(t) = 0$. Thus we apparently have a significant difference in analyzing these error dynamics. To overcome this, we consider a slight variation in the algorithm.

[1] Although, as discussed in [1] this step **must** proceed from fine-to-coarse and, hence, must use the backward model (2.6) for the prediction step.

Specifically, we define what we will refer to as the *ML filter* by setting the $P_x^{-1}(t)$ terms in (2.10)–(2.18) to zero. The resulting filter recursions are then given by the following.

Measurement Update:

$$\hat{x}_{ML}(t|t) = \hat{x}_{ML}(t|t+) + K_{ML}(t)[y(t) - C(t)\hat{x}_{ML}(t|t+)] \tag{3.1}$$

$$K_{ML}(t) = P_{ML}(t|t+)C^T(t)V_{ML}^{-1}(t) \tag{3.2}$$

$$V_{ML}(t) = C(t)P_{ML}(t|t+)C^T(t) + R(t) \tag{3.3}$$

$$P_{ML}(t|t) = [I - K_{ML}(t)C(t)]P_{ML}(t|t+) \tag{3.4}$$

One-Step Prediction:

$$\hat{x}_{ML}(t\bar{\gamma}|t) = A^{-1}(t)\hat{x}_{ML}(t|t) \tag{3.5}$$

$$P_{ML}(t\bar{\gamma}|t) = A^{-1}(t)P_{ML}(t|t)A^{-T}(t) + A^{-1}(t)B(t)A^{-T}(t) \tag{3.6}$$

Merge Step:

$$\hat{x}_{ML}(t|t+) = P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)\hat{x}_{ML}(t|t\alpha) + P_{ML}^{-1}(t|t\beta)\hat{x}_{ML}(t|t\beta)] \tag{3.7}$$

$$P_{ML}^{-1}(t|t+) = P_{ML}^{-1}(t|t\alpha) + P_{ML}^{-1}(t|t\beta) \tag{3.8}$$

The key differences, here are the absence of a $P_x^{-1}(t)$ term in (3.8) (compare to (2.18)) and the changes to the prediction step.

As shown in Appendix A, the ML estimate of $x(t)$ based on Y_t does indeed satisfy (3.1)–(3.8), and standard results [4] on the relationship between ML and Bayesian estimates yield

$$\hat{x}(t|t) = P(t|t)P_{ML}^{-1}(t|t)\hat{x}_{ML}(t|t) \tag{3.9}$$

$$P^{-1}(t|t) = P_{ML}^{-1}(t|t) + P_x^{-1}(t). \tag{3.10}$$

Note that this provides us with an alternative RTS-like algorithm: we apply the fine-to-coarse ML filter (3.1)–(3.8) from the finest scale M up to the top of the tree, i.e., through the computation of $\hat{x}_{ML}(0|0), P_{ML}(0|0)$. We then incorporate prior information at the top of the tree, using (3.9), (3.10) to yield $\hat{x}_s(0) = \hat{x}(0|0)$ and $P_s(0) = P(0|0)$. The downward smoothing sweep is then computed by adapting (2.19)–(2.21) [using (3.9), (3.10)] so that the ML estimator computed in the ML filtering sweep is used in the smoothing step. Specifically, as shown in [9]

$$\hat{x}_s(t) = \hat{x}_{ML}(t|t) + J(t)[\hat{x}_s(t\bar{\gamma}) - \hat{x}_{ML}(t\bar{\gamma}|t)] \tag{3.11}$$

$$P_s(t) = P_{ML}(t|t) + J(t)[P_s(t\bar{\gamma}) - P_{ML}(t\bar{\gamma}|t)]J^T(t) \tag{3.12}$$

$$J(t) = P_{ML}(t|t)A^{-T}(t)P_{ML}^{-1}(t\bar{\gamma}|t). \tag{3.13}$$

Note that one can perform exactly analogous calculations (without the merge step) for standard Kalman filtering problems, although in the present context we have the additional motivation of obtaining a form that yields an explicit error dynamic equation. Also as in the standard case, the ML filtering equations (3.1)–(3.8) cannot be directly used at the initial levels of recursion—i.e., for the finest level M and perhaps several levels above this—until the ML covariance is well defined. Rather the information form of this filter must be used, and this is also described in Appendix A. Note that as one might expect and as will be used in Section V, observability plays a central role in guaranteeing that the error covariance does become well defined. Also, in Appendix B we present an alternate viewpoint for the derivation of RTS-like algorithms, namely using the Hamiltonian equations for our estimation problem. As discussed in [6] [7], diagonalization of the Hamiltonian for standard state-space models leads to two-filter smoothing algorithms, while triangularization leads to the RTS algorithm. In our case, the structure of the tree adds a fundamental asymmetry to the Hamiltonian, which precludes diagonalization, but whose triangularization is possible, leading to the ML form of the RTS algorithm we have just described.

Finally, let us show that we can use the ML filter to obtain a dynamic representation for the filtering error that is decoupled from the state dynamics itself. Specifically, from (3.1)–(3.8) we can derive the following ML filter recursion

$$\begin{aligned}\hat{x}_{ML}(t|t) = {} & [I - K_{ML}(t)C(t)]P_{ML}(t|t+) \\ & \cdot [P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)\hat{x}(t\alpha|t\alpha) \\ & + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)\hat{x}(t\beta|\beta)] \\ & + K_{ML}(t)y(t). \end{aligned} \tag{3.14}$$

Also, from (2.1)

$$x(t) = A^{-1}(t\alpha)x(t\alpha) - A^{-1}(t\alpha)B(t\alpha)w(t\alpha) \tag{3.15}$$

with an analogous equation with $t\alpha$ replaced by $t\beta$, and thus, using (3.8)

$$\begin{aligned}x(t) &= P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)x(t) + P_{ML}^{-1}(t|t\beta)x(t)] \\ &= P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)x(t\alpha) \\ &\quad + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)x(t\beta)] \\ &\quad - P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)B(t\alpha)w(t\alpha) \\ &\quad + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)B(t\beta)w(t\beta)] \end{aligned} \tag{3.16}$$

and thus defining $\tilde{x}_{ML}(t|t) = x(t) - \hat{x}_{ML}(t|t)$, we obtain

$$\begin{aligned}\tilde{x}_{ML}(t|t) = {} & [I - K_{ML}(t)C(t)]P_{ML}(t|t+) \\ & \cdot [P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)\tilde{x}(t\alpha|t\alpha) \\ & + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)\tilde{x}(t\beta)\tilde{x}(t\beta|t|)] \\ & - P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)B(t\alpha)w(t\alpha) \\ & + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)B(t\beta)w(t\beta)] \\ & - K_{ML}(t)v(t) \end{aligned} \tag{3.17}$$

Equation (3.17) represents the filtering error as the state of such a fine-to-coarse system, as in (2.3), driven by white process and measurement noise. It is the stability of this system—in the scale-varying case—that is investigated in Section V.

IV. System-Theoretic Concepts for Fine-to-Coarse Dynamic Models

In this section we introduce and investigate the several system-theoretic concepts for dynamic systems on dyadic trees that are needed in Section V for the asymptotic analysis of the fine-to-coarse filtering algorithm. In particular, we focus here on the scale-varying version of the fine-to-coarse model (2.2), (2.3), namely

$$\begin{aligned}x(t) = {} & F(m(t)+1)[x(t\alpha) + x(t\beta)] \\ & + G(m(t)+1)[w(t\alpha) + w(t\beta)] \end{aligned} \tag{4.1}$$

$$y(t) = C(m(t))x(t). \tag{4.2}$$

Since we focus on deterministic properties in this section, $w(t)$ should be viewed as an input, and we have eliminated the measurement noise from (4.2). To simplify the discussion, we assume the $F(m)$ is invertible for all m.

A. Reachability and Observability

The first property we wish to investigate is reachability for the model (4.1), i.e., the ability to drive the system from any fine-scale initial condition to any coarse-scale target. Note that the number of descendent nodes below any node t_0 grows geometrically with scale: there are 2^M "initial conditions" affecting $x(t_0)$ and at a scale M levels finer than $x(t_0)$. Specifically, let

$$X_{M,t_0} \triangleq [x^T(t_0\alpha^M), x^T(t_0\beta\alpha^{M-1}), \cdots x^T(t_0\beta^M)]^T \tag{4.3}$$

$$W_{M,t_0} \triangleq [w^T(t_0\alpha)w^T(t_0\beta)\cdots w^T(t_0\alpha^M)\cdots w^T(t_0\beta^M)]^T. \tag{4.4}$$

X_{M,t_0} contains the 2^M points at the Mth level down that influence the value of $x(t_0)$. The vector W_{M,t_0} contains all inputs that influence $x(t_0)$ starting from initial condition X_{M,t_0}, i.e., $w(t)$ in the entire subtree down to M levels from t_0.

As always, in studying reachability, we can set $X_{M,t_0} = 0$, so that

$$x(t_0) = \mathcal{G}W_{M,t_0} \tag{4.5}$$

$$\begin{aligned}\mathcal{G} \triangleq {} & [\Psi(0)\Psi(0)\Psi(1)\Psi(1)\Psi(1)\Psi(1)\cdots \\ & \cdot \underbrace{\Psi(M-2)\cdots\Psi(M-2)}_{2^{M-1}\text{times}} \cdot \underbrace{\Psi(M-1)\cdots\Psi(M-1)}_{2^M\text{times}}] \end{aligned} \tag{4.6}$$

$$\Psi(i) \triangleq \phi(m(t_0), m(t_0)+i)G(m(t_0)+i+1) \tag{4.7}$$

$$\phi(m_1, m_2) \triangleq \begin{cases} I & m_1 = m_2 \\ F(m_1+1)\phi(m_1+1, m_2) & m_1 < m_2 \end{cases} \tag{4.8}$$

$$\phi(m-1, m) \triangleq F(m). \tag{4.9}$$

Let us define the *reachability Gramian* as

$$\begin{aligned} \mathcal{R}(t_0, M) &\triangleq \mathcal{G}\mathcal{G}^T \\ &= \sum_{i=0}^{M-1} 2^{i+1}\phi(m(t_0), m(t_0)+i) \\ &\quad \cdot G(m(t_0)+i+1) \\ &\quad \times G^T(m(t_0)+i+1) \\ &\quad \cdot \phi^T(m(t_0), m(t_0)+i). \end{aligned} \tag{4.10}$$

Since the rank of $\mathcal{G}$ equals the rank of $\mathcal{G}\mathcal{G}^T$, we see that we can reach any $x(t_0)$ from any X_{M,t_0} if and only if $\mathcal{R}(t_0, M)$ is invertible. Also we will refer to the system (4.1) as being *uniformly reachable* if there exists γ, $M_0 > 0$ so that

$$\mathcal{R}(t, M_0) \geq \gamma I \quad \text{for all } t. \tag{4.11}$$

Note that $\mathcal{R}(t_0, M)$ is the standard reachability gramian for the system

$$\begin{aligned} x(m) = \sqrt{2}F(m+1)x(m+1) \\ +\sqrt{2}G(m+1)u(m+1). \end{aligned} \tag{4.12}$$

The factor of $\sqrt{2}$ in (4.12) does not effect either reachability *or* uniform reachability. Thus, the usual conditions for temporal state-space models apply here as well. For example, if F and G are constant, then reachability and uniform reachability are equivalent to the usual condition, i.e., $\operatorname{rank}[G|FG|\cdots F^{n-1}G] = n$.

It is interesting to note that the structure of the tree adds a substantial level of asymmetry to the analysis of coarse-to-fine and fine-to-coarse systems. For example, for standard temporal systems there are two closely related notions, namely reachability (i.e., the ability to reach any state from any state) and controllability (i.e., the ability to reach *zero* from any state). If the state dynamic matrix is invertible, these are equivalent, and this is also true for the fine-to-coarse model (4.1). This is *not* true, however, for the coarse-to-fine model (*e.g.*, (2.1) or its scale-varying specialization). In particular, reachability for a coarse-to-fine model involves driving a single initial condition $x(t_0)$ to *any* possible value of the 2^M-point set of values in X_{M,t_0}. This is an extremely strong condition, in contrast to the condition of controllability, i.e., driving $x(t_0)$ to $X_{M,t_0} = 0$. While this is of no direct interest to us here (and we refer the reader to [9] for details), the dual of this property is.

Specifically, let us turn to the problem of determining the state given the knowledge of the input and output. In the standard temporal case, there are two notions—observability (i.e., the ability to determine the initial condition) and reconstructibility (i.e., the ability to determine the final state)—which coincide if the state dynamic matrix is invertible. The asymmetry of the tree certainly leads to a substantial difference for us. For coarse-to-fine dynamics, observability (i.e., determining the single coarse state from the subtree of data beneath it) is a much *weaker* notion that reconstructibility (i.e., determining the 2^M states at a fine scale based on the subtree of data above it). The exact opposite conditions hold for the fine-to-coarse model (4.1), (4.2) (i.e., reconstructing $x(t_0)$ based on the subtree of data below it is a much weaker condition than determining the 2^M states in X_{M,t_0} based on the data in the subtree above it). Fortunately for us, it is the weaker of these notions that we require here. Thus we focus on that case here and refer the reader to [9] for a full treatment.

Let us define

$$\begin{aligned} Y_{M,t_0} \triangleq [y^T(t_0)|y^T(t_0\alpha), y^T(t_0\beta)|\cdots \\ |y^T(t_0\alpha^M), \cdots y^T(t_0\beta^M)]^T. \end{aligned} \tag{4.13}$$

As always in studying reconstructibility and observability, superposition allows us to focus on the case when $W_{M,t_0} = 0$ in which case

$$Y_{M,t_0} = \mathcal{H}_M X_{M,t_0} \tag{4.14}$$

where the level-to-level partitioned form of $\mathcal{H}_M$ is

$$\left[\begin{array}{cccccccccccc}
\Theta(0) & \Theta(0) & \cdots & & & & & & & & \cdots & \Theta(0) \\
\Theta(1) & \cdots & & & \cdots & \Theta(1) & 0 & \cdots & & & \cdots & 0 \\
0 & \cdots & & & \cdots & 0 & \Theta(1) & & & \cdots & \cdots & \Theta(1) \\
\Theta(2) & \cdots & \Theta(2) & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & \Theta(2) & \cdots & \Theta(2) & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 & \Theta(2) & \cdots & \Theta(2) & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \Theta(2) & \cdots & \Theta(2) \\
 & & & & \vdots & & & \vdots & & & & \\
 & & & & \vdots & & & \vdots & & & & \\
\Theta(M) & 0 & \cdots & & & & & & & & \cdots & 0 \\
0 & \Theta(M) & \cdots & & & & & & & & \cdots & 0 \\
\vdots & & & & & & & & & & & \vdots \\
0 & 0 & \cdots & & & & & & & & \cdots & \Theta(M)
\end{array}\right] \tag{4.15}$$

where

$$\Theta(i) \triangleq C(m(t_0)+i)\phi(m(t_0)+i, m(t_0)+M). \tag{4.16}$$

That is, at level i, there are 2^i measurements each of which provides information about the *sum* of a block of 2^{M-i} components of X_{M,t_0}. Note that this makes clear that observability is indeed a very strong condition: since successively larger blocks of X_{M,t_0} are summed as we move up the tree, subsequent measurements provide no information about the differences among the values that have been summed. The situation for *reconstructibility*, however, is very different. Specifically, if $W_{M,t_0} = 0$,

$$x(t_0) = \phi(m(t_0), m(t_0)+M)I_M X_{M,t_0} \tag{4.17}$$

$$I_M = \underbrace{[I|I|\cdots|I]}_{2^M \text{times}} \tag{4.18}$$

and each I is an $n \times n$ identity matrix.

Reconstructibility is equivalent to requiring that any vector in the nullspace of (4.14) is also in the nullspace of (4.17).

Since $\phi(m_1, m_2)$ is invertible, this is equivalent to being able to uniquely determine $I_M X_{M,t_0}$, i.e., the sum of the components of X_{M,t_0} from Y_{M,t_0}. We then have Theorem 4.1.

Theorem 4.1: The system (4.1), (4.2) is reconstructible iff $\mathcal{N}(\mathcal{H}_M) \subseteq \mathcal{N}(I_M)$, which is equivalent to the invertibility of the reconstructibility gramian $\mathcal{O}(t_0, M)$:

$$\begin{aligned}\mathcal{O}(t_0, M) &= I_M \mathcal{H}_M^T \mathcal{H}_M I_M^T \\ &= \sum_{i=0}^{M} 2^{2M-i} \phi^T(m(t_0)+i, m(t_0)+M) \\ &\quad \cdot C^T(m(t_0)+i) \\ &\quad \cdot C(m(t_0)+i)\phi(m(t_0)+i, m(t_0)+M).\end{aligned} \tag{4.19}$$

Proof: We must show that $\mathcal{N}(\mathcal{H}_M) \subseteq \mathcal{N}(I_M)$ is equivalent to the invertibility of $\mathcal{O}(t_0, M)$. Suppose first that $\mathcal{O}(t_0, M)$ is not invertible. Then there exists $y \neq 0$ so that $H_M z = 0$ where $z = I_M^T y$. Since I_M^T is one-to-one, $z \neq 0$, which implies that $I_M z = I_M I_M^T y \neq 0$ contradicting $\mathcal{N}(\mathcal{H}_M) \subseteq \mathcal{N}(I_M)$. If, on the other hand, $\mathcal{N}(\mathcal{H}_M)$ is not included in $\mathcal{N}(\mathcal{I}_M)$, choose x such that $\mathcal{H}_M x = 0$ and $I_M x \neq 0$. Since $x \epsilon \mathcal{R}(I_M^T(t_0)) \oplus \mathcal{N}(I_M(t_0))$, we can write $x = I_M^T y + z$ where $y \neq 0$ and $z \epsilon \mathcal{N}(I_M)$. Substituting this into $\mathcal{H}_M x = 0$ and left-multiplying by $I_M \mathcal{H}_M^T$, we get

$$I_M \mathcal{H}_M^T \mathcal{H}_M I_M^T y + I_M \mathcal{H}_M^T \mathcal{H}_M z = 0. \tag{4.20}$$

A straightforward but tedious calculation [9] yields

$$I_M \mathcal{H}_M^T \mathcal{H}_M = \overline{\Lambda} I_M \tag{4.21}$$

where $\overline{\Lambda}$ is an nxn matrix. Equation (4.21) indicates that the column of $\mathcal{I}_M^T$ form a block-eigenspace for $\mathcal{H}_M^T \mathcal{H}_M$. Indeed, as discussed in detail in [9], $\mathcal{H}_M^T \mathcal{H}_M$ is block diagonalized by the (vector) Haar transform, and (4.21) represents the coarsest scale component of that transform. If we now substitute (4.21) into (4.20) and use the fact that $z \epsilon \mathcal{N}(H_M)$, we see that $\Phi(t_0) \mathcal{H}_M^T \mathcal{H}_M \Phi^t(t_0) y = 0$ for some $y \neq 0$, implying that $y^T \Phi(t_0) \mathcal{H}_M^T \mathcal{H}_M \Phi^t(t_0) y = 0$, contradicting the invertibility of $\mathcal{O}(t_0, M)$. □

Also, (4.1), (4.2) is uniformly reconstructible if there exist $\delta, M_0 > 0$ so that

$$\mathcal{O}(t, M_0) \geq \delta I \quad \text{for all } t. \tag{4.22}$$

Note that $\mathcal{O}(t_0, M)$ is the standard observability gramian for the system.

$$x(m) = \frac{1}{2} F(m+1) x(m+1) + \frac{1}{2} G(m+1) u(m+1) \tag{4.23}$$

$$y(m) = \sqrt{2} C(m) x(m) \tag{4.24}$$

Thus if F and C are constant, then (since F is assumed to be invertible) reconstructibility and uniform reconstructibility are equivalent to the usual condition for F and C to be an observable pair.

B. Stability

Next we examine asymptotic stability for the autonomous version of (4.1). Since $x(t)$ is influenced by a geometrically increasing number of nodes at the initial level and ($x(t)$ depends on $\{x(t\alpha), x(t\beta)\}$ or, alternatively on $\{x(t\alpha^2), x(t\beta\alpha), x(t\alpha\beta), x(t\beta^2)\}$, etc., it is necessary to consider an infinite tree, with an infinite set of nodes at each level. Also, we adopt a change of notation to a more standard form by changing the sense of our index of recursion so that m increases as we move up the tree. In particular we arbitrarily choose a level of the tree to be our "initial" level, i.e., level 0, we now index the points on this initial level as $z_i(0)$ for $i \in \mathcal{Z}$. Points at the mth level up from level 0 are denoted $z_i(m)$ for $i \in \mathcal{Z}$. The dynamical equation we then wish to consider is of the form

$$z_i(m) = F(m-1)(z_{2i}(m-1) + z_{2i+1}(m-1)). \tag{4.25}$$

Let $Z(m)$ denote that set $\{z_i(m), i \in \mathcal{Z}\}$, with p-norm inherited from the p-norms of its components:

$$\|Z(m)\|_p \triangleq \left(\sum_i \|z_i(m)\|_p^p \right)^{1/p}. \tag{4.26}$$

Definition 4.1: A system is l_p-exponentially stable if there exists $0 \leq \alpha < 1$ and $C > 0$ so that given any initial sequence $Z(0)$ such that $\|Z(0)\|_p < \infty$

$$\|Z(m)\|_p \leq C\alpha^m \|Z(0)\|_p. \tag{4.27}$$

From (4.25) we can immediately write the following.

$$z_i(m) = \Phi(m, 0) \sum_{j \in O_{m,i}} z_j(0) \tag{4.28}$$

where the cardinality of $O_{m,i}$ is 2^m and $\Phi(m, 0)$ is the transition matrix for $F(m)$.

Theorem 4.2: The system (4.25) is l_p-exponentially stable if and only if

$$2^{m/q} \|\Phi(m, 0)\|_p \leq K' \gamma^m \quad \text{for all } m \tag{4.29}$$

where $0 \leq \gamma < 1$ and K' is a positive constant, and

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{4.30}$$

Proof: Let us first show necessity. Specifically, suppose that for any $K > 0, 0 \leq \gamma < 1$, and $M \geq 0$ we can find a vector z and an $m \geq M$ so that

$$\|\Phi(m, 0) z\|_p > K \gamma^m 2^{-m/q} \|z\|_p. \tag{4.31}$$

Let z and m be such a vector and integer for some choice of K, γ, and M, and define an initial sequence as follows. Let $\rho_0, \rho_1, \rho_2, \cdots$ be a sequence with

$$\sum_{i=0}^{\infty} \rho_i^p = 1. \tag{4.32}$$

Then let

$$z_i(0) = \rho_j z, \qquad j2^m \leq i < (j+1)2^m, \quad j = 0, 1, \cdots. \tag{4.33}$$

Note that

$$\|Z(0)\|_p^p = 2^m \|z\|_p^p. \tag{4.34}$$

Thus, using (4.28), (4.31)–(4.34)

$$\begin{aligned}\|Z(m)\|_p^p &= 2^{mp}\|\Phi(m,\, 0)z\|_p^p \\ &> 2^{mp}K^p\gamma^{mp}2^{-mp/q}\|z\|_p^p \\ &= 2^{mp}K^p\gamma^{mp}2^{-mp/q}2^{-m}\|Z(0)\|_p^p \\ &= K^p\gamma^{mp}\|Z(0)\|_p^p. \end{aligned} \tag{4.35}$$

Hence for any $K, 0 \leq \gamma < 1$ and $M \geq 0$ we can find $Z(0)$ and $m \geq M$ so that

$$\|Z(m)\|_p > K\gamma^m \|Z(0)\|_p \tag{4.36}$$

so that the system cannot be l_p-exponentially stable.

To prove sufficiency we use two simple facts. First, (4.25) is exponentially stable if there exist $0 \leq \beta < 1$ and $K > 0$ so that for each i

$$\|z_i(m)\|_p \leq K\beta^m \left(\sum_{j \in O_{m,\, i}} \|z_j(0)\|_p^p \right)^{1/p}. \tag{4.37}$$

This follows by raising (4.37) to the pth power and summing over i. Secondly, for any sequence of vectors x_i and any m and j

$$\left\| \sum_{i \in \mathcal{I}_{m,\, j}} x_i \right\|_p \leq 2^{m/q} \left(\sum_{i \in \mathcal{I}_{m,\, j}} \|x_i\|_p^p \right)^{1/p} \tag{4.38}$$

where $\mathcal{I}_{m,\, j} = \{j, j+1, \cdots, j+2^m-1\}$. To show this, we use the fact

$$\|a+b\|_p \leq 2^{1/q}(\|a\|_p^p + \|b\|_p^p)^{1/p} \tag{4.39}$$

together with induction on m. Note first that (4.38) is trivially true for $m = 0$. Suppose then that for all j (4.38) holds for a particular value of m. If we then sum x_i over the two sets $I_{m,\, j_1}$ and $I_{m,\, j_2}$ where $j_2 = j_1 + 2^m$ we get

$$\begin{aligned}&\left\| \left(\sum_{i \in \mathcal{I}_{m,\, j_1}} x_i + \sum_{i \in \mathcal{I}_{m,\, j_2}} x_i \right) \right\|_p \\ &\leq 2^{1/q} \left(\left\| \left(\sum_{i \in \mathcal{I}_{m,\, j_1}} x_i \right) \right\|_p^p + \left\| \left(\sum_{i \in \mathcal{I}_{m,\, j_2}} x_i \right) \right\|_p^p \right)^{1/p}. \end{aligned} \tag{4.40}$$

Then by substituting into (4.38) into (4.40) we get

$$\begin{aligned}&\left\| \sum_{i \in \mathcal{I}_{m,\, j_1} \cup \mathcal{I}_{m,\, j_2}} x_i \right\|_p \\ &\leq 2^{(m+1)/q} \left(\left\| \left(\sum_{i \in \mathcal{I}_{m,\, j_1}} x_i \right) \right\|_p^p + \left\| \left(\sum_{i \in \mathcal{I}_{m,\, j_2}} x_i \right) \right\|_p^p \right)^{1/p} \end{aligned} \tag{4.41}$$

and applying (4.39), we find that (4.38) holds for $m+1$ as well.

By taking the p-norm of (4.28), using Cauchy–Schwarz and (4.38), we obtain

$$\|z_i(m)\|_p \leq \|\Phi(m,\, 0)\|_p 2^{m/q} \left(\sum_{j \in \mathcal{I}_{m,\, i}} \|z_j(0)\|_p^p \right)^{1/p}. \tag{4.42}$$

If we then assume that (4.29) holds this, together with (4.42) yields

$$\|z_i(m)\|_p \leq K'\gamma^m \left(\sum_{j \in \mathcal{I}_{m,\, i}} \|z_j(0)\|_p^p \right)^{1/p} \tag{4.43}$$

from which we conclude that the system is l_p-exponentially stable. □

Note that from this result we see that the l_p-exponential stability of (4.25) is equivalent to the usual exponential stability of the system

$$\xi(m) = 2^{1/p}F(m-1)\xi(m-1). \tag{4.44}$$

For example for $p = 2$, and F is constant, this reduces to all eigenvalues of F having magnitude less than $(\sqrt{2}/2)$.

V. Bounds, Stability, and Steady-State Behavior

In this section we develop several system-theoretic results for our fine-to-coarse filtering algorithm, paralleling those for standard Kalman filtering, but with several key differences due to the structure of the dyadic tree. We focus in this section on the scale-varying case, i.e., the case in which all system parameters vary with scale only. In this case straightforward analysis of the filtering algorithm of Section II verifies that the fine-to-coarse Kalman filter parameters also depend only on scale, i.e. $K(t) = K(m(t))$, $P(t|t+) = P(m(t)|m(t)+)$, etc., resulting in the filter

$$\hat{x}(t|t) = \hat{x}(t|t+) + K(m(t))[y(t) - C(m(t))\hat{x}(t|t+)] \tag{5.1}$$

$$\hat{x}(t\bar{\gamma}t) = F(m(t))\hat{x}(t|t) \tag{5.2}$$

$$\begin{aligned}\hat{x}(t|t+) = P(m(t)|m(t)+)&P^{-1}(m(t)|m(t)+1) \\ &\cdot[\hat{x}(t|t\alpha) + \hat{x}(t|t\beta)] \end{aligned} \tag{5.3}$$

with the scale-varying Riccati equation

$$\begin{aligned}P(m|m+1) &= F(m+1)P(m+1|m+1)F^T(m+1) \\ &\quad + G(m+1)\tilde{Q}(m+1)G^T(m+1) \end{aligned} \tag{5.4}$$

$$\begin{aligned}P^{-1}(m|m) &= 2P^{-1}(m|m+1) \\ &\quad + C^T(m)R^{-1}(m)C(m) - P_x^{-1}(m) \end{aligned} \tag{5.5}$$

where we have combined the update and fusion steps in (5.5). Also $F(m(t))$ and $\tilde{Q}(m(t))$ are given by (2.7), (2.9) in the scale-varying case and

$$G(m) = A^{-1}(m)B(m). \tag{5.6}$$

Furthermore, the remaining quantities needed in (5.1)–(5.2) are

$$P^{-1}(m|m+) = 2P^{-1}(m|m+1) - P_x^{-1}(m) \tag{5.7}$$

$$K(m) = P(m|m)C^T(m)R^{-1}(m). \tag{5.8}$$

In the ML case, with P_x^{-1} set to zero we obtain a further simplification:

$$\begin{aligned}\hat{x}_{ML}(t|t) = &\frac{1}{2}(I - K_{ML}(m(t))C(m(t))A^{-1}(m(t)+1)\\ &\cdot(\hat{x}_{ML}(t\alpha|t\alpha) + \hat{x}_{ML}(t\beta|t\beta))\\ &+ K_{ML}(m(t))y(t).\end{aligned} \tag{5.9}$$

Similarly we have the following simplified form of (3.17) for the ML filter error:

$$\begin{aligned}\tilde{x}_{ML}(t|t) = &\frac{1}{2}(I - K_{ML}(m(t))C(m(t)))A^{-1}(m(t)+1)\\ &\cdot(\tilde{x}_{ML}(t\alpha|t\alpha) + \tilde{x}_{ML}(t\beta|t\beta))\\ &- \frac{1}{2}(I - K_{ML}(m(t))C(m(t)))G(m(t)+1)\\ &\cdot(w(t\alpha) + w(t\beta)) - K_{ML}(m(t))v(t).\end{aligned} \tag{5.10}$$

The ML Riccati equation in this case becomes

$$\begin{aligned}P_{ML}(m|m+1) = &A^{-1}(m+1)P(m+1|m+1)\\ &\cdot A^{-T}(m+1) + G(m+1)G^T(m+1)\end{aligned} \tag{5.11}$$

$$P_{ML}^{-1}(m|m) = 2P_{ML}^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m). \tag{5.12}$$

Also

$$K_{ML}(m) = P_{ML}(m|m)C^T(m)R^{-1}(m) \tag{5.13}$$

$$P_{ML}(m|m+) = \frac{1}{2}P_{ML}(m|m+1) \tag{5.14}$$

and (5.13), (5.14) together yield

$$\frac{1}{2}[I - K_{ML}(m)C(m)] = P_{ML}(m|m)P_{ML}^{-1}(m|m+1). \tag{5.15}$$

A. Bounds on the Error Covariance

As is the case for standard Kalman filtering, [3], [8], reachability and reconstructibility conditions are key in deriving upper and lower bounds on the error covariances $P(m|m)$ and $P_{ML}(m|m)$. The system to be analyzed is the following backward model, obtained directly from (2.6)–(2.9) in the scale-varying case:

$$\begin{aligned}x(t) = &\frac{1}{2}F(m(t)+1)[x(t\alpha) + x(t\beta)]\\ &+\frac{1}{2}G(m(t)+1)[\tilde{w}(t\alpha) + \tilde{w}(t\beta)]\end{aligned} \tag{5.16}$$

together with the measurements (2.2). To begin, we define the gramians:

$$\begin{aligned}\overline{\mathcal{R}}(t, M) \triangleq &\sum_{i=0}^{M-1} 2^{-i-1}\phi(m(t), m(t)+i)G(m(t)+i+1)\\ &\cdot\tilde{Q}(m(t)+i+1)G^T(m(t)+i+1)\\ &\cdot\phi^T(m(t), m(t)+i)\end{aligned} \tag{5.17}$$

$$\begin{aligned}\overline{\mathcal{O}}(t, M) \triangleq &\sum_{i=0}^{M} 2^i\phi^T(m(t)+i, m(t)+M)C^T(m(t)+i)\\ &\cdot R^{-1}(m(t)+i)C(m(t)+i)\\ &\cdot\phi(m(t)+i, m(t)+M)\end{aligned} \tag{5.18}$$

where the state transition matrix is given by (4.8)–(4.9). We also assume that $A(m), A^{-1}(m), B(m), P_x^{-1}(m), C(m), R(m)$, and $R^{-1}(m)$ are bounded functions of m, implying that for any $M_0 > 0$ we can find $\alpha, \beta > 0$ so that

$$\overline{\mathcal{R}}(t, M_0) \le \alpha I \quad \text{for all } t \tag{5.19}$$

$$\overline{\mathcal{O}}(t, M_0) \le \beta I \quad \text{for all } t. \tag{5.20}$$

Also uniform reachability corresponds to the existence of $\gamma, M_0 > 0$ so that

$$\overline{\mathcal{R}}(t, M_0) \ge \gamma I \quad \text{for all } t \tag{5.21}$$

while uniform reconstructibility corresponds to the existence of $\delta, M_0 > 0$ so that

$$\overline{\mathcal{O}}(t, M_0) \ge \delta I \quad \text{for all } t. \tag{5.22}$$

These conditions coincide with those in Section IV-A with the replacement of $F(m)$ by $(1/2)F(m)$, $G(m)$ by $(1/2)G(m)\tilde{Q}^{1/2}(m)$, and $C(m)$ by $R^{-1/2}(m)C(m)$. To derive an upper bound for the optimal filter error covariance, the key is to make a comparison between the Riccati equation for our optimal filter and the Riccati equation for the standard Kalman filters.

Lemma 5.1: Let $P(m|m)$ be the solution to the Riccati equation (5.4)–(5.5), and let $\overline{P}(m|m)$ satisfy the second Riccati equation

$$\begin{aligned}\overline{P}(m|m+1) = &F(m+1)\overline{P}(m+1|m+1)F^T(m+1)\\ &+G(m+1)\tilde{Q}(m+1)G^T(m+1)\end{aligned} \tag{5.23}$$

$$\overline{P}^{-1}(m|m) = \overline{P}^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m). \tag{5.24}$$

Then

$$\overline{P}^{-1}(m|m) \le P^{-1}(m|m). \tag{5.25}$$

Proof: First note that (5.5) can be rewritten as

$$\begin{aligned}P^{-1}(m|m) = &P^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m)\\ &+D^T(m)D(m)\end{aligned} \tag{5.26}$$

since $P(m|m+1) \le P_x(m)$. Also, (5.23) and (5.24) characterize the error covariance for the optimal filter corresponding to the following standard filtering problem.

$$x(m) = F(m+1)x(m+1) + G(m+1)w(m+1) \tag{5.27}$$

$$y(m) = C(m)x(m) + v(m) \tag{5.28}$$

where $w(m)$ and $v(m)$ have covariances $\tilde{Q}(m)$ and $R(m)$, respectively. Equation (5.25) then follows by observing that

(5.26) characterizes the error covariance for the filtering problem involving the same state equation but with augmented measurements

$$\tilde{y}(m) = \begin{bmatrix} C(m) \\ D(m) \end{bmatrix} x(m) + u(m) \quad (5.29)$$

$$E[u(m)u^T(m)] = \begin{bmatrix} R(m) & 0 \\ 0 & I \end{bmatrix}. \quad (5.30)$$

□

Theorem 5.1: Suppose there exists β, δ, $M_0 > 0$ so that (5.20) and (5.22) are satisfied. Then there exists $\kappa > 0$ such that for all m at least M_0 levels from the initial level $P(m|m) \leq \kappa I$.

Proof: As we have discussed, (5.20) and (5.22) are equivalent to the existence of analogous uniform upper and lower bounds on the observability gramian for (5.27). Thus standard Kalman filtering results imply that there exists a $\kappa > 0$ such that $\overline{P}(m|m) \leq \kappa I$ or $\overline{P}^{-1}(m|m) \geq \kappa^{-1} I$. Lemma 5.1 then yields the desired result. □

We can easily apply the previous ideas to derive an upper bound for $P_{ML}(m|m)$ as well: Specifically note that the identical idea used in Lemma 5.1 yields an analogous result for the ML Riccati equation (5.11) and (5.12), i.e.,

$$\hat{P}^{-1}(m|m) \leq P_{ML}^{-1}(m|m) \quad (5.31)$$

where $\hat{P}(m|m)$ is the solution of a Riccati equation as in (5.23) and (5.24), but with F and $\tilde{Q}$ replaced by A^{-1} and I, respectively. Since (5.20) and (5.22) are equivalent to analogous conditions on the usual observability gramian for the pair $(R^{-1/2}(m)C(m), A^{-1}(m))$, we obtain an upper bound on $\hat{P}(m|m)$, which, with (5.31), yields the following theorem.

Theorem 5.2: Suppose that there exists β, δ, $M_0 > 0$ so that (5.20) and (5.22) are satisfied. Then there exists $\kappa' > 0$ such that for all m at least M_0 levels from the initial level $P_{ML}(m|m) \leq \kappa' I$.

We now turn to the lower bound for $P(m|m)$. We begin with the following lemma.

Lemma 5.2: Let

$$\overline{S}(m|m) \triangleq \frac{1}{2}(P^{-1}(m|m) - C^T(m)R^{-1}(m) \cdot C(m) + P_x^{-1}(m)) \quad (5.32)$$

$$\overline{S}(m|m+1) \triangleq F^{-T}(m+1)P^{-1}(m+1|m+1) \cdot F^{-1}(m+1). \quad (5.33)$$

Consider also the Riccati equation

$$S^*(m|m+1) = 2F^{-T}(m+1)S^*(m+1|m+1) \cdot F^{-1}(m+1) + F^{-T}(m+1)C^T(m) \cdot R^{-1}(m)C(m)F^{-1}(m+1) \quad (5.34)$$

$$S^{*-1}(m|m) = S^{*-1}(m|m+1) + G(m+1)\tilde{Q}(m+1)G^T(m+1) \quad (5.35)$$

where $\overline{S}(0|0) = S^*(0|0)$. Then for all m, $S^*(m|m) \geq \overline{S}(m|m)$.

Proof: Straightforward calculations using (5.4), (5.5), (5.32), and (5.33) yield

$$\begin{aligned} \overline{S}(m|m) &= P^{-1}(m|m+1) \\ &= [\overline{S}^{-1}(m|m+1) + G(m+1) \\ &\quad \cdot \tilde{Q}(m+1)G^T(m+1)]^{-1}. \end{aligned} \quad (5.36)$$

Also, by substituting (5.32) into (5.33) and collecting terms we obtain

$$\begin{aligned} \overline{S}(m|m+1) = 2F^{-T}(m+1)\overline{S}(m+1|m+1) \\ \cdot F^{-1}(m+1) + F^{-T}(m+1)C^T(m) \\ \cdot R^{-1}(m)C(m)F^{-1}(m+1) \\ - F^{-T}(m+1)P_x^{-1}(m)F^{-1}(m+1). \end{aligned} \quad (5.37)$$

Now we prove by induction that for all $mS^*(m|m) \geq \overline{S}(m|m)$. Obviously, $S^*(0|0) \geq \overline{S}(0|0)$. As an induction hypothesis we assume $S^*(i+1|i+1) \geq \overline{S}(i+1|i+1)$. From (5.37), (5.34), and the fact that $F^{-T}(m+1)P_x^{-1}(m)F^{-1}(m+1) \geq 0$, we get that

$$S^{*-1}(i|i+1) \leq \overline{S}^{-1}(i|i). \quad (5.38)$$

Combining (5.35), (5.36) and (5.38) yields $S^*(i|i) \geq \overline{S}(i|i)$. □

Theorem 5.3: Suppose that there exists α, γ, $M_0 > 0$ so that (5.19) and (5.21) are satisfied. Then there exists $L > 0$ such that for all m at least M_0 levels from the initial level $P(m|m) \geq LI$.

Proof: From standard Kalman filtering results we know that the solution to the standard Riccati equation (5.34), (5.35) satisfies $S^*(m|m) \leq NI$. For some $N > 0$ if the pair $(\tilde{Q}^{T/2}(m)G^T(m), F^{-T}(m))$ is bounded and uniformly observable. By standard duality results and the boundedness of F, however, this is equivalent to the boundedness and uniform reachability of $(F(m), G(m)\tilde{Q}^{1/2}(m))$, which in turn are equivalent to (5.19) and (5.21). Then from Lemma 5.2 we conclude that $\overline{S}(m|m) \leq NI$, and (5.32) together with the boundedness assumption yields the desired result. □

Using analogous arguments we can derive a lower bound for $P_{ML}(m|m)$. Note that with the following definitions for $\overline{S}$ and (5.34), (5.35) where the matrices F and $\tilde{Q}$ are replaced with the matrices A^{-1} and I, respectively, Lemma 5.2 still applies.

$$\overline{S}(m|m) \triangleq \frac{1}{2}(P_{ML}^{-1}(m|m) - C^T(m)R^{-1}(m)C(m)) \quad (5.39)$$

$$\overline{S}(m|m+1) \triangleq A^T(m+1)P_{ML}^{-1}(m+1|m+1)A(m+1) \quad (5.40)$$

Using the same argument as in the proof of Theorem 5.3 we find that

$$\frac{1}{2}(P_{ML}^{-1}(m|m) - C^T(m)R^{-1}(m)C(m)) \leq NI \quad (5.41)$$

for $N > 0$, and the boundedness assumption then yields

Theorem 5.4: Suppose that there exists α, γ, $M_0 > 0$ so that (5.19) and (5.21) are satisfied. Then there exists $L' > 0$ such that for all m $P_{ML}(m|m) \geq L'I$.

B. Filter Stability

We first analyze the ML filter error dynamics (5.10). Using (5.15), we examine the asymptotic stability of the autonomous error dynamics

$$\xi(t) = P_{ML}(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t)+1) \cdot A^{-1}(m(t)+1)[\xi(t\alpha)+\xi(t\beta)]. \quad (5.42)$$

Theorem 5.5: Suppose that (5.19)–(5.22) are satisfied. Then, the ML error dynamics (5.10), or equivalently (5.42) are l_2-exponentially stable.

Proof: Based on Section IV-B, we wish to show that the following system is stable:

$$z(m) = P_{ML}(m|m)P_{ML}^{-1}(m|m+-1) \cdot \sqrt{2}A^{-1}(m+1)z(m+1). \quad (5.43)$$

The analysis follows the line of reasoning used in [3]. Specifically, thanks to Theorem 5.2 and 5.4, we can define the following Lyapunov function

$$V(z(m), m) \triangleq z^T(m)P_{ML}^{-1}(m|m)z(m). \quad (5.44)$$

Let us also define the following quantity.

$$\tilde{z}(m) \triangleq \sqrt{2}A^{-1}(m+1)z(m+1). \quad (5.45)$$

Substituting (5.12) into (5.44), using (5.45), and performing some algebra (see [9]) yields

$$\begin{aligned} V(z(m), m) &- V(z(m+1), m+1) \\ &\le -\left(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}}\right)^T P_{ML}^{-1}(m|m+1) \\ &\quad \cdot \left(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}}\right) \\ &\quad - z^T(m)C^T(m)R^{-1}(m)C(m)z(m). \end{aligned} \quad (5.46)$$

Stability follows from (5.46) since $(R^{-1/2}(m)C(m), A^{-1}(m))$ is uniformly observable. □

The full estimation error, after incorporating prior statistics, is given by

$$\tilde{x}(t|t) = P(m(t)|m(t))(P_{ML}^{-1}(m(t)|m(t))\tilde{x}_{ML}(t|t) + P_x^{-1}(m(t))x(t)). \quad (5.47)$$

Thus $\tilde{x}(t|t)$ is a linear combination of the states of two upward-evolving systems, one for $\tilde{x}_{ML}(t|t)$ and one for $P_x^{-1}(m(t))x(t)$. Note that since $P(m|m) \le P_{ML}(m|m)$

$$\|P(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t))\tilde{x}_{ML}(t|t)\| \le \|\tilde{x}_{ML}(t|t)\| \quad (5.48)$$

and we already have the stability of the $\tilde{x}_{ML}(t|t)$ dynamics from Theorem 5.5. Next, note that the covariance of $P_x^{-1}(m(t))x(t)$ is simply $P_x^{-1}(m(t))$. By uniform reachability $P_x^{-1}(m(t))$ is bounded above. Thus, since $P(m(t)|m(t))$ is bounded, the contribution to the error of the second term in (5.47) is bounded.

C. Steady-State Filter

In this section we focus on the constant parameter case and analyze the asymptotic properties of the filter. Specifically, we have the following theorem.

Theorem 5.6: Consider the following system defined on a tree.

$$x(t) = Ax(t\bar{\gamma}) + Bw(t) \quad (5.49)$$

$$y(t) = Cx(t) + v(t) \quad (5.50)$$

with independent white noises w and v having covariances I and R, respectively. Suppose that (A, B) is a reachable pair and that (C, A) is observable. Then, the error covariance for the ML estimator, $P_{ML}(m|m)$, converges as $m \to -\infty$ to $\overline{P}_\infty$, which is the unique positive definite solution to

$$\begin{aligned} \overline{P}_\infty &= \frac{1}{2}A^{-1}\overline{P}_\infty A^{-T} + \frac{1}{2}GG^T \\ &- K_\infty\left(\frac{1}{2}CA^{-1}\overline{P}_\infty A^{-T}C^T + \frac{1}{2}CGG^TC^T + R\right)K_\infty^T \end{aligned} \quad (5.51)$$

where

$$K_\infty = \overline{P}_\infty C^T R^{-1}. \quad (5.52)$$

Moreover, the autonomous dynamics of the steady-state ML filter, i.e,

$$e(t) = \frac{1}{2}(I - K_\infty C)A^{-1}(e(t\alpha) + e(t\beta)) \quad (5.53)$$

are l_2-stable, i.e., $(1/2)(I - K_\infty C)A^{-1}$ has eigenvalues less than $\sqrt{2}/2$ in magnitude.

Proof: The convergence of $P_{ML}(m|m)$ will be established if we can show that 1) $P_{ML}(m|m)$ is monotone-nonincreasing as $m \to -\infty$ and 2) $P_{ML}(m|m)$ is bounded below. The second of these conditions comes directly from the assumptions of reachability and observability. The monotonicity of $P_{ML}(m|m)$ follows from an argument analogous to that used in the standard case (see [9]. Let $\overline{P}_\infty$ denote the limit. It is straightforward to see that $\overline{P}_\infty$ must satisfy (5.51), which is the steady-state version of the constant-coefficient ML Riccati equation (5.11), (5.12). Furthermore, by Theorem 5.4, $\overline{P}_\infty$ must be positive definite.

We next show that if $\overline{P}_\infty$ is any positive definite solution to (5.51), then each eigenvalue of $(\sqrt{2}/2)(I - K_\infty C)A^{-1}$ has magnitude less than one, where K_∞ is given by (5.52). The approach is a variation of the proof for the standard Riccati equations [8]. Specifically, suppose that there exists an eigenvalue with $|\lambda| \ge 1$. Then letting x be the associated eigenvector of $[(\sqrt{2}/2)(I - K_\infty C)A^{-1}]^T$, some algebra using (5.51) yields

$$x^H\overline{P}_\infty x = |\lambda|^2 x^H P_\infty x + |\lambda|^2 x^H BB^T x + x^H K_\infty R K_\infty^T x. \quad (5.54)$$

Since $\overline{P}_\infty > 0$ and $|\lambda| > 1$, we conclude from (5.54) that $x^H B = 0$ and $x^H K_\infty = 0$, the latter of which implies $x^H A^{-1} = \sqrt{2}\lambda^H x$. These in turn imply that (A^{-1}, B) is not a reachable pair which contradicts the assumption that (A, B) is reachable.

Finally, suppose P_1 and P_2 are both positive definite and satisfy (5.51) and (5.52). Using (5.51) and (5.52) for both P_1

and P_2 then yields

$$P_1 - P_2 = \frac{\sqrt{2}}{2}(I - K_1C)A^{-1}(P_1 - P_2) \cdot \left(\frac{\sqrt{2}}{2}(I - K_1C)A^{-1}\right)^T + \Delta \quad (5.55)$$

$$\Delta = (K_1 - K_2)\left[\frac{1}{2}CA^{-1}P_2A^{-T}C^T + \frac{1}{2}CGG^TC^T + R\right](K_1 - K_2)^T > 0. \quad (5.56)$$

Since $(\sqrt{2}/2)(I - K_1C)A^{-1}$ has eigenvalues within the unit circle, standard system theory yields $P_1 - P_2 \geq 0$. Reversing indices yields $P_2 - P_1 \geq 0$, proving uniqueness. □

Let us comment on the asymptotic behavior of the Bayesian error covariance $P(m|m)$, which is given by

$$P(m|m) = [P_{ML}^{-1}(m|m) + P_x^{-1}(m)]^{-1}. \quad (5.57)$$

Since the original state process is defined evolving from coarse-to-fine while the recursion of the ML filter is in the opposite direction, we need to be a bit careful about defining exactly what we mean by the asymptotic behavior of (5.57). Specifically, what we mean here is its asymptotic behavior at a finite value of m as both the bottom *and* top levels of the tree recede. Note that while the convergence of $P_x(m)$ depends upon the stability of A, the convergence of $P_x^{-1}(m)$ does not. Specifically, since (A, B) is reachable, it is easily seen (e.g., by examining the Riccati equation for $P_x^{-1}(m)$ obtained from (2.5)) that $P_x^{-1}(m)$ does converge as m increases.[2] Thus, if we let S_x denote that limiting value, then $P(m|m)$ converges to $[\overline{P}_{\infty}^{-1} + S_x]^{-1}$.

VI. Conclusion

In this paper we have analyzed in detail the new class of multiscale filtering and smoothing algorithms developed in [1], based on dynamic models defined on dyadic trees in which each level in the tree corresponds to a different resolution of signal representation. In particular, this framework leads to an extremely efficient and highly parallelizable scale-recursive optimal estimation algorithm generalizing the Rauch–Tung–Striebel smoothing algorithm to the dyadic tree. This algorithm involves a variation on the Kalman filtering algorithm in that, in addition to the usual measurement update and (fine-to-coarse) prediction steps, there is also a data fusion step. This in turn leads to a new Riccati equation. As we have seen, the presence of the data fusion step leads to a complication in filter and Riccati equation analysis, and this motivated the derivation in this paper of an alternative ML algorithm which leads in turn to a variation on the RTS procedure corresponding to the triangularization of the Hamiltonian description of the optimal smoother.

This paper focuses on the development of system-theoretic concepts of reachability, reconstructibility, and stability for fine-to-coarse dynamic models which we then used to analyze the asymptotic stability of the multiscale Kalman filter error dynamics and the steady-state convergence of the Riccati equation in the constant parameter case. As we have seen, the structure of the dyadic tree leads to differences in these system-theoretic concepts and results as compared to their counterparts for standard state-space models.

As we discuss in [1], multiresolution methods of signal and image analysis are of considerable interest in research and in numerous applications. One of our objectives in [1], the present paper, and our paper on multiresolution realization theory [2] is to demonstrate that there is a substantial role for the systems and control community in this field. Indeed it is our opinion that there are a broad range of opportunities for further work in both theory and application (including, for example, exploring the relationship between the methods and framework described here and well-known singular and regular perturbation methods of multiple time scale analysis), and it is our hope that our work will help to stimulate activity in this fascinating and important area.

Appendix A

To verify that the ML estimate for $x(t)$ given Y_t does indeed satisfy (3.1)–(3.8), note first that $\hat{x}_{ML}(t\overline{\gamma}|t)$ is the ML estimate based on Y_t together with one additional "measurement," namely the dynamical relation (2.1) between $x(t)$ and $x(t\overline{\gamma})$. Using results on recursive ML estimation [4], $\hat{x}_{ML}(t\overline{\gamma}|t)$ is, equivalently, the ML estimate of $x(t\overline{\gamma})$ given the "measurement"

$$\hat{x}_{ML}(t|t) = A(t)x(t\overline{\gamma}) + w(t) + \tilde{x}_{ML}(t|t) \quad (A.1)$$

where the estimation error $\tilde{x}_{ML}(t|t)$ is zero-mean, independent of $w(t)$, and with covariance $P_{ML}(t|t)$. Equations (3.5)–(3.6) follow directly from this. The fusion step [(3.7), (3.8)] then follows from standard results [4] on the fusion of ML estimates based on disjoint data sets with independent noises, since $\hat{x}_{ML}(t|t\alpha)$ is the ML estimate based on $Y_{t\alpha}$ together with (2.1) evaluated at $t\alpha$, while $\hat{x}_{ML}(t|t\beta)$ is based on $Y_{t\beta}$ and (2.1) evaluated at $t\beta$. The update step (3.1)–(3.4 follows from the standard result on incorporating a new, independent measurement.

Also, straightforward calculations using (3.1)–(3.10) lead to an information filter version of the ML algorithm. Specifically, let S denote the inverse covariance and z the state of the information filter, i.e.,

$$S(t|t) = P_{ML}^{-1}(t|t), \qquad S(t|t+) = P_{ML}^{-1}(t|t+), \text{ etc.} \quad (A.2)$$

$$\hat{z}(t|t) = S(t|t)\hat{x}_{ML}(t|t),$$

$$\hat{z}(t|t+) = S(t|t+)\hat{x}_{ML}(t|t+), \text{ etc..} \quad (A.3)$$

Then we have the following algorithm

$$\hat{z}(t|t) = \hat{z}(t|t+) + C^T(t)R^{-1}(t)y(t) \quad (A.4)$$

$$\hat{z}(t\overline{\gamma}|t) = J^T(t)\hat{z}(t|t) \quad (A.5)$$

[2] The two extreme cases are: A stable, so that $P_x^{-1}(m) \to P_x^{-1}$ where P_x is the positive-definite solution of the algebra Lyapunov equation, and A^{-1} stable, so that $P_x^{-1}(m) \to 0$.

$$\hat{z}(t|t+) = \hat{z}(t|t\alpha) + \hat{z}(t|t\beta) \tag{A.6}$$

$$S(t|t) = S(t|t+) + C^T(t)R^{-1}(t)C(t) \tag{A.7}$$

$$J(t) = \{I - B(t)[B^T(t)S(t|t)B(t) + I]^{-1} \cdot B^T(t)S(t|t)\}A(t) \tag{A.8}$$

$$S(t\bar{\gamma}|t) = J^T(t)S(t|t)A(t) \tag{A.9}$$

$$S(t|t+) = S(t|t\alpha) + S(t|t\beta). \tag{A.10}$$

Note the simple form of the fusion step (A.5), (A.10), emphasizing that the independent sets of information are being combined. Also this algorithm is well defined when S is singular, i.e., when insufficient information has been collected for x to be estimable. In particular, initialization of the algorithm is given by

$$\hat{z}(t|t+) = 0, \qquad S(t|t+) = 0$$

for all t such that $m(t) = M$. (A.11)

In addition, further algebra yields the corresponding smoothing step

$$\hat{x}_x(t) = J(t)\hat{x}_s(t\bar{\gamma}) + J(t)A^{-1}(t)B(t)B^T(t)\hat{z}(t|t) \tag{A.12}$$

$$P_s(t) = J(t)P_s(t\bar{\gamma})J^T(t) + J(t)A^{-1}(t)B(t)B^T(t)J^T(t) \tag{A.13}$$

where this is initialized at the top of the tree with

$$\hat{x}_s(0) = P_s(0)\hat{z}(0|0) \tag{A.14}$$

$$P_s(0) = [S(0|0) + P_x^{-1}(0)]^{-1}. \tag{A.15}$$

APPENDIX B

In this appendix we introduce and analyze the Hamiltonian form of the smoothing equations on the tree. For simplicity we focus on the constant parameter case. The extension to the general case is straightforward. Specifically, consider the model (2.1), (2.2) defined on an M-level tree with a single root node 0, with A, B, C constant and where w and v are white-noise processes with variances I and R respectively. The Hamiltonian form of the smoothing equations can be derived either using the complementary model construction (e.g., as in [6]) or, as we do here, by computing the $x(t)$-trajectory that has maximum posterior probability. Specifically, with $x(0)$ having prior mean of 0 and prior covariance of $P_x(0)$, by straightforward adaptation of standard results we find that the optimal smoothed estimate $\hat{x}_s(t)$ is obtained by minimizing

$$\begin{aligned} H = &\sum_t \frac{1}{2}(y(t) - Cx(t))^T R^{-1}(y(t) - Cx(t)) \\ &+ \sum_{t\neq 0} \frac{1}{2} w^T(t)w(t) \\ &+ \frac{1}{2}x^T(0)P_x^{-1}(0)x(0) \\ &+ \sum_{t\neq 0} \lambda^T(t)(x(t) - Ax(t\bar{\gamma}) - Bw(t)) \end{aligned} \tag{B.1}$$

with respect to the state x, the noise w, and the Lagrange multiplier λ^T.

As in the standard case, after we set to zero the derivatives of H with respect to x, w, and λ, we find that we can eliminate w by expressing it as a function of λ, yielding the following optimal smoothing equations for $m(t) = 1, \cdots, M$:

$$\hat{\lambda}(t) = A^T[\hat{\lambda}(t\alpha) + \hat{\lambda}(t\beta)] - C^T R^{-1} C\hat{x}_s(t) + C^T R^{-1} y(t) \tag{B.2}$$

$$\hat{x}_s(t) = A\hat{x}_s(t\bar{\gamma}) + BB^T\hat{\lambda}(t) \tag{B.3}$$

and the boundary conditions[3]

$$\hat{x}_s(0) = [P_x(0) + C^T R^{-1} C]^{-1} \cdot\{A^T[\hat{\lambda}(0\alpha) + \hat{\lambda}(0\beta)] + C^T R^{-1} y(0)\} \tag{B.4}$$

$$\hat{\lambda}(t) = 0, \qquad m(t) = M + 1. \tag{B.5}$$

Note that, as in the standard case, the dual dynamics for $\tilde{\lambda}$ run in the opposite direction to the x-dynamics. In this case, thanks to the asymmetry of the tree, the dual dynamics (B.2) are in the form of fine-to-coarse dynamics which merge values as we progress up the tree. Also, by organizing the dynamics (B.2), (B.3) we can obtain the Hamiltonian form of the dynamics for $m(t) = 1, \cdots, M$;

$$\mathcal{A}\begin{bmatrix}\hat{x}_s \\ \hat{\lambda}\end{bmatrix}_\alpha + \Theta_\alpha \begin{bmatrix}\hat{x}_s \\ \hat{\lambda}\end{bmatrix}_{t\alpha} + \Theta_\beta \begin{bmatrix}\hat{x}_s \\ \hat{\lambda}\end{bmatrix}_{t\beta} = \begin{bmatrix} 0 \\ 0 \\ C^T R^{-1} y(t)\end{bmatrix} \tag{B.6}$$

where A, Θ_α, Θ_β can be determined from (B.2) and (B.3).

While the dynamics strongly resemble the standard Hamiltonian equations, there is a substantial difference due to the fact that the number of points double as we move from one level to the next finer level, i.e., (B.6) involves one node t but two nodes, $t\alpha$ and $t\beta$, at the next level. This asymmetry in the number of variables in (B.6) makes it impossible to "diagonalize" the Hamiltonian, i.e., to decouple the dynamics and boundary conditions into separate upward and downward dynamics driven by $y(t)$, and thus there is no two-filter algorithm as in [6] and [7]. We can, however, triangularize these dynamics and boundary conditions to obtain an RTS algorithm.

Specifically, drawing inspiration from [6] and [7], consider a t-varying transformation

$$\begin{bmatrix} x^u \\ \hat{x} \end{bmatrix}_t = T_{m(t)} \begin{bmatrix} \hat{x} \\ \hat{\lambda} \end{bmatrix}_t \tag{B.7}$$

$$T_m = \begin{bmatrix} \Gamma_m & I \\ I & 0 \end{bmatrix}. \tag{B.8}$$

where we wish to transform the Hamiltonian dynamics and boundary conditions into a form in which there is an upward recursion for x^u followed by a downward recursion for $\hat{x}_s$. Note that we are free to multiply (B.6) on the left by an

[3]Note that, as is typically done in the standard case, we have added an $(M+1)$st level to $\hat{\lambda}(t)$ to simplify the form of the boundary condition.

invertible matrix, $S_{m(t)}$. Thus, we wish to transform the dynamics

$$S_{m(t)}\mathcal{A}T^{-1}_{m(t)}\begin{bmatrix}x^u\\ \hat{x}\end{bmatrix}_t + S_{m(t)}\Theta_\alpha T^{-1}_{m(t)+1}\begin{bmatrix}x^u\\ \hat{x}\end{bmatrix}_{\alpha t} + S_{m(t)}\Theta_\beta T^{-1}_{m(t)+1}\begin{bmatrix}x^u\\ \hat{x}\end{bmatrix}_{\beta t} = \begin{bmatrix}C^TR^{-1}y(t)\\ 0\\ 0\end{bmatrix} \quad \text{(B.9)}$$

where, the desired forms of the various matrices are as follows:

$$S_m = \begin{bmatrix}-P_m^{-1}A^{-1} & -P_m^{-1}A^{-1} & I\\ 0 & I & 0\\ I & 0 & 0\end{bmatrix} \quad \text{(B.10)}$$

$$S_m\mathcal{A}T_m^{-1} = \begin{bmatrix}I & 0\\ L_1 & L_2\\ L_3 & L_4\end{bmatrix} \quad \text{(B.11)}$$

$$S_m\Theta_\alpha T_{m+1}^{-1} = \begin{bmatrix}F_{m+1} & 0\\ 0 & 0\\ N & G_{m+1}\end{bmatrix} \quad \text{(B.12)}$$

$$S_m\Theta_\beta T_{m+1}^{-1} = \begin{bmatrix}F_{m+1} & 0\\ N & G_{m+1}\\ 0 & 0\end{bmatrix} \quad \text{(B.13)}$$

and after some algebra [9] we find $L_1 = L_3 = 0, L_2 = L_4 = -A, N = -BB^T$,

$$P_m^{-1} = P_{ML}^{-1}(m|m+1) \quad \text{(B.14)}$$

with the boundary condition $P_M^{-1} = 0$,

$$\Gamma_m = P_{ML}^{-1}(m|m) \quad \text{(B.15)}$$

and $F_m = -J^T(m)$ and $G_m = AJ^{-1}(m)$, where J is defined in Section II.

Equations (B.9)–(B.13) yield a fine-to-coarse filtering recursion given by

$$x^u(t) = J^T(m(t)+1)[x^u(t\alpha) + x^u(t\beta)] + C^TR^{-1}y(t), \qquad m(t) = 0, \cdots, M-1 \quad \text{(B.16)}$$

with initial conditions

$$x^u(t) = C^TR^{-1}y(t), \qquad m(t) = M. \quad \text{(B.17)}$$

Also, using the boundary conditions at $t = 0$ yields the initial condition

$$\hat{x}_s(0) = [\Gamma_0 + P_x^{-1}(0)]^{-1}x^u(0) \quad \text{(B.18)}$$

for the downward recursion, which we obtain directly from (B.9)–(B.13):

$$\hat{x}_s(t) = J(m(t))\hat{x}_s(t) + J(m(t))A^{-1}BB^Tx^u(t). \quad \text{(B.19)}$$

Finally, comparing (B.16)–(B.19) to (A.4)–(A.14), we see that this triangularization yields the information filter form of the ML RTS algorithm.

REFERENCES

[1] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automat. Contr.*, vol. 39, no. 3, pp. 464–478, 1994.

[2] A. Benveniste, R. Nikoukhah, and A. S. Willsky, "Multiscale system theory," in *Proc. 29th IEEE Conference Decision Cont.*, Honolulu, HI, Dec. 1990.

[3] J. Deyst, and C. Price, "Conditions for asymptotic stability of the discrete, minimum variance, linear, estimator," *IEEE Trans. Autom. Contr.*, vol. AC-13, pp. 702–705, Dec. 1968.

[4] R. Nikoukhah, A. S. Willsky, and B. C. Levy, "Kalman filtering and Riccati equations for descriptor systems," *IEEE Trans. Automat. Contr.*, vol. 37, no. 9, pp. 1325–1342, 1992.

[5] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.

[6] M. B. Adams, A. S. Willsky, and B. C. Levy, "Linear estimation of boundary value stochastic processes: Part I: The role and construction of complementary models; Part II: 1-D smoothing problems," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. no. 9, 803–821, Sept. 1984.

[7] T. Kailath and L. Ljung, "Two filter smoothing formulas by diagonalization of the Hamiltonian equations," *Int'l J. Contr.*, vol. 36, pp. 663–673, 1982.

[8] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

[9] K. C. Chou, "A stochastic modeling approach to multiscale signal processing," Ph.D. dissertation, MIT Dep. Elec. Eng. and Com. Sci., Cambridge, MA, May 1991.

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

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

[12] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. on Inf. Theory,* vol. 38, no. 2, pp. 910–917, Mar. 1992.

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

[14] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intel.,* vol. 11, pp. 674–693, July 1989.

[15] Y. Meyer, "L'analyse par Ondelettes," *Pour le Science,* Sept. 1987.

[16] W. Hackebush and V. Trottenberg, eds. *Multigrid Methods.* New York: Springer-Verlag, 1982.

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

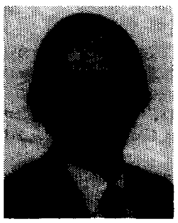
[18] M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Efficient multiscale regularization with application to the computation of optical flow," submitted for publication.

[19] M. R. Luettgen, W. C. Karl, A. S. Willsky and R. R. Tenney, "Multiscale representations of markov random fields," *IEEE Trans. Sign. Proc.,* to appear.

Kenneth C. Chou (S'90-M'91) received the S.B., S.M., E.E., and Ph.D degrees in electrical engineering from the Massachusetts Institute of Technology, (MIT) Cambridge, MA in 1985, 1987, 1987, and 1991, respectively.

He is currently a research engineer at SRI International in the Acoustics and Radar Technology Laboratory, Menlo Park, CA. His current research interests include detection and estimation theory involving multiscale stochastic processes, multiscale system identification, and applications of multiscale signal processing to machine monitoring and radar imaging. His interests also include MIMO robust control of active noise control systems.

Dr. Chou is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.

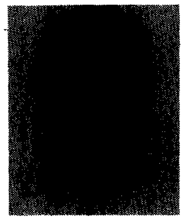


Alan S. Willsky (S'70, M'73, SM'82, F'86) received both the S.B. degree and the Ph.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 1969 and 1973, respectively.

He joined the MIT faculty in 1973, and he is presently Professor of Electrical Engineering. From 1974 to 1981, Dr. Willsky served as Assistant Director of the MIT Laboratory for Information and Decision Systems. He is also a founder and member of the board of directors of Alphatech, Inc. He has held visiting positions at Imperial College, London, L'Université de Paris-Sud, and the Institut de Recherche en Informatique et Systèmes Aléatoires in Rennes, France.

Dr. Willsky was program chairman for the 17th IEEE Conference on Decision and Control, has been an associate editor of several journals including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, has served as a member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society, was program chairman for the 1981 Bilateral Seminar on Control Systems held in the People's Republic of China, and was special guest editor of the 1992 special issue of the IEEE TRANSACTIONS ON INFORMATION THEORY on wavelet transforms and multiresolution signal analysis. Also in 1988 he was made a Distinguished Member of the IEEE Control Systems Society. In addition, he has given several plenary lectures at major scientific meetings including the 20th IEEE Conference on Decision and Control, the 1991 IEEE International Conference on Systems Engineering, the SIAM Conference on Applied Linear Algebra, 1991, and the 1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems, Canberra, Australia.

He is the author of the research monograph *Digital Signal Processing and Control and Estimation Theory* and is co-author of the undergraduate text *Signals and Systems.* In 1975 he received the Donald P. Eckman Award from the American Automatic Control Council. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph. His present research interests are in problems involving multidimensional and multiresolution estimation and imaging, discrete-event systems, and the asymptotic analysis of control and estimation systems.



Ramine Nikoukhah was born in Teheran, Iran, on March 30, 1961. He received the B.S. degree from the University of California, Davis, in 1982, the degree of Maitrise from the University of Paris-Sud, Orsay, France, in 1983, and the M.S. and Engineer's degrees in 1986 and the Ph.D. degree in 1988 from Massachusetts Institute of Technology, (MIT) Cambridge, MA all in electrical engineering and computer science.

During 1983 has was a Research Engineer at TAU Co., Los Gatos, CA, where he worked in the area of flight control. Since 1988, he has been with the Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt, France. He has also done consulting for Matra Co. and has held teaching positions at various engineering schools in France. He is currently involved in the development of the control and signal processing software package Scilab at INRIA.

R. Nikoukhah's research interests are in the areas of control, signal processing, and algebraic system theory.