LIDS-P-2152

MULTISCALE SYSTEMS, KALMAN FILTERS, AND RICCATI EQUATIONS

Kenneth C. Chou[1]
Alan S. Willsky[1]
Ramine Nikoukhah[2]

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.

[1]Laboratory for Information and Decision Systems and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge,MA 02139, USA. The work of these authors was also supported in part by the Air Force Office of Scientific Research under Grant AFOSR-88-0032, 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. In addition some of this research was performed while KCC and ASW were visitors at IRISA and while ASW received support from INRIA. K.C. Chou is presently with SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025.

[2]INRIA, Domaine de Voluceau, Rocquencourt, BP105, 78153 Le Chesnay, CEDEX, FRANCE.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multiscale Systems, Kalman Filters, and Riccati Equations	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,Laboratory for Information and Decision Systems,77 Massachusetts Avenue,Cambridge,MA,02139-4307**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**47**	

1 Introduction

In a companion paper [1] we introduce 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. Such pyramidal representations for signals and images have been and continue to be of considerable interest, both in research and in application, since they suggest efficient and highly parallelizable computational structures and also appear to be natural forms of representation for many phenomena including those with fractal or self-similar features. The framework introduced in [1] had as its motivation the development of a rational framework for statistical modeling and optimal processing based on such pyramidal representation, and the potential of this framework was illustrated in [1] 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.

One of the other contributions of this work, we feel, is in identifying the significant role that systems and control researchers can have in this area, as multiresolution modeling and analysis problems have a strong systems flavor. For example, the optimal estimation algorithm [1] can be viewed as 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 theory of multiresolution modeling, requiring techniques for realization and identification, and the detailed system-theoretic analysis of the filtering algorithms developed in [1]. The objective of this paper is to tackle this latter problem, while an initial investigation of multiscale 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]. As we discuss, the objective of error and stability analysis for multiscale filtering leads directly to a variation on this algorithm which we develop in Section 3. 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 4 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 5 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.

2 State Space Models and Multiscale Estimation on Dyadic Trees

In this section we briefly review the formulation and results in [1]. As illustrated in Figure 1, the basic data structure for multiresolution modeling is the dyadic tree. Here each node t in the tree 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. As illustrated in the figure, it is useful to visualize T as having horizontal levels corresponding to different scales, where increasing m corresponds to moving to finer resolutions. While we will find it convenient to use the more compact notation t for nodes on T, rather than the scale-translation pair (m,n), we will on occassion wish to refer to the scale of a particular node t, which we denote by $m(t)$. Also, again as illustrated in the figure, we will define our dynamic operations in terms of basic shift operators, namely the unique backward shift $\overline{\gamma}$ 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 $t\overline{\gamma} = (m-1, [\frac{n}{2}])$. 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 (e.g. low-pass filtered) representation (sec [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 :

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

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

The term $A(t)x(t\overline{\gamma})$ in (2.1) represents a coarse–to–fine prediction or interpolation, $B(t)w(t)$ represents the higher resolution detail added in going from one scale to the next finer scale, and $y(t)$ is the measured variable (if any) at the particular scale m and location n represented by t. This model serves as the basis for the 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

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

Note that the general models (2.1)-(2.3) allow full t-dependence of all the system matrices, and several of the applications described in [1] require this general dependence. An important special case is that in which 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. For simplicity we focus the detailed covariance analysis and stability results on this case, while our investigation of steady-state behavior, of course, looks at the further specialization to constant–parameter models.

In [1] we analyze the second–order statistics of (2.1) when $w(t)$ and $v(t)$ are independent, zero-mean white noise processes with covariances I and $R(t)$, respectively. We also assume that $w(t)$ is independent of the "past" of x, i.e. $\{x(\tau)|m(\tau) < m(t)\}$. 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 Figure 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.

The covariance $P_x(t) = E[x(t)x^T(t)]$ evolves according to a Lyapunov equation on the tree:

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

In the scale-varying model, *i.e.* the case in which the model parameters vary in scale only, if at some scale $P_x(t)$ is constant, then this holds at each scale, so that by an abuse of notation $P_x(t) = P_x(m(t))$, and we have a scale–to–scale Lyapunov equation:

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

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:

$$P_x = AP_xA^T + BB^T \tag{2.6}$$

In our development and analysis of smoothing algorithms, we encounter the need for fine–to–coarse prediction and recursion. In particular, the reversal of (2.1), *i.e.* a model representing $x(t\overline{\gamma})$ as a linear function of $x(t)$ and a noise that is uncorrelated with $x(t)$ is given by

$$x(t\overline{\gamma}) = F(t)x(t) - A^{-1}(t)B(t)\tilde{w}(t) \tag{2.7}$$

with

$$\begin{aligned} F(t) &= A^{-1}(t)[I - B(t)B^T(t)P_x^{-1}(t)] \\ &= P_x(t\overline{\gamma})A^T(t)P_x^{-1}(t) \end{aligned} \tag{2.8}$$

and where

$$\begin{aligned} \tilde{w}(t) &= w(t) - E[w(t)|x(t)] && (2.9) \\ E[\tilde{w}(t)\tilde{w}^T(t)] &= I - B^T(t)P_x^{-1}(t)B(t) \\ &\triangleq \tilde{Q}(t) && (2.10) \end{aligned}$$

In [1] we derive a generalization of the Rauch-Tung-Striebel smoothing algorithm consisting of a fine–to–coarse Kalman filtering step followed by coarse–to–fine smoothing step. Specifically, let $\hat{x}(s|t)$ denote the optimal estimate of $x(s)$ based on data at or "below" node t (i.e. $y(\tau)$ for $\tau = t$ or τ a descendant 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

$$\begin{aligned} \hat{x}(t|t) &= \hat{x}(t|t+) + K(t)[y(t) - C(t)\hat{x}(t|t+)] && (2.11) \\ K(t) &= P(t|t+)C^T(t)V^{-1}(t) && (2.12) \\ V(t) &= C(t)P(t|t+)C^T(t) + R(t) && (2.13) \end{aligned}$$

$$P(t|t) = [I - K(t)C(t)]P(t|t+) \quad (2.14)$$

a coarse–to–fine one-step prediction step:

$$\begin{aligned} \hat{x}(t|t\alpha) &= F(t\alpha)\hat{x}(t\alpha|t\alpha) && (2.15) \\ \hat{x}(t|t\beta) &= F(t\beta)\hat{x}(t\beta|t\beta) && (2.16) \end{aligned}$$

with corresponding error covariances given by

$$\begin{aligned} P(t|t\alpha) &= F(t\alpha)P(t\alpha|t\alpha)F^T(t\alpha) + \mathcal{Q}(t\alpha) && (2.17) \\ \mathcal{Q}(t\alpha) &= A^{-1}(t\alpha)B(t\alpha)\tilde{Q}(t\alpha)B^T(t\alpha)A^{-T}(t\alpha) && (2.18) \\ P(t|t\beta) &= F(t\beta)P(t\beta|t\beta)F^T(t\beta) + \mathcal{Q}(t\beta) && (2.19) \\ \mathcal{Q}(t\beta) &= A^{-1}(t\beta)B(t\beta)\tilde{Q}(t\beta)B^T(t\beta)A^{-T}(t\beta) && (2.20) \end{aligned}$$

and a fusion step to merge the estimates (2.15) and (2.16), to form $\hat{x}(t|t+)$:

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

(2.21)

$$P(t|t+) = [P^{-1}(t|t\alpha) + P^{-1}(t|t\beta) - P_x^{-1}(t)]^{-1} \quad (2.22)$$

This filtering algorithm has an obvious pyramidal structure, allowing substantial parallelization. Note that while the update and prediction steps are analogous to the corresponding steps in usual Kalman filtering (although, as discussed in [1], this step **must** proceed from fine–to–coarse and hence must use the backward model (2.7) for the prediction step), the fusion step has no counterpart in the standard case, and, as we'll see this leads to some interesting differences in our analysis of the filtering algorithm.

Finally, let $\hat{x}_s(t)$ denote the optimal estimate of $x(t)$ based on all available data on a finite subtree with root node 0 and M scales below it. Once the Kalman filter has reached the root node at the top of the tree, we have computed $\hat{x}_s(0) = \hat{x}(0|0)$, which serves as the initial condition for the coarse–to–fine RTS smoothing sweep which also has a parallel, pyramidal structure :

$$\hat{x}_s(t) = \hat{x}(t|t) + J(t)\,[\hat{x}_s(t\overline{\gamma}) - \hat{x}(t\overline{\gamma}|t)] \quad (2.23)$$

$$J(t) \triangleq P(t|t)F^T(t)P^{-1}(t\overline{\gamma}|t) \quad (2.24)$$

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

$$P_s(t) = P(t|t) + J(t)[P_s(t\overline{\gamma}) - P(t\overline{\gamma}|t)]J^T(t) \quad (2.25)$$

3 The ML Filter

The fine–to–coarse filtering equations presented in the preceding section have several significant differences with standard Kalman filtering analysis and present some difficulties in analysis that provide motivation for a slightly different algorithm. Specifically the Riccati equation (2.12)-(2.14), (2.17)-(2.20), (2.22) for our optimal filter, 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 (2.22). The latter of these is intrinsic to our Riccati equations and, as we will see, 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 P_x does not appear. Specifically, in standard Kalman filtering analysis the Riccati equation for the error covariance can be viewed simply as the covariance of the error equations, which can be analyzed directly without explicitly examining the state dynamics since the error evolves as a state process itself. This is apparently **not** the case here because of the explicit presence of $P_x(t)$ in (2.22) and in the backward model parameters (2.7)-(2.10) that enter into the fine–to–coarse prediction step (2.17)-(2.20). On first examination, this might not appear to be a new problem, as backward models for standard temporal models also involve the state covariance. However, the present situation is not as simple. First of all, as discussed in [1], the driving noises in (2.7) are *not* white (except along fine–to–coarse paths). Also, and more importantly, the new fusion step adds a new twist. In particular, if we examine the backward model (2.7)-(2.10) and the Kalman filter (2.11), (2.15), (2.16), (2.21), 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, and, in particular, their stability. To overcome this, we consider a slight variation in the filtering and RTS algorithm.

Specifically, we define what we will refer to as the *ML filter* by setting the $P_x^{-1}(t)$ terms in (2.11)-(2.22) to zero. The resulting filter recursions are then given by

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:

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

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

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 difference here are the absence of a $P_x^{-1}(t)$ term in (3.8) (compare to (2.22)), and the changes to the prediction step (3.5)–(3.6) due to the simpler form of the backward model (2.7)- (2.10) when $P_x^{-1}(t) = 0$.

As shown in Appendix A, $\hat{x}_{ML}(t|t)$ has the interpretation as the ML estimate of $x(t)$, viewed as an unknown vector, based on the measurements Y_t. Thus the Bayesian estimate of the preceding section and its covariance can be computed as follows:

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

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

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 (2.23)- (2.25) (using (3.9), (3.10)) so that the ML estimator computed in the ML filtering sweep are used in the smoothing step. Specifically, as shown in Appendix A

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

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

where

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

Also as in standard Kalman filtering, 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 5, 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 through analysis of the Hamiltonian equations for our estimation problem. The Hamiltonian and the two-point boundary-value problem associated with it plays a central role in the theory of smoothing for standard state space models. For example, as discussed in [6], [7], triangularization of the Hamiltonian 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. This is developed, for simplicity in the constant-parameter case in Appendix B.

Finally, as mentioned at the beginning of this subsection one of the motivations for introducing the ML filter is that its use allows us 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

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

Also, from (2.1)

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

$$x(t) = A^{-1}(t\beta)x(t\beta) - A^{-1}(t\beta)B(t\beta)w(t\beta) \tag{3.16}$$

and thus, using (3.8)

$$\begin{aligned} x(t) = {} & P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)x(t\alpha) + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)x(t\beta)] \\ & -P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)B(t\alpha)w(t\alpha) \\ + \; & P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)B(t\beta)w(t\beta)] \end{aligned} \tag{3.17}$$

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

$$\begin{aligned} & \tilde{x}_{ML}(t|t) \\ = \; & [I - K_{ML}(t)C(t)]P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)\tilde{x}(t\alpha|t\alpha) + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)\tilde{x}(t\beta|t\beta)] \\ - \; & P_{ML}(t|t+)[P_{ML}^{-1}(t|t\alpha)A^{-1}(t\alpha)B(t\alpha)w(t\alpha) + P_{ML}^{-1}(t|t\beta)A^{-1}(t\beta)B(t\beta)w(t\beta)] \\ - \; & K_{ML}(t)v(t) \end{aligned} \tag{3.18}$$

Note that (3.14), (3.17), and (3.18) each represents a fine-to-coarse system of the form of (2.3), and in particular, (3.18) represents the filtering error as the state of such a system driven by white process and measurement noise. It is the stability of this system–in the scale-varying case–that is investigated in Section 5.

4 System–Theoretic Concepts for Fine–To–Coarse Dynamic Models

In this section we introduce and investigate several system–theoretic concepts for dynamic systems on dyadic trees. The structure of the tree leads to several important differences with standard–state space system theory, and furthermore this setting appears to be a natural one in which to develop a theory for multiresolution modeling and realization. Our goals here, however, are far more modest. In particular we refer the reader to [2] for a first step in developing such a multiscale realization theory and

focus here on the specific constructs and results needed in Section 5 for the asymptotic analysis of the fine–to–coarse filtering algorithm described in the preceding section.

In particular, we focus here on the fine–to–coarse model (2.2), (2.3). Moreover, the analysis of Section 5 focuses for scale–varying systems, and thus we focus here on the analogous specialization of (2.2),(2.3), namely

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

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

where, since we focus in this section on deterministic properties, $w(t)$ in (4.1) should be viewed as an input, and we have eliminated the measurement noise from the observation (4.2). Furthermore, to simplify the discussion we assume the $F(m)$ is invertible for all m.

4.1 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— i.e., there are 2 descendants one scale finer than t_0, 4 descendants two scales finer, etc. Thus there are 2^m "initial conditions" affecting $x(t_0)$ and at a scale M levels finer than $x(t_0)$. Thus let us define the following vectors,

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

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

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

Definition 4.1 *The system (4.1) is* **reachable** *from X_{M,t_0} to $x(t_0)$ if given any $\overline{X}_{M,t_0}$ and any desired $\overline{x}(t_0)$, it is possible to specify W_{M,t_0} so if $X_{M,t_0} = \overline{X}_{M,t_0}$, then $x(t_0) = \overline{x}(t_0)$.*

As always, in studying conditions for reachability, we can set $X_{M,t_0} = 0$, in which case

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

where

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

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

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

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

Let us define the *reachability Gramian*

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

Thus since the rank of $\mathcal{G}$ equals the rank of $\mathcal{G}\mathcal{G}^T$, we see that the system (4.1) is reachable from X_{M,t_0} to $x(t_0)$ if $\mathcal{R}(t_0, M)$ is invertible. Also we can now define a notion of uniform reachability:

Definition 4.2 *The system (4.1) is* **uniformly reachable** *if there exists $\gamma, M_0 > 0$ so that*

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

Note that $\mathcal{R}(t_0, M)$ bears a strong similarity to the standard reachability gramian for the following system.

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

Furthermore the factor of $\sqrt{2}$ in (4.12) certainly affects the absolute magnitude of the reachability gramian, but it does *not* affect 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. $rank[G|FG|...|F^{n-1}G] = n$.

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). However, this is *not* true for the coarse–to–fine model (*e.g.* (2.1) or it s 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 than 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

$$Y_{M,t_0} \triangleq [\ y^T(t_0)|\ \ y^T(t_0\alpha),\ \ y^T(t_0\beta)|\ \ \ldots\ \ |y^T(t_0\alpha^M),\ldots y^T(t_0\beta^M)\ ]^T \tag{4.13}$$

Definition 4.3 *The system (4.1),(4.2) is* **reconstructible** *from X_{M,t_0} to $x(t_0)$ if given knowledge of W_{M,t_0} and Y_{M,t_0}, we can uniquely determine $x(t_0)$.*

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 $\mathcal{H}_M$ is most easily visualized if we partition it compatibly with the levels of the observations in Y_{M,t_0}:

$$\mathcal{H}_M \triangleq \left[\begin{array}{cccccccccccc}
\overbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}^{2^M \text{ blocks}} & & & & & & & & & & & \\
\Theta(0) & \Theta(0) & \ldots & & & & & & & & \ldots & \Theta(0) \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
\Theta(1) & \ldots & & & \ldots & \Theta(1) & 0 & \ldots & & & \ldots & 0 \\
0 & \ldots & & & \ldots & 0 & \Theta(1) & \ldots & & & \ldots & \Theta(1) \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
\Theta(2) & \ldots & \Theta(2) & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & \ldots & 0 & \Theta(2) & \ldots & \Theta(2) & 0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & \ldots & 0 & \Theta(2) & \ldots & \Theta(2) & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & \Theta(2) & \ldots & \Theta(2) \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
 & & & & \vdots & & & \vdots & & & & \\
 & & & & \vdots & & & \vdots & & & & \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
\Theta(M) & 0 & \ldots & & & & & & & & \ldots & 0 \\
0 & \Theta(M) & \ldots & & & & & & & & \ldots & 0 \\
\vdots & & & & & & & & & & & \vdots \\
0 & 0 & \ldots & & & & & & & & \ldots & \Theta(M)
\end{array}\right] \tag{4.15}$$

Here

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

As a simple example to help clarify the structure of the matrix $\mathcal{H}_M$ consider the matrix $\mathcal{H}_2$ for the scale-invariant case, i.e. where $F(m) = F$, $C(m) = C$.

$$\mathcal{H}_\epsilon = \begin{bmatrix} CF^2 & CF^2 & CF^2 & CF^2 \\ CF & CF & 0 & 0 \\ 0 & 0 & CF & CF \\ C & 0 & 0 & 0 \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix} \tag{4.17}$$

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 upward *observability* is indeed a very strong condition. Specifically, 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. However, the situation for *reconstructibility* is very different. Specifically, if $W_{M,t_0} = 0$, then

$$x(t_0) = \phi(m(t_0), m(t_0) + M) I_M X_{M,t_0} \tag{4.18}$$

where

$$I_M = \underbrace{[I|I|...|I]}_{2^M \text{ times}} \tag{4.19}$$

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

Since the condition of reconstructibility only requires being able to uniquely determine the single point $x(t_0)$ from the measurements in the subtree, we guarantee this condition by requiring that any vector in the nullspace, of (4.14) is also in the nullspace of (4.18). 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} (which is all that affects $x(t_0)$) from Y_{M,t_0}. We then have

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* $\mathcal{O}(t_0, M)$*:*

$$\begin{aligned}\mathcal{O}(t_0, M) &= I_M \mathcal{H}_M^T \mathcal{H}_M I_M^T \\ &= \sum_{i=0}^{M} 2^{2M-i} \phi^T(m(t_0)+i, m(t_0)+M) C^T(m(t_0)+i) \\ &\times\ C(m(t_0)+i)\phi(m(t_0)+i, m(t_0)+M) \qquad (4.20)\end{aligned}$$

Proof: We need only show that $\mathcal{N}(\mathcal{H}_M) \subseteq \mathcal{N}(I_M)$ is equivalent to the invertibility of $\mathcal{O}(t_0, M) = I_M \mathcal{H}_M^T \mathcal{H}_M I_M^T$. Suppose first that $\mathcal{O}(t_0, M)$ is not invertible. Then there exists $y \neq 0$ so that $\mathcal{H}_M z = 0$ where $z = I_M^T y$. Note that I_M^T is one-to-one so that $z \neq 0$ which in turn implies that $I_M z = I_M I_M^T y \neq 0$ which contradicts $\mathcal{N}(\mathcal{H}_M) \subseteq \mathcal{N}(I_M)$. If on the other hand $\mathcal{N}(\mathcal{H}_M)$ is not included in $\mathcal{N}(\mathcal{I}_M)$, then there exists an 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 $z \epsilon \mathcal{N}(I_M)$. Making this substitution into $\mathcal{H}_M x = 0$ and left-multiplying by $I_M \mathcal{H}_M^T$, we get

$$I_M \mathcal{H}_M^T \mathcal{H}_M I_M^T y + I_M \mathcal{H}_M^T \mathcal{H}_M z = 0 \qquad (4.21)$$

However, a straightforward but tedious calculation [9] yields

$$I_M \mathcal{H}_M^T \mathcal{H}_M = \overline{\Lambda}_T I_M \qquad (4.22)$$

where $\overline{\Lambda}$ is an nxn matrix computed from the elements of $\mathcal{H}_M$ in (4.15). Equation (4.22) is a consequence of the special structure of $\mathcal{H}_M^T \mathcal{H}_M$. In particular it indicates that the columns 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 (4.22) represents the coarsest scale component of that transform.

If we now substitute (4.22) into (4.21) 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, as in the case of reachability, it is useful to define a notion of uniformity:

Definition 4.4 *The system (4.1), (4.2) is* **uniformly reconstructible** *if there exists* $\delta, M_0 > 0$ *so that*

$$\mathcal{O}(t, M_0) \geq \delta I \text{ for all } t \qquad (4.23)$$

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

$$x(m) = \frac{1}{2}F(m+1)x(m+1) + \frac{1}{2}G(m+1)u(m+1) \tag{4.24}$$

$$y(m) = \sqrt{2}C(m)x(m) \tag{4.25}$$

Thus as for reachability, the conditions for reconstructibility and uniform reconstructibility for our model is the same as the usual notions for the pair $F(m), C(m)$. For example 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.

4.2 Stability

Finally, let us examine the question of asymptotic stability for the autonomous version of (4.1), i.e.

$$z(t) = F(m(t)+1)[z(t\alpha) + z(t\beta)] + \mathcal{G}(m(t))u(t) \tag{4.26}$$

as the dynamics propagate up the tree. Intuitively what we would like stability to mean is that $z(t) \to 0$ as we propagate farther and farther away from the initial level of the tree. Note, however, that as we move up the tree, $z(t)$ is influenced by a geometrically increasing number of nodes at the initial level. For example, $z(t)$ depends on $\{z(t\alpha), z(t\beta)\}$ or, alternatively on $\{z(t\alpha^2), z(t\beta\alpha), z(t\alpha\beta), z(t\beta^2)\}$, etc. Thus in order to study asymptotic stability it is necessary to consider an infinite dyadic tree, with an infinite set of initial conditions corresponding to all nodes at the initial level.

For the remainder of this discussion, we adopt a change of notation to a more standard form for stability analysis of dynamic systems. Specifically, we change 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

$$z_i(m) = F(m-1)(z_{2i}(m-1) + z_{2i+1}(m-1)) \tag{4.27}$$

Let $Z(m)$ denote the infinite sequence at level m, i.e. the set $\{z_i(m)\ , i \in \mathcal{Z}\}$. The p-norm on such a sequence is defined as

$$\|Z(m)\|_p \triangleq (\sum_i \|z_i(m)\|_p^p)^{\frac{1}{p}} \tag{4.28}$$

where $\|z_i(m)\|_p$ is the standard p-norm for the finite dimensional vector $z_i(m)$.

Definition 4.5 *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 $\|Z(0)\|_p < \infty$,*

$$\|Z(m)\|_p \leq C\alpha^m \|Z(0)\|_p \tag{4.29}$$

From (4.27) we can immediately write the following.

$$z_i(m) = \Phi(m,0) \sum_{j \in O_{m,i}} z_j(0) \tag{4.30}$$

where the cardinality of the set $O_{m,i}$ is 2^m and $\Phi(m,0)$ is the state transition matrix associated with $F(m)$. As one would expect, it is this matrix that controls the stability properties of (4.27), although the structure of the tree leads to important differences.

Theorem 4.2 *The system defined in eq.(4.27) is l_p-exponentially stable if and only if*

$$2^{\frac{m}{q}} \|\Phi(m,0)\|_p \leq K'\gamma^m \quad \textit{for all } m \tag{4.31}$$

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

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{4.32}$$

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

$$\|\Phi(m,0)z\|_p > K\gamma^m 2^{-\frac{m}{q}}\|z\|_p \tag{4.33}$$

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 $\rho_0, \rho_1, \rho_2, \ldots$ be a sequence with

$$\sum_{i=0}^{\infty} \rho_i^p = 1 \tag{4.34}$$

Then let

$$z_i(0) = \begin{cases} \rho_0 z & 0 \leq i < 2^m \\ \rho_1 z & 2^m \leq i < 2 \cdot 2^m \\ \vdots & \\ \rho_i z & j2^m \leq i < (j+1)2^m \\ \vdots & \end{cases} \tag{4.35}$$

Note that

$$\|Z(0)\|_p^p = 2^m\|z\|_p^p \tag{4.36}$$

Thus, using (4.33)- (4.36)

$$\begin{aligned} \|Z(m)\|_p^p &= 2^{mp}\|\Phi(m,0)z\|_p^p \\ &> 2^{mp}K^p\gamma^{mp}2^{\frac{-mp}{q}}\|z\|_p^p \\ &= 2^{mp}K^p\gamma^{mp}2^{\frac{-mp}{q}}2^{-m}\|Z(0)\|_p^p \\ &= K^p\gamma^{mp}\|Z(0)\|_p^p \end{aligned} \tag{4.37}$$

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

$$\|Z(m)\|_p > K\gamma^m\|Z(0)\|_p \tag{4.38}$$

so that the system cannot be l_p-exponentially stable.

To prove sufficiency we use two simple facts. First, (4.27) is exponentially stable if there exist $0 \leq \beta < 1$ and $K > 0$ so that for each i

$$\|z_i(m)\|_p \leq K\beta^m \Big(\sum_{j \in O_{m,i}} \|z_j(0)\|_p^p\Big)^{\frac{1}{p}} \tag{4.39}$$

This follows by raising (4.39) to the pth power and summing over i. Secondly, for any sequence of vectors x_i and any m and j

$$\Big\|\sum_{i \in \mathcal{I}_{m,j}} x_i\Big\|_p \leq 2^{\frac{m}{q}} \Big(\sum_{i \in \mathcal{I}_{m,j}} \|x_i\|_p^p\Big)^{\frac{1}{p}} \tag{4.40}$$

where $\mathcal{I}_{m,j} = \{j, j+1, \ldots j+2^m-1\}$. To show this, note that

$$\|a+b\|_p \leq 2^{\frac{1}{q}} (\|a\|_p^p + \|b\|_p^p)^{\frac{1}{p}} \tag{4.41}$$

Specifically, since $\|\cdot\|_p^p$ is a convex function, we can write

$$\|(\frac{1}{2})a + (1-\frac{1}{2})b\|_p^p \leq (\frac{1}{2})\|a\|_p^p + (1-\frac{1}{2})\|b\|_p^p \tag{4.42}$$

from which eq.(4.41) follows immediately. Using this we can show (4.40) by induction on m. Note first that (4.40) is trivially true for $m = 0$. Suppose then that for all j (4.40) 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

$$\Big\|\Big(\sum_{i \in \mathcal{I}_{m,j_1}} x_i + \sum_{i \in \mathcal{I}_{m,j_2}} x_i\Big)\Big\|_p \leq 2^{\frac{1}{q}} \Big(\Big\|\Big(\sum_{i \in \mathcal{I}_{m,j_1}} x_i\Big\|_p^p + \Big\|\Big(\sum_{i \in \mathcal{I}_{m,j_2}} x_i\Big\|_p^p\Big)^{\frac{1}{p}} \tag{4.43}$$

Then by substituting into (4.40) into (4.43) we get

$$\Big\|\sum_{i \in \mathcal{I}_{m,j_1} \cup \mathcal{I}_{m,j_2}} x_i\Big\|_p \leq 2^{\frac{(m+1)}{q}} \Big(\Big\|\Big(\sum_{i \in \mathcal{I}_{m,j_1}} x_i\Big\|_p^p + \Big\|\Big(\sum_{i \in \mathcal{I}_{m,j_2}} x_i\Big\|_p^p\Big)^{\frac{1}{p}} \tag{4.44}$$

and apply (4.41), we find that (4.40) holds for $m+1$ as well.

We can now complete the proof of the theorem. By applying the p-norm to eq.(4.30), using the Cauchy-Schwarz inequality, and then (4.40) we obtain

$$\|z_i(m)\|_p \leq \|\Phi(m,0)\|_p 2^{\frac{m}{q}} \Big(\sum_{j \in \mathcal{I}_{m,i}} \|z_j(0)\|_p^p\Big)^{\frac{1}{p}} \tag{4.45}$$

If we then assume that (4.31) holds this, together with (4.45) yields

$$\|z_i(m)\|_p \leq K'\gamma^m(\sum_{j\in\mathcal{I}_{m,i}} \|z_j(0)\|_p^p)^{\frac{1}{p}} \tag{4.46}$$

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 (4.27) is equivalent to the usual exponential stability of the system

$$\xi(m) = 2^{\frac{1}{q}}F(m-1)\xi(m-1) \tag{4.47}$$

For example for $p = 2$, we are interested in the exponential stability of

$$\xi(m) = \sqrt{2}F(m-1)\xi(m-1) \tag{4.48}$$

If F is constant this is equivalent to requiring F to have eigenvalues with magnitudes smaller than $\frac{\sqrt{2}}{2}$.

5 Covariance Bounds, Stability, and Steady-State Behavior

In this section we develop several system-thematic results for our fine-to-coarse filtering algorithms, 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 the Bayesian filtering algorithm of Section 2 becomes

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

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

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

with the scale-varying Riccati equation

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

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

where we have combined the update and fusion steps in (5.5). Also $F(m(t))$ and $\tilde{Q}(m(t))$ are given by (2.8), (2.10) in the scale-varying case and

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

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

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

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

In the ML case, with P_x^{-1} set to zero we obtain a further simplification of these equations or, equivalently, of (3.14):

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

Similarly we have the following simplified form of (3.18) for the ML filter error:

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

The ML Riccati equation in this case becomes

$$P_{ML}(m|m+1) = A^{-1}(m+1)P(m+1|m+1)A^{-T}(m+1)+G(m+1)G^T(m+1) \quad (5.11)$$

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

Also

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

and also, for future reference,

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

and this, together with (5.13) yield

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

5.1 Bounds on the Error Covariance

In this section we derive upper and lower bounds on the error covariances $P(m|m)$ and $P_{ML}(m|m)$. As is the case for standard Kalman filtering, [3,8], reachability and reconstructibility conditions are key in this analysis. In this case the system to be analyzed is the following backward model, obtained directly from (2.7)-(2.10) in the scale-varying case:

$$x(t) = \frac{1}{2}F(m(t)+1)[x(t\alpha) + x(t\beta)] + \frac{1}{2}G(m(t)+1)[\tilde{w}(t\alpha) + \tilde{w}(t\beta)] \quad (5.16)$$

together with the measurements (2.2). In this case, accounting for the scaling factor of $\frac{1}{2}$ in (5.16) and the covariances of $\tilde{w}$ and v, we define the stochastic reachability Grammian:

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

and the stochastic reconstructibility grammian:

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

where the state transition matrix is given by (4.8)-(4.9). In our analysis 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. In terms of our reachability and reconstructibility grammians these assumptions mean that for any $M_0 > 0$ we can find $\alpha, \beta > 0$ so that

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

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

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

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

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

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

These conditions coincide with those in Section 4.1 with the replacement of $F(m)$ by $\frac{1}{2}F(m)$, $G(m)$ by $\frac{1}{2}G(m)\tilde{Q}^{\frac{1}{2}}(m)$, and $C(m)$ by $R^{-\frac{1}{2}}(m)C(m)$. Furthermore, thanks to the boundedness assumptions, the relationship between the original model (2.1) in the scale-varying case and the reverse model (2.7)–(2.10), and the analysis in Section 4.1, it is straightforward to show that the uniform reachability and reconstructibility conditions are equivalent to the usual conditions for the pairs $(A^{-1}(m), G(m))$ and $(R^{-\frac{1}{2}}(m)C(m), A^{-1}(m))$, respectively.

We are now in a position to derive an upper bound for the optimal filter error covariance, $P(m|m)$. The general idea in deriving this bound is to make a careful comparison between the Riccati equations for our optimal filter and the Riccati equations for the standard Kalman filter. First consider the following lemma.

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*

$$\begin{aligned} \overline{P}(m|m+1) &= F(m+1)\overline{P}(m+1|m+1)F^T(m+1) \\ &+ G(m+1)\tilde{Q}(m+1)G^T(m+1) \qquad (5.23) \\ \overline{P}^{-1}(m|m) &= \overline{P}^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) \qquad (5.24) \end{aligned}$$

Then

$$\overline{P}^{-1}(m|m) \leq P^{-1}(m|m) \qquad (5.25)$$

Proof

First note that (5.5) can be rewritten as

$$P^{-1}(m|m) = P^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) + D^T(m)D(m) \qquad (5.26)$$

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

$$\begin{aligned} x(m) &= F(m+1)x(m+1) + G(m+1)w(m+1) \qquad (5.27) \\ E[w(m)w^T(m)] &= \tilde{Q}(m) \qquad (5.28) \\ y(m) &= C(m)x(m) + v(m) \qquad (5.29) \\ E[v(m)v^T(m)] &= R(m) \qquad (5.30) \end{aligned}$$

Similarly, the Riccati equation,(5.4), (5.26), characterizes the error covariance for the optimal filter corresponding to the filtering problem involving the same state equation but with the following augmented measurement equation.

$$\begin{aligned} \tilde{y}(m) &= \begin{bmatrix} C(m) \\ D(m) \end{bmatrix} x(m) + u(m) \qquad (5.31) \\ E[u(m)u^T(m)] &= \begin{bmatrix} R(m) & 0 \\ 0 & I \end{bmatrix} \qquad (5.32) \end{aligned}$$

so that (5.25) follows immediately.

□

We now state and prove the following theorem concerning an upper bound for $P(m|m)$.

Theorem 5.1 *Suppose there exists $\beta, \delta, M_0 > 0$ 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), (5.22) are equivalent to the existence of analogous uniform upper and lower bounds on the observability Gramian for (5.27)–(5.30). Thus standard Kalman filtering results imply that there exists a $\kappa > 0$ such that $\overline{P}(m|m) \leq \kappa I$ or $\overline{P}^{-1}(m|m) \geq \kappa^{-1} I$. From Lemma 5.1 we then have that $P^{-1}(m|m) \geq \kappa^{-1} I$ or $P(m|m) \leq \kappa I$.

□

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), (5.12), i.e.

$$\hat{P}^{-1}(m|m) \leq P_{ML}^{-1}(m|m) \tag{5.33}$$

where $\hat{P}(m|m)$ is the solution of a Riccati equation as in (5.23), (5.24) but with F and $\tilde{Q}$ replaced by A^{-1} and I, respectively. Then, as we have discussed, the conditions (5.20), (5.22) are equivalent to the analogous conditions on the usual observability gramian for the pair $(R^{-\frac{1}{2}}(m)C(m), A^{-1}(m))$. This in turn yields an upper bound on $\hat{P}(m|m)$. Using (5.33) we then have the following

Theorem 5.2 *Suppose that there exists $\beta, \delta, M_0 > 0$ 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 5.2 *Let*

$$\overline{S}(m|m) \triangleq \frac{1}{2}(P^{-1}(m|m) - C^T(m)R^{-1}(m)C(m) + P_x^{-1}(m)) \quad (5.34)$$

$$\overline{S}(m|m+1) \triangleq F^{-T}(m+1)P^{-1}(m+1|m+1)F^{-1}(m+1) \quad (5.35)$$

Consider also the Riccati equation

$$\begin{aligned} S^*(m|m+1) &= 2F^{-T}(m+1)S^*(m+1|m+1)F^{-1}(m+1) \\ &+ F^{-T}(m+1)C^T(m)R^{-1}(m)C(m)F^{-1}(m+1) \quad (5.36) \\ S^{*^{-1}}(m|m) &= S^{*^{-1}}(m|m+1) + G(m+1)\tilde{Q}(m+1)G^T(m+1) \quad (5.37) \end{aligned}$$

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

Proof

A straightforward calculation using (5.5) and (5.34) yields

$$\overline{S}(m|m) = P^{-1}(m|m+1) \quad (5.38)$$

Then using (5.4) we arrive at

$$\begin{aligned} \overline{S}(m|m) &= [F(m+1)P(m+1|m+1)F^T(m+1) \\ &+ G(m+1)\tilde{Q}(m+1)G^T(m+1)]^{-1} \\ &= [\overline{S}^{-1}(m|m+1) + G(m+1)\tilde{Q}(m+1)G^T(m+1)]^{-1} \quad (5.39) \end{aligned}$$

where the the last equality results from the substitution of (5.35). Also, by substituting (5.34) into (5.35) and collecting terms we obtain

$$\begin{aligned} \overline{S}(m|m+1) &= 2F^{-T}(m+1)\overline{S}(m+1|m+1)F^{-1}(m+1) \\ &+ F^{-T}(m+1)C^T(m)R^{-1}(m)C(m)F^{-1}(m+1) \\ &- F^{-T}(m+1)P_x^{-1}(m)F^{-1}(m+1) \quad (5.40) \end{aligned}$$

Now we prove by induction that for all m $S^*(m|m) \geq \overline{S}(m|m)$. 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 (5.40), (5.36), and the fact that $F^{-T}(m+1)P_x^{-1}(m)F^{-1}(m+1) \geq 0$ we get that

$$S^{*^{-1}}(i|i+1) \leq \overline{S}^{-1}(i|i) \quad (5.41)$$

Substituting (5.37) and (5.39) into (5.41) and canceling terms we arrive at $S^{*^{-1}}(i|i) \leq \overline{S}^{-1}(i|i)$, i.e. $S^*(i|i) \geq \overline{S}(i|i)$.

□

Theorem 5.3 *Suppose that there exists $\alpha, \gamma, M_0 > 0$ 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 (5.36), (5.37) satisfies $S^*(m|m) \leq NI$. for some $N > 0$ if the pair $(\tilde{Q}^{\frac{T}{2}}(m)G^T(m), F^{-T}(m))$ is bounded and uniformly observable. However, by standard duality results and the boundedness of F, this is equivalent to the boundedness and uniform reachability of $(F(m), G(m)\tilde{Q}^{\frac{1}{2}}(m))$, which in turn are equivalent to (5.19), (5.21) . Then from Lemma 5.2 we conclude that $\overline{S}(m|m) \leq NI$.

Thus (5.34) together with the boundedness assumption yields

$$P^{-1}(m|m) \leq L^{-1}I \tag{5.42}$$

Using analogous arguments we can derive a lower bound for $P_{ML}(m|m)$. Note that with the following definitions for $\overline{S}$ and (5.36), (5.37) where the matrices F and $\tilde{Q}$ are replaced with the matrices A^{-1} and I, respectively, Lemma 5.2 still applies.

$$\overline{S}(m|m) \triangleq \frac{1}{2}(P_{ML}^{-1}(m|m) - C^T(m)R^{-1}(m)C(m)) \tag{5.43}$$

$$\overline{S}(m|m+1) \triangleq A^T(m+1)P_{ML}^{-1}(m+1|m+1)A(m+1) \tag{5.44}$$

Using the same argument as in the proof of Theorem 5.3 with our current definitions for $\overline{S}$ we get that

$$\frac{1}{2}(P_{ML}^{-1}(m|m) - C^T(m)R^{-1}(m)C(m)) \leq NI \tag{5.45}$$

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

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

5.2 Filter Stability

We are now in a position to analyze the internal stability of the ML filter error dynamics (5.10) , i.e. using (5.15) we are interested in investigating the asymptotic stability of the autonomous error dynamics

$$\xi(t) = P_{ML}(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t)+1)A^{-1}(m(t)+1)[\xi(t\alpha)+\xi(t\beta)] \quad (5.46)$$

In particular, we have the following

Theorem 5.5 *Suppose that (5.19)-(5.22) are satisfied. Then, the ML error dynamics (5.10), or equivalently (5.46), are l_2-exponentially stable.*

Proof

Based on the analysis in Section 4.2, we see that we wish to show that the following causal system is stable in the standard sense:

$$z(m) = P_{ML}(m|m)P_{ML}^{-1}(m|m+-1)\sqrt{2}A^{-1}(m+1)z(m+1) \quad (5.47)$$

The remainder of the analysis follows the line of reasoning used in [3]. Specifically, thanks to Theorems 5.2 and 5.4, $P_{ML}(m|m)$ is bounded above and below by positive definite matrices. Thus we can define the following Lyapunov function.

$$V(z(m),m) \stackrel{\Delta}{=} z^T(m)P_{ML}^{-1}(m|m)z(m) \quad (5.48)$$

Let us also define the following quantity.

$$\begin{aligned} \tilde{z}(m) &\stackrel{\Delta}{=} \sqrt{2}A^{-1}(m+1)z(m+1) \\ &= P_{ML}(m|m+1)P_{ML}^{-1}(m|m)z(m) \end{aligned} \quad (5.49)$$

Substituting (5.12) into (5.48) followed by algebraic manipulations, one gets

$$\begin{aligned} V(z(m),m) &= z^T(m)(2P_{ML}^{-1}(m|m+1)+C^T(m)R^{-1}(m)C(m))z(m) \\ &= 2z^T(m)(P_{ML}^{-1}(m|m)-2P_{ML}^{-1}(m|m+1))z(m)-z^T(m)C^T(m)R^{-1}(m)C(m))z(m) \end{aligned}$$

$$
\begin{aligned}
&+ \quad z^T(m)(2P_{ML}^{-1}(m|m+1))z(m) \\
&+ \quad \frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m+1)\frac{\tilde{z}(m)}{\sqrt{2}} - \frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m+1)\frac{\tilde{z}(m)}{\sqrt{2}} \\
&= \quad -(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}})^T P_{ML}^{-1}(m|m+1)(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}}) \\
&- \quad z^T(m)C^T(m)R^{-1}(m)C(m)z(m) + \frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m+1)\frac{\tilde{z}(m)}{\sqrt{2}} \qquad (5.50)
\end{aligned}
$$

However, using (5.49) we have that

$$
\begin{aligned}
\frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m+1)\frac{\tilde{z}(m)}{\sqrt{2}} &= z^T(m+1)A^{-T}(m+1)P_{ML}^{-1}(m|m+1)A^{-1}(m+1)z(m+1) \\
&\leq z^T(m+1)P_{ML}^{-1}(m+1|m+1)z(m+1) \\
&= V(z(m+1), m+1) \qquad (5.51)
\end{aligned}
$$

where the inequality follows from (5.11) and application of the matrix inversion lemma. Thus it follows that

$$
\begin{aligned}
V(z(m),m) - V(z(m+1),m+1) &\leq -(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}})^T P_{ML}^{-1}(m|m+1)(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}}) \\
&- z^T(m)C^T(m)R^{-1}(m)C(m)z(m) \qquad (5.52)
\end{aligned}
$$

Stability follows from (5.52) under the condition of uniform observability of the pair $(R^{-\frac{1}{2}}(m)C(m), A^{-1}(m))$

□

Let us now examine the full estimation error after incorporating prior statistics. It is straightforward to see that

$$\tilde{x}(t|t) = P(m(t)|m(t))(P_{ML}^{-1}(m(t)|m(t))\tilde{x}_{ML}(t|t) + P_x^{-1}(m(t))x(t)) \qquad (5.53)$$

Thus we can view $\tilde{x}(t|t)$ as 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 first that since $P(m|m) \leq P_{ML}(m|m)$

$$\|P(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t))\tilde{x}_{ML}(t|t)\| \leq \|\tilde{x}_{ML}(t|t)\| \qquad (5.54)$$

and we already have the stability of the $\tilde{x}_{ML}(t|t)$ dynamics from Theorem 5.5. Turning to the second term in (5.53), 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 (5.53) is bounded. Finally, our analysis also allows us to conclude that the full, driven $\tilde{x}_{ML}(t|t)$ dynamics (5.10) are bounded-input, bounded-output stable from inputs w and v to output $\tilde{x}_{ML}(t|t)$.

5.3 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 5.6 *Consider the following system defined on a tree.*

$$x(t) = Ax(t\overline{\gamma}) + Bw(t) \tag{5.55}$$

$$y(t) = Cx(t) + v(t) \tag{5.56}$$

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

$$\begin{aligned} \overline{P}_\infty &= \frac{1}{2}A^{-1}\overline{P}_\infty A^{-T} + \frac{1}{2}GG^T \\ &- K_\infty(\frac{1}{2}CA^{-1}\overline{P}_\infty A^{-T}C^T + \frac{1}{2}CGG^TC^T + R)K_\infty^T \end{aligned} \tag{5.57}$$

where

$$K_\infty = \overline{P}_\infty C^T R^{-1} \tag{5.58}$$

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

$$e(t) = \frac{1}{2}(I - K_\infty C)A^{-1}(e(t\alpha) + e(t\beta)) \tag{5.59}$$

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

Proof

Note first that the reachability of (A, B) and observability of (C, A) are equivalent to the reachability of (A^{-1}, G) and observability of $(R^{-\frac{1}{2}}C, A^{-1})$, respectively.

Convergence of $P_{ML}(m|m)$

This will be established if we can show that a) $P_{ML}(m|m)$ is monotone-nonincreasing as $m \to -\infty$ and b) $P_{ML}(m|m)$ 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 [8]. Specifically, let $P(m; m')$ denote the solution to the scale-invariant version of the ML Riccati equation (5.11), (5.12) initialized at m' with[1] $P(m'; m') = \infty$. That is $P(m; m')$ equals $P_{ML}(m|m)$ if the coarse level at which we begin is m'. Since the parameters of (5.11), (5.12) are constant, we immediately see that

$$P(m; m') = P(m - m') \tag{5.60}$$

so that the monotonicity result we wish to show is equivalent to showing that if $m_1 > m_2$, then

$$P(m; m_1) \leq P(m; m_2) \text{ for all } m \leq m_2 \tag{5.61}$$

However, the scale-invariant Riccati equation certainly preserves positive definite orderings so that the inequality in (5.61) holds for $m = m_0$, then it must hold for all $m \leq m_0$. However at $m = m_2$, $P(m_2; m_1) \leq \infty = P(m_2; m_2)$, so that (5.61) is in fact true.

Having established the convergence of $P_{ML}(m|m)$, let us denote the limit as follows.

$$\lim_{m \to \infty} P_{ML}(m|m) \triangleq \overline{P}_\infty \tag{5.62}$$

[1]To be precise here we should use the information form of (5.11), (5.12) (see Appendices A,B). However, thanks to observability and reachability, for $|m - m'|$ sufficiently large $P(m; m')$ is well-defined and invertible. Since we are interested in the asymptotic behavior as $m - m' \to -\infty$, the argument given above is valid.

It is straightforward to see that $\overline{P}_\infty$ must satisfy (5.57), 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.

Exponential Stability of $\frac{1}{2}(I - K_\infty C)A^{-1}$

What we need to show is that if $\overline{P}_\infty$ is any positive definite solution to (5.57), then each eigenvalue, λ, of $\frac{\sqrt{2}}{2}(I-K_\infty C)A^{-1}$ has magnitude less than 1, where K_∞ is given by (5.58). The approach is a variation of the proof for the standard Riccati equations [8]. Specifically, some algebra shows that we can rewrite the Riccati equation (5.57) in the following form:

$$\overline{P}_\infty = [\frac{\sqrt{2}}{2}(I-K_\infty C)A^{-1}]P_\infty[\frac{\sqrt{2}}{2}(I-K_\infty C)A^{-1}]^T+\frac{1}{2}(I-K_\infty C)GG^T(I-K_\infty C)^T+K_\infty RK_\infty^T \tag{5.63}$$

Suppose that there exists an eigenvalue with $|\lambda| \geq 1$. Then letting x be the associated eigenvector of $[\frac{\sqrt{2}}{2}(I - K_\infty C)A^{-1}]^T$, we see that

$$x^H\overline{P}_\infty x = |\lambda|^2 x^H P_\infty x + |\lambda|^2 x^H BB^T x + x^H K_\infty RK_\infty^T x \tag{5.64}$$

where x^H is the conjugate transpose of x and we have used the fact that $G = A^{-1}B$. Since $\overline{P}_\infty > 0$ and $|\lambda| > 1$, we can conclude from (5.64) that $B^T x = 0$ and $K_\infty^T x = 0$, but the latter of these implies that $\frac{\sqrt{2}}{2}A^{-T}x = \lambda x$. That is, we have a vector x so that

$$x^H A^{-1} = \sqrt{2}\lambda^H x, \qquad x^H B = 0 \tag{5.65}$$

which implies that (A^{-1}, B) is not a reachable pair which in turn contradicts the assumption that $(A^{-1}, G) = (A^{-1}, A^{-1}B)$, or, equivalently, (A, B) is reachable.

Uniqueness of $\overline{P}_\infty$

Consider P_1 and P_2, both of which are positive definite and satisfy (5.57). Thus, for $i = 1, 2$

$$\begin{aligned} P_i &= \frac{1}{2}A^{-1}P_iA^{-T} + \frac{1}{2}GG^T \\ &- K_i(\frac{1}{2}CA^{-1}P_iA^{-T}C^T + \frac{1}{2}CGG^TC^T + R)K_i^T \end{aligned} \tag{5.66}$$

$$K_i = P_iC^TR^{-1} \tag{5.67}$$

Subtracting (5.66) with $i = 2$ from (5.66) with $i = 1$ yields

$$\begin{aligned} P_1 - P_2 &= \frac{\sqrt{2}}{2}(I - K_1C)A^{-1}(P_1 - P_2)(\frac{\sqrt{2}}{2}(I - K_1C)A^{-1})^T \\ &+ \Delta \end{aligned} \tag{5.68}$$

where

$$\Delta = (K_1 - K_2)[\frac{1}{2}CA^{-1}P_2A^{-T}C^T + \frac{1}{2}CGG^TC^T + R](K_1 - K_2)^T > 0 \tag{5.69}$$

Note that we have established the fact that $\frac{\sqrt{2}}{2}(I - K_1C)A^{-1}$ has eigenvalues within the unit circle. From standard system theory this tells us that $P_1 - P_2 \geq 0$. Reversing indices yields $P_2 - P_1 \geq 0$, proving uniqueness.

□

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

$$P(m|m) = [P_{ML}^{-1}(m|m) + P_x^{-1}(m)]^{-1} \tag{5.70}$$

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 (5.70). 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 $[\overline{P}_\infty^{-1} + S_x]^{-1}$.

6 Conclusions

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

[2]The two extreme cases being when A is stable, so that $P_x^{-1}(m) \to P_x^{-1}$ where P_x is the positive-definite solution of (2.6), and when A^{-1} is stable, in which case $P_x^{-1}(m) \to 0$.

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.

The major emphasis of this paper is 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 multiscale Kalman filter error dynamics and Riccati equation. Specifically, as we have seen, the structure of the dyadic tree leads to significant differences in these system–theoretic concepts as compared to their counterparts for standard state–space models. Using these concepts, we have determined reconstructibility and controllability conditions under which the solution to the Riccati equation is bounded above and below, the Kalman filter error dynamics are asymptotically stable, and, in the constant-parameter case, the Riccati equation solution converges to a unique, steady-state solution.

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 [2] on multiresolution realization theory 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, and it is our hope that our work will help to stimulate activity in this fascinating and important area.

A Appendix

In this appendix we derive several results related to the ML filter described in Section 3. The first is to show that $\hat{x}_{ML}(t|t)$ as defined in this section is indeed the ML estimate of $x(t)$ based on Y_t and also that it satisfies (3.9) - (3.10). To do this, we start by writing

$$Y_t = \mathcal{H}_t x(t) + \theta(t) \tag{A.1}$$

where $\theta(t)$ is a zero-mean noise vector constructed from process and measurement noises in the subtree under t using (2.1), (2.2). The ML estimate of $x(t)$ based on (A.1) is precisely $\hat{x}_{ML}(t|t)$, while, using standard ML estimation results [4] $\hat{x}(t|t)$ is the ML estimate of $x(t)$ based on (A.1) together with one additional "measurement".

$$0 = x(t) = \eta(t) \tag{A.2}$$

where $\eta(t)$ is a zero-mean noise vector, independent of $\theta(t)$ and with covariance $P_x(t)$. From this it is straightforward to verify (3.9)- (3.10).

To verify the recursive formulae (3.1)- (3.8) note 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".

$$\hat{x}_{ML}(t|t) = A(t)x(t\overline{\gamma}) + w(t) + \tilde{x}_{ML}(t|t) \tag{A.3}$$

where the estimation error $\tilde{x}_{ML}(t|t)$ is zero-mean, independent of $w(t)$, and with covariance $P_{ML}(t|t)$. Eqs. (3.5)- (3.6) follow directly from this. The fusion step (3.7), (3.8) then follows directly from standard ML 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 $t\beta$. Similarly the update step (3.1)- (3.4) follows from the standard result on incorporating a new, independent measurement (namely (2.1)).

The second result to be derived is the equivalence of (2.23) - (2.25) and (3.11) - (3.13). We begin with the equivalence of (2.24) and (3.13). That is, using (2.8), (3.11) and its counterpart for $P(t\overline{\gamma}|t)$ we wish to verify that

$$[P_{ML}^{-1}(t|t)+P_x^{-1}(t)]P_x^{-1}(t)A(t)P_x(t\overline{\gamma})[P_{ML}^{-1}(t\overline{\gamma}|t)+P_x^{-1}(t)] = P_{ML}(t|t)A^{-T}(t)P_{ML}(t\overline{\gamma}|t) \tag{A.4}$$

Algebraic manipulation of this relationship yields the equivalent form

$$P_{ML}(t\overline{\gamma}|t) = A^{-1}(t)P_{ML}(t|t)A^{-T}(t) + A^{-1}(t)P_x(t)A^{-T}(t) - P_x(t\overline{\gamma}) \tag{A.5}$$

Eqs. (2.4) and (3.6) then verify this equality. Secondly, to verify the equivalence of (2.23) and (3.11) we must show that

$$(I - J(t)F(t))\hat{x}(t|t) = (I - J(t)A^{-1}(t))\hat{x}_{ML}(t|t) \tag{A.6}$$

(where we have expressed one-step predicted estimates in terms of updated estimates.) Using (3.9), we must show that

$$(I - J(t)F(t))(P_{ML}^{-1}(t|t) + P_x^{-1}(t))^{-1}P_{ML}^{-1}(t|t) = I - J(t)A^{-1}(t) \tag{A.7}$$

Rearranging and using (2.8) and (3.13), we find that this is equivalent to (A.5), which verifies (A.6). Next, we must verify the equivalence of (2.25) and (3.12). i.e. we must show that

$$P(t|t) - J(t)P(t\overline{\gamma}|t)J^T(t) = P_{ML}(t|t) - J(t)P_{ML}(t\overline{\gamma}|t)J^T(t) \tag{A.8}$$

Again algebraic manipulations reduce (A.8) to (A.5), finishing this verification.

Finally, straightforward algebraic manipulations on (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.

$$S(t|t) = P_{ML}^{-1}(t|t), S(t|t+) = P_{ML}^{-1}(t|t+), \text{ etc.} \tag{A.9}$$

$$\hat{z}(t|t) = S(t|t)\hat{x}_{ML}(t|t),\ \hat{z}(t|t+) = S(t|t+)\hat{x}_{ML}(t|t+), \text{ etc.} \tag{A.10}$$

Then, we have the following algorithm

$$\hat{z}(t|t) = \hat{z}(t|t+) + C^T(t)R^{-1}(t)y(t) \tag{A.11}$$

$$\hat{z}(t\overline{\gamma}|t) = J^T(t)\hat{z}(t|t) \tag{A.12}$$

$$\hat{z}(t|t+) = \hat{z}(t|t\alpha) + \hat{z}(t|t\beta) \tag{A.13}$$

where the recursions for the inverse covariances and a corresponding equivalent expression for $J(t)$ are:

$$S(t|t) = S(t|t+) + C^T(t)R^{-1}(t)C(t) \tag{A.14}$$

$$J(t) = \{I - B(t)[B^T(t)S(t|t)B(t) + I]^{-1}B^T(t)S(t|t)\}A(t) \tag{A.15}$$

$$S(t\overline{\gamma}|t) = J^T(t)S(t|t)A(t) \tag{A.16}$$

$$S(t|t+) = S(t|t\alpha) + S(t|t\beta) \tag{A.17}$$

Note in particular the simple form of the fusion calculations (A.12) , (A.17), emphasizing the fact that independent sets of information are being combined. Also, as indicated in Section 3, this algorithm, is well-defined when S is singular, i.e. when insufficient information has been collected for x to be estimable. In particular, the initialization of the ML algorithm at the finest level is given by

$$\hat{z}(t|t+) = 0 \ , \ S(t|t+) = 0 \text{ for all t such that m(t)=M} \tag{A.18}$$

In addition, further algebra yields the corresponding version of the smoothing step (3.11) - (3.12), using only the information filter quantities calculated during the upward sweep:

$$\hat{x}_s(t) = J(t)\hat{x}_s(t\overline{\gamma}) + J(t)A^{-1}(t)B(t)B^T(t)\hat{z}(t|t) \tag{A.19}$$

$$P_s(t) = J(t)P_s(t\overline{\gamma})J^T(t) + J(t)A^{-1}(t)B(t)B^T(t)J^T(t) \tag{A.20}$$

where this is initialized at the top of the tree with

$$\hat{x}_s(0) = P_s(0)\hat{z}(0|0) \tag{A.21}$$

$$P_s(0) = [S(0|0) + P_x^{-1}(0)]^{-1} \tag{A.22}$$

B Appendix

In this appendix we describe an alternate derivation of the RTS smoothing algorithm by introducing the Hamiltonian form of the smoothing equations on the tree. For simplicity in exposition and notation we focus here exclusively on the constant parameter case. The extension to the general case is straightforward.

Specifically, consider the model

$$x(t) = Ax(t\overline{\gamma}) + Bw(t) \tag{B.1}$$

$$y(t) = Cx(t) + v(t) \tag{B.2}$$

where w and v are white-noise processes with variances I and R respectively, and (B.1), (B.2) are defined on an M-level tree, i.e. $m(t) = 0, \ldots, M$, with a single root node which we denote by 0. The Hamiltonian form of the smoothing equations can be derived in several ways: using the complementary model construction as described, for example, in [6] or by examining the minimization problem in computing the $x(t)$-trajectory that has maximum posterior probability given the data, the prior statistics and those of the noises, and the dynamic constraint (B.1). We follow the latter approach here. 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 trajectory $\hat{x}_s(t)$ is obtained by minimizing the following

Hamiltonian

$$H = \sum_t \frac{1}{2}(y(t) - Cx(t))^T R^{-1}(y(t) - Cx(t)) + \sum_{t\neq 0} \frac{1}{2} w^T(t)w(t) \qquad \text{(B.3)}$$
$$+ \frac{1}{2}x^T(0)P_x^{-1}(0)x(0) + \sum_{t\neq 0} \lambda^T(t)(x(t) - Ax(t\overline{\gamma}) - Bw(t))$$

with respect to the state x, the noise w, and the Lagrange multiplier $\lambda^T(t)$.

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 $m(t) = 1, \ldots, M$:

$$\hat{\lambda}(t) = A^T[\hat{\lambda}(t\alpha) + \hat{\lambda}(t\beta)] - C^T R^{-1} C\hat{x}_s(t) + C^T R^{-1} y(t) \qquad \text{(B.4)}$$

$$\hat{x}_s(t) = A\hat{x}_s(t\overline{\gamma}) + BB^T\hat{\lambda}(t) \qquad \text{(B.5)}$$

and the boundary conditions[3]

$$\hat{x}_s(0) = [P_x(0) + C^T R^{-1} C]^{-1}\{A^T[\hat{\lambda}(0\alpha) + \hat{\lambda}(0\beta)] + C^T R^{-1} y(0)\} \qquad \text{(B.6)}$$

$$\hat{\lambda}(t) = 0 \ , \ m(t) = M + 1 \qquad \text{(B.7)}$$

Let us note several points concerning these equations. First note that, as in the standard case, the dual dynamics for $\hat{\lambda}$ run in the opposite direction to the x-dynamics. In this case, thanks to the asymmetry of the tree, the dual dynamics (B.4) are in the form of fine-to-coarse dynamics which merge values as we progress up the tree (4.1). Secondly, by organizing the dynamics (B.5), (B.6) we can obtain the Hamiltonian form of the dynamics for $m(t) = 1, \ldots, M$:

$$\mathcal{A}\begin{bmatrix} \hat{x}_s \\ \hat{\lambda} \end{bmatrix}_\alpha + \Theta_t \begin{bmatrix} \hat{x}_s \\ \hat{\lambda} \end{bmatrix}_{t\alpha} + \Theta_\beta \begin{bmatrix} \hat{x}_s \\ \hat{\lambda} \end{bmatrix}_{t\beta} = \begin{bmatrix} 0 \\ 0 \\ C^T R^{-1} y(t) \end{bmatrix} \qquad \text{(B.8)}$$

[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.

with boundary conditions given by (B.6) and (B.7). Also, the matrices in (B.8) are

$$\mathcal{A} = \begin{bmatrix} -A & 0 \\ -A & 0 \\ C^T R^{-1} C & I \end{bmatrix} \tag{B.9}$$

$$\Theta_\alpha = \begin{bmatrix} I & -BB^T \\ 0 & 0 \\ 0 & -A^T \end{bmatrix} \tag{B.10}$$

$$\Theta_\beta = \begin{bmatrix} 0 & 0 \\ I & -BB^T \\ 0 & -A^T \end{bmatrix} \tag{B.11}$$

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.8) 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.8) 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], [7]. However, we can triangularize these dynamics and boundary conditions to obtain an RTS algorithm.

Specifically, drawing inspiration from [6],[7] consider a time-varying transformation of the following form

$$\begin{bmatrix} x^u \\ \hat{x} \end{bmatrix}_t = T_{m(t)} \begin{bmatrix} \hat{x} \\ \hat{\lambda} \end{bmatrix}_t \tag{B.12}$$

where

$$T_m = \begin{bmatrix} \Gamma_m & I \\ I & 0 \end{bmatrix} \tag{B.13}$$

With respect to the transformed variables x^u and $\hat{x}$ we now 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.8) on the left by an invertible matrix, $S_{m(t)}$, without losing

information. By doing so, we wish to transform the dynamics into the following structure.

$$S_{m(t)}\mathcal{A}T^{-1}_{m(t)}\begin{bmatrix} x^u \\ \hat{x} \end{bmatrix}_t + S_{m(t)}\Theta_\alpha T^{-1}_{m(t)+1}\begin{bmatrix} x^u \\ \hat{x} \end{bmatrix}_{\alpha t} + S_{m(t)}\Theta_\beta T^{-1}_{m(t)+1}\begin{bmatrix} x^u \\ \hat{x} \end{bmatrix}_{\beta t} = \begin{bmatrix} C^T R^{-1} y(t) \\ 0 \\ 0 \end{bmatrix} \tag{B.14}$$

where

$$S_m = \begin{bmatrix} -P_m^{-1}A^{-1} & -P_m^{-1}A^{-1} & I \\ 0 & I & 0 \\ I & 0 & 0 \end{bmatrix} \tag{B.15}$$

$$S_m\mathcal{A}T_m^{-1} = \begin{bmatrix} I & 0 \\ L_1 & L_2 \\ L_3 & L_4 \end{bmatrix} \tag{B.16}$$

$$S_m\Theta_\alpha T^{-1}_{m+1} = \begin{bmatrix} F_{m+1} & 0 \\ 0 & 0 \\ N & G_{m+1} \end{bmatrix} \tag{B.17}$$

$$S_m\Theta_\beta T^{-1}_{m+1} = \begin{bmatrix} F_{m+1} & 0 \\ N & G_{m+1} \\ 0 & 0 \end{bmatrix} \tag{B.18}$$

Substituting the forms of (B.13), (B.15) into (B.16)- (B.18) yields the following constraints for $L_1 - L_4$, N, F_m, G_m, P_m, and Γ_m:

$$L_1 = L_3 = 0\ ,\ L_2 = L_4 = -A \tag{B.19}$$

$$N = -BB^T \tag{B.20}$$

$$\Gamma_m = 2P_m^{-1} + C^T R^{-1} C \tag{B.21}$$

$$F_{m+1}\Gamma_{m+1} = -P_m^{-1}A^{-1} \tag{B.22}$$

$$F_{m+1} = P_m^{-1} A^{-1} B B^T - A^T \tag{B.23}$$

$$G_{m+1} = I + B B^T \Gamma_{m+1} \tag{B.24}$$

Combining (B.21)- (B.23) yields a recursion for P_m^{-1}:

$$P_m^{-1} = (A^{-1}(2P_{m+1}^{-1} + C^T R^{-1} C)^{-1} A^{-T} + A^{-1} B B^T A^{-T})^{-1} \tag{B.25}$$

which is exactly the same as the information form of the ML Riccati equation (5.11), (5.12) in the constant parameter case– i.e. if we set

$$P_m^{-1} = P_{ML}^{-1}(m|m+1) \tag{B.26}$$

then P_m^{-1} satisfies (B.25) together with the boundary condition

$$P_M^{-1} = 0 \tag{B.27}$$

Furthermore in this case from (B.21) we see that

$$\Gamma_m = P_{ML}^{-1}(m|m) \tag{B.28}$$

and, using these identifications plus (B.23), (B.24), yields

$$F_{m+1} = -P_m^{-1} A^{-1} \Gamma_{m+1}^{-1} = -P_{ML}^{-1}(m|m+1) A^{-1} P_{ML}(m+1|m+1) \tag{B.29}$$

$$G_{m+1} = A P_m A^T \Gamma_{m+1} = A P_{ML}(m|m+1) A^T P_{ML}^{-1}(m+1|m+1) \tag{B.30}$$

so that $F_m = -J^T(m)$ and $G_m = AJ^{-1}(m)$, where J is defined in Section 2.

Finally, using these expressions and the dynamics (B.14) - (B.18) yields the following algorithm. The filtering recursion is given by

$$x^u(t) = J^T(m(t)+1)[x^u(t\alpha) + x^u(t\beta)] + C^T R^{-1} y(t) \ , \ m(t) = 0, \ldots, M-1 \tag{B.31}$$

with initial conditions (using (B.12), (B.13), (B.21) for $m = M$, (B.27), and (B.7)

$$x^u(t) = C^T R^{-1} y(t) \ , \ m(t) = M \tag{B.32}$$

Using the boundary conditions at $t = 0$ yields the initial condition

$$\hat{x}_s(0) = [\Gamma_0 + P_x^{-1}(0)]^{-1} x^u(0) \tag{B.33}$$

for the downward recursion, which we directly have from (B.14)- (B.18):

$$\hat{x}_s(t) = J(m(t))\hat{x}_s(t) + J(m(t))A^{-1}BB^T x^u(t) \tag{B.34}$$

Finally, comparing (B.31)- (B.34) to (A.8)- (A.10), (A.16), (A.18), (A.19), 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", submitted for publication.

[2] A. Benveniste, R. Nikoukhah, and A.S. Willsky, "Multiscale System Theory", Proceedings of the 29th IEEE Conference on Decision and Control, Honolulu, HI, December 1990.

[3] J. Deyst, C. Price, "Conditions for Asymptotic Stability of the Discrete, Minimum Variance, Linear, Estimator," *IEEE Trans. on Automatic Control*, vol. 13, pp. 702– 705, Dec. 1968.

[4] R. Nikoukhah, A.S. Willsky and B.C. Levy, "Kalman Filtering and Riccati Equations for Descriptor Systems", submitted to it IEEE Trans. on Automatic Control.

[5] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems," AIAA Journal, Vol. 3, No. 8, Aug. 1965, pp. 1445-1450.

[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" and Part II: 1-D Smoothing Problems," *IEEE Trans. on Automatic Control*, Vol 29, Sept. 1984, pp. 803-821.

[7] T. Kailath and L. Ljung, "Two Filter Smoothing Formulas by Diagonalization of the Hamiltonian Equations," *Int'l J. Control*, Vol. 36, 1982, pp. 663-673.

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

[9] K.C. Chou, "A Stochastic Modeling Approach to Multiscale Signal Processing," M.I.T. Department of Electrical Engineering and Computer Science, Ph.D. Thesis, May 1991.