Hierarchical Adaptive Kalman Filtering for Interplanetary Orbit Determination

WASSIM S. CHAER, Member, IEEE
SDT

ROBERT H. BISHOP, Member, IEEE

JOYDEEP GHOSH, Member, IEEE
University of Texas at Austin

A modular and flexible approach to adaptive Kalman filtering has recently been introduced using the framework of a mixture-of-experts regulated by a gating network. Each expert is a Kalman filter modeled with a different realization of the unknown system parameters. The unknown or uncertain parameters can include elements of the state transition matrix, observation mapping matrix, process noise covariance matrix, and measurement noise covariance matrix. The gating network performs on-line adaptation of the weights given to individual filters based on performance. The mixture-of-experts approach is extended here to a hierarchical architecture which involves multiple levels of gating. The proposed architecture provides a multilevel hypothesis testing capability. The utility of the hierarchical architecture is illustrated via the problem of interplanetary navigation (Mars Pathfinder) using simulated radiometric data. It serves as a useful tool for assisting navigation teams in the process of selecting the parameters of the navigational filter over various operating regimes. It is shown that the scheme has the capability of detecting changes in the system parameters and switching filters appropriately for optimal performance. Furthermore, the expectation-maximization (EM) algorithm is shown to be applicable in the proposed framework.

I. INTRODUCTION

For a Kalman filter to operate in an optimal fashion, all filter parameters should be known exactly [1, 2]. Incorrect selection of these parameters will result in large estimation errors or divergence [3, 4]. This has motivated the design of a bank of Kalman filters with each filter using different parameters. An example is the classic Magill filter bank wherein the most suitable filter is selected through a sequential maximum *a posteriori* estimation procedure [5]. We have recently shown that the mixture-of-experts framework provides a more flexible and powerful alternative to traditional filter bank approaches [6]. The proposed solution was motivated by modular networks [7–9] whose general structure is shown in Fig. 1. In the suggested scheme, each module or expert is a particular Kalman filter realization. That is, each Kalman filter is modeled with a different realization of the unknown or uncertain parameter vector. Typically, the parameter vector contains the elements of the process noise and the measurement noise covariance matrices. The parameter vector can also contain various unknown or uncertain elements of the state transition and measurement mapping matrices. A gating network is used to regulate the Kalman filter bank by adaptively assigning weights to the individual filters based on their performance. It has been shown that the modular architecture regulated by a gating network has the ability to 1) select the best filter realization as measurements are processed, 2) rapidly change selection in case of a change in the operating regime, 3) adapt individual filters to better match incoming data, and 4) avoid numerical instabilities (usually encountered with Magill's scheme, for example). A complete description of the suggested approach, mathematical derivations, adaptation of the parameter vectors, and a number of illustrative examples can be found in [6].

In this work, the adaptive Kalman filtering architecture in [6] is extended to a hierarchical structure which involves multiple levels of gating [8, 10]. The proposed hierarchy has the ability to detect parameter changes and allows the examination of parameter changes on an individual basis and in combination with other parameters. Also, the expectation-maximization (EM) algorithm [11, 12] is shown to be applicable to the proposed hierarchy of adaptive Kalman filter banks, provided off-line solutions (i.e., batch) are acceptable.

The mixture-of-experts hierarchical architecture provides a structured approach to conducting the necessary numerical experiments and to clarify the outcomes. The workings and power of the hierarchical adaptive Kalman filtering architecture are illustrated using the interplanetary orbit determination problem [13]. The available measurements are simulated radiometric (Doppler) generated by Deep Space

z : input vector
x_i : output of expert i
g_i : gating weight of expert i
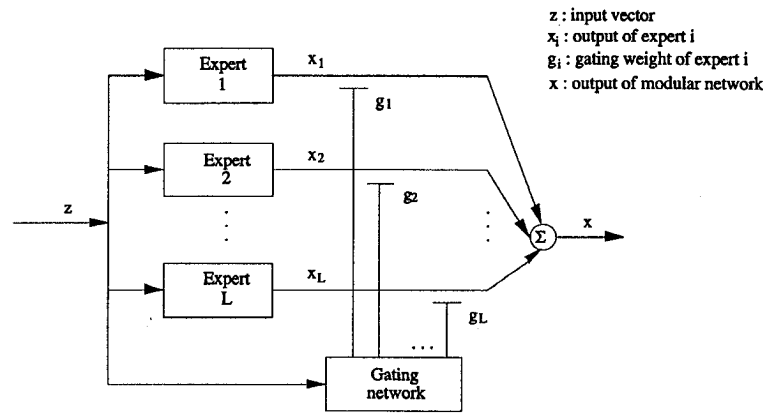x : output of modular network

Fig. 1.   General architecture of modular network.

Network stations. The spacecraft trajectory is modeled after the Mars Pathfinder mission [14]. The results for this problem show that the proposed hierarchy can be used to 1) identify individual parameter changes by using an on-line algorithm, such as gradient ascent, 2) divide the data set into statistically coherent subsets, and 3) find an "optimal" set of weights utilizing the EM algorithm. It can be envisioned that the proposed approach might be used by navigation teams seeking to determine the various unknown or uncertain parameters of the navigational filter. Generally, these parameters are selected using ad-hoc filter tuning methods.

II.   KALMAN FILTER

The general system model considered is represented by

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_k \mathbf{x}_k + \mathbf{w}_k$$
$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{1}$$

where $\mathbf{x}$ is the n-dimensional state vector, $\mathbf{\Phi}_k$ is the state transition matrix, $\mathbf{z}_k$ is the m-dimensional measurement vector, and $\mathbf{H}_k$ is the observation mapping matrix. The $\mathbf{w}_k$ and $\mathbf{v}_k$ vectors are assumed to be zero-mean, white sequences with known covariance matrices. The covariance matrices for the $\mathbf{w}_k$ and $\mathbf{v}_k$ vectors are given by

$$E[\mathbf{w}_k \mathbf{w}_i^T] = \mathbf{Q}_k \delta_{ki} \quad \text{and} \quad E[\mathbf{v}_k \mathbf{v}_i^T] = \mathbf{R}_k \delta_{ki}$$

where the Kronecker delta $\delta_{ki} = 0$ for $k \neq i$ and $\delta_{ki} = 1$ for $k = i$. We also assume that the two noises are uncorrelated as

$$E[\mathbf{w}_k \mathbf{v}_i^T] = \mathbf{0}, \quad \forall \ i,k.$$

The estimate of the state can be obtained sequentially with the Kalman filter. The filter equations, depicted in Fig. 2, are divided into an update and propagate parts and represent the standard recursive Kalman filter. The update equations are used to incorporate the new measurement while the propagate equations

Update equations:	$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \mathbf{r}_k$
	$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-$
Propagate equations:	$\hat{\mathbf{x}}_{k+1}^- = \mathbf{\Phi}_k \hat{\mathbf{x}}_k^+$
	$\mathbf{P}_{k+1}^- = \mathbf{\Phi}_k \mathbf{P}_k^+ \mathbf{\Phi}_k^T + \mathbf{Q}_k$
Kalman gain:	$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{W}_k^{-1}$
Residual:	$\mathbf{r}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-$
Residual covariance:	$\mathbf{W}_k = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k$

Fig. 2.   Kalman filter equations.

are used in between measurements. The "^" denotes the estimated state, "+" denotes after measurement incorporation, and "−" prior to measurement incorporation. The optimal Kalman filter requires an accurate knowledge of the process noise covariance matrix $\mathbf{Q}_k$, the measurement noise covariance matrix $\mathbf{R}_k$, and the parameters of $\mathbf{H}_k$ and $\mathbf{\Phi}_k$. The unknown or uncertain elements of these matrices are included in a parameter vector α. That is, the vector α is used in the remainder of this work to denote the uncertain filter parameters.

III.   HIERARCHICAL KALMAN FILTERING ARCHITECTURE

We propose an adaptive hierarchical Kalman filtering architecture. The suggested setup is shown in Fig. 3. It is a generalization of the framework proposed in [6] and follows the development in [8]. The architecture consists of K adaptive Kalman filter banks, each of which contains L Kalman filters. At time step t_k, the bank processes the measurement $\mathbf{z}_k$ and each filter (or expert) provides its updated state estimate $\hat{\mathbf{x}}_{ji,k}^+$, where $j = 1, 2, \ldots, L$ and $i = 1, 2, \ldots, K$. The state estimate of the ith bank is a weighted sum of the individual filter estimates in the bank with the gating network providing the weights $g_{j|i}$. The overall state estimate $\hat{\mathbf{x}}_{opt,k}$ is a weighted sum of the filter bank estimates, with the top-level gating network providing the weights g_i. In our design, all gating networks consist of a single layer of computational
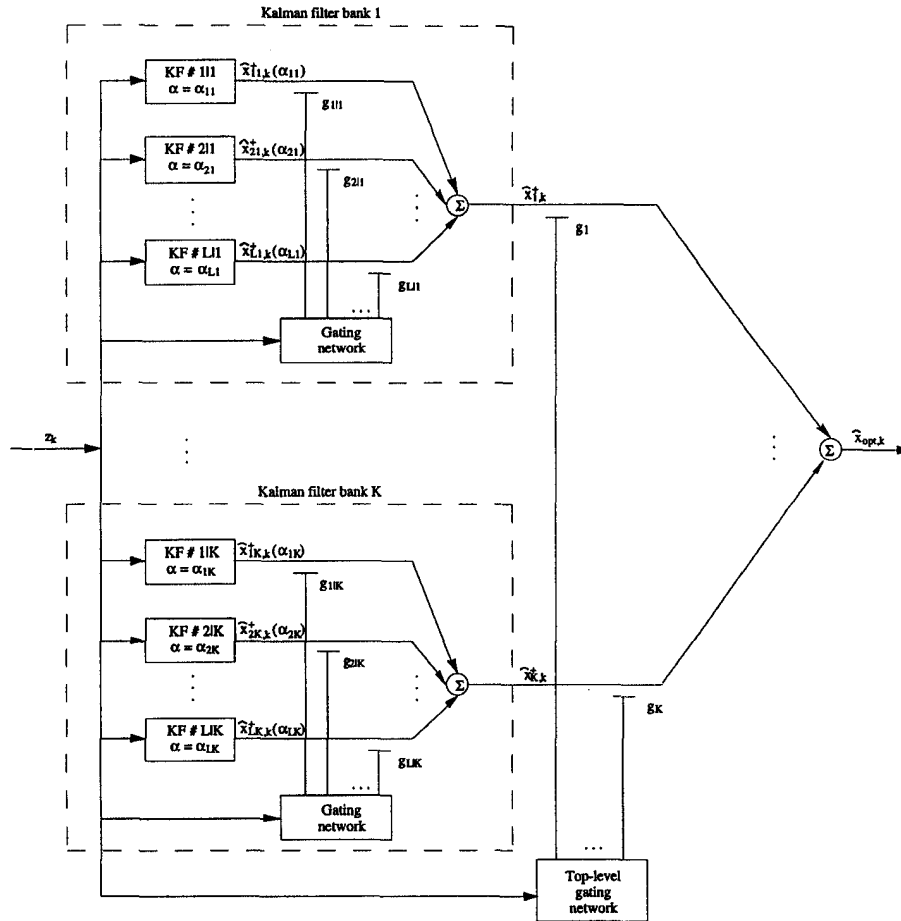
Fig. 3. Adaptive hierarchical Kalman filter banks.

cells or "neurons". The activation g_i of the ith output cell of the top-level gating network is defined using "softmax" [15, 16]:

$$g_i = \frac{e^{u_i}}{\sum_{j=1}^{K} e^{u_j}} \qquad (2)$$

where

$$u_i = \mathbf{z}_k^T \mathbf{a}_i \qquad (3)$$

is the weighted sum of the inputs applied to that cell and $\mathbf{a}_i$ is the weight vector of cell i, representing the ith bank in the hierarchical architecture. A schematic of the top-level gating network is shown in Fig. 4. In a similar fashion, we define the activation of the jth output neuron in the ith Kalman filter bank as

$$g_{j|i} = \frac{e^{u_{j|i}}}{\sum_{n=1}^{L} e^{u_{n|i}}} \qquad (4)$$

where

$$u_{j|i} = \mathbf{z}_k^T \mathbf{a}_{j|i} \qquad (5)$$

is the weighted sum of the inputs applied to this particular cell in the ith Kalman filter bank and $\mathbf{a}_j$ is the weight vector of cell $j \mid i$, representing the jth

filter in the ith bank. The use of softmax in (2) and (4) is motivated by the need to use a differentiable operator which is a generalization of a "winner takes all" strategy. This means that within a bank the best performing filter will be given a weight close to unity by the gating network which regulates the bank. Similarly, at the top-level the best performing bank will be given a weight close to unity by the top-level gating network.

By interpreting the gs as *prior* probabilities of the corresponding gated components [17] (i.e., probability given expert is chosen before using filter performance), the probability distribution of the next measurement, as estimated by the current hierarchical architecture can be written as

$$f(\mathbf{z}_k) = \sum_{i=1}^{K} g_i \sum_{j=1}^{L} f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i} \qquad (6)$$

where $f(\mathbf{z}_k \mid \alpha_{j|i})$ is the probability distribution indicated by the jth filter in the ith Kalman filter bank. As in [8], we let each component distribution be normal. The Gaussian distribution of the $j \mid i$th filter has mean $\mathbf{H}_k \hat{\mathbf{x}}_{ji,k}^-$ and covariance $\mathbf{W}_{ji,k}$. In other words, the distribution of the measurement vector $\mathbf{z}_k$ of the
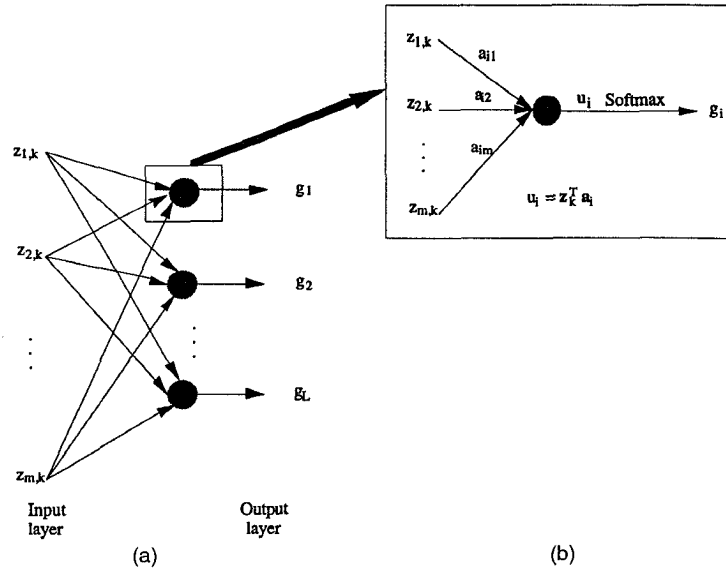
Fig. 4. Top-level gating network.

$j \mid i$th filter is given by

$$f(\mathbf{z}_k \mid \alpha_{j|i}) = \frac{1}{\sqrt{|2\pi \mathbf{W}_{ji,k}|}} \exp(\tfrac{1}{2}\mathbf{r}_{ji,k}^T \mathbf{W}_{ji,k}^{-1} \mathbf{r}_{ji,k}). \quad (7)$$

Learning consists of updating the state estimates $\hat{\mathbf{x}}_{ji,k}^-$ and the gating weights g_is and $g_{j|i}$s every time a measurement $\mathbf{z}_k$ is observed.

Equation (6) can be viewed as a likelihood function. It is better to work with the natural logarithm of $f(\mathbf{z}_k)$ rather than $f(\mathbf{z}_k)$ itself. We define the log-likelihood function as

$$l = \ln \sum_{i=1}^{K} g_i \sum_{j=1}^{L} f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i}. \quad (8)$$

To formulate the learning algorithm, it is first convenient to define h_i, the a posteriori probability associated with the ith Kalman filter bank, as

$$h_i = \frac{g_i \sum_{j=1}^{L} f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i}}{\sum_{i=1}^{K} g_i \sum_{j=1}^{L} f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i}} \quad (9)$$

and $h_{j|i}$, the a posteriori probability associated with the jth Kalman filter in the ith bank, as

$$h_{j|i} = \frac{f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i}}{\sum_{j=1}^{L} f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i}}. \quad (10)$$

To derive the update equations for the top-level gating network and the gating networks within the banks, we follow the same procedure described in [6]. Using $\mathbf{a}_i$ to denote the weight vector of the ith output cell of the top-level gating network and $\mathbf{a}_{j|i}$ to denote the weight of the jth output cell of the gating network in the ith bank, we find

$$\frac{\partial l}{\partial \mathbf{a}_i} = \frac{\partial l}{\partial u_i}\frac{\partial u_i}{\partial \mathbf{a}_i} = (h_i - g_i)\mathbf{z}_k \quad (11)$$

and

$$\frac{\partial l}{\partial \mathbf{a}_{j|i}} = \frac{\partial l}{\partial u_{j|i}}\frac{\partial u_{j|i}}{\partial \mathbf{a}_{j|i}} = h_i(h_{j|i} - g_{j|i})\mathbf{z}_k. \quad (12)$$

The modifications to the weight vectors $\mathbf{a}_i$ and $\mathbf{a}_{j|i}$ are then defined as

$$\Delta \mathbf{a}_i = \eta \frac{\partial l}{\partial \mathbf{a}_i} = \eta(h_i - g_i)\mathbf{z}_k \quad (13)$$

and

$$\Delta \mathbf{a}_{j|i} = \eta \frac{\partial l}{\partial \mathbf{a}_{j|i}} = \eta h_i(h_{j|i} - g_{j|i})\mathbf{z}_k. \quad (14)$$

Using (13) and (14), the update equations for the weight vectors are given by

$$\mathbf{a}_i \leftarrow \mathbf{a}_i + \eta(h_i - g_i)\mathbf{z}_k \quad (15)$$

$$\mathbf{a}_{j|i} \leftarrow \mathbf{a}_{j|i} + \eta h_i(h_{j|i} - g_{j|i})\mathbf{z}_k \quad (16)$$

where η is a learning-rate parameter. This parameter need not be the same for all gating networks, if desired. Note that the update equations are effectively instantaneous gradient ascent procedures which seek to maximize (8). In other words, the weight vectors $\mathbf{a}_i$ and $\mathbf{a}_{j|i}$ are updated every time a new measurement vector $\mathbf{z}_k$ is processed. The learning rules given by (15) and (16) can be used for real-time and post-processed data type applications. The weight vectors can be initialized either randomly (i.e., different) or with the same small values (i.e., so as not to overly favor any filter or bank initially).

A hierarchical structure with a small number of filters per bank can be exploited to study the effect of one particular parameter at a time. For example, in a hierarchy composed of two filter banks, we can make one bank nominal in process noise covariance matrix and the other one nominal in measurement noise covariance matrix. That is, all Kalman filters in the first bank operate with the same process noise

while having different measurement noise and vice versa for the second bank. This allows the study of the effects of each parameter in isolation as well as in combination with other parameters. The proposed hierarchy provides a multilevel hypothesis testing capability.

Other well-known multiple model estimation approaches are the Magill filter bank [5] and the interacting multiple model (IMM) algorithm [18–23]. IMM provides an improvement over Magill's scheme. The IMM algorithm assumes that 1) the unknown parameter vector takes a finite number of values, 2) according to a known transition probability matrix, and 3) requires a filter for the unknown parameter vector. Also, IMM is more computationally demanding than Magill's scheme but is numerically stable. Unlike the Magill scheme, the mixture-of-experts framework does not assume that the optimal filter is included in the bank. Moreover, it does not require the assumptions of IMM mentioned above, and is computationally faster but needs a tuning of the learning-rate parameter.

IV. APPLYING EM TO HIERARCHICAL KALMAN FILTERING ARCHITECTURE

The EM algorithm [11, 12] is an iterative maximum-likelihood estimation technique, where each iteration consists of 1) an estimation (E) step and 2) a maximization (M) step. The batch iterative nature of EM makes the algorithm suitable only for post-processed data type applications.

The use of EM in the hierarchical filtering structure follows the approach and nomenclature of [8] where EM was used in a hierarchical mixture-of-experts setup. For the hierarchical Kalman filtering architecture, the parameter vector Θ is composed of the gating network parameters $\mathbf{a}_i$ and $\mathbf{a}_{j|i}$ defined in the previous section. We assume that the observable data set $\mathcal{O}$ is of size N. The log-likelihood function of the data set is obtained by taking the log of the product of N densities of the form of (8), which yields the following log-likelihood:

$$l(\Theta;\mathcal{O}) = \sum_{k=1}^{N} \ln \sum_{i=1}^{K} g_{i,k} \sum_{j=1}^{L} g_{j|i,k} f(\mathbf{z}_k \mid \alpha_{j|i}) \quad (17)$$

where k indexes measurement collected at equal time intervals and $\mathcal{O}$ refers to the observable data which is termed as the "incomplete data" for obvious reasons as shown later in this section. Hence, (17) is called the incomplete likelihood function. Also, it is useful to define the joint posterior probability h_{ij} as the product of h_i and $h_{j|i}$ defined in (9) and (10), respectively. This is given by

$$h_{ij} = \frac{f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i} g_i}{\sum_{i=1}^{K} g_i \sum_{j=1}^{L} f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i}}. \quad (18)$$

The density functions, $f(\mathbf{z}_k \mid \alpha_{j|i})$, are indicative of the Kalman filters performance. In general, $f(\mathbf{z}_k \mid \alpha_{j|i})$ is consistently larger for the best filter realization.

To develop an EM algorithm for the hierarchical architecture, we must identify the "missing variables" so as to simplify the likelihood function $l(\Theta;\mathcal{O})$. As in [8], we postulate the indicator variables d_i and $d_{j|i}$ such that for each input data point, a unique d_i is equal to one, a unique $d_{j|i}$ is equal to one, and all other d_is, $d_{j|i}$s are zeros. The *unknown* indicator variables, denoted $\mathcal{M}$, labels the decisions in the probability model (i.e., Kalman filter bank). The indicator variable $d_{ij} = d_i d_{j|i}$ specifies the expert in the probability model (i.e., Kalman filter) that is performing best. Since $d_{i,j,k}$ is an indicator variable, we can specify a probability model that links the missing variables to the observable data. The model is written as

$$f(\mathbf{z}_k, d_{i,j,k} \mid \Theta) = g_{i,k} g_{j|i,k} f(\mathbf{z}_k \mid \alpha_{j|i})$$

$$= \prod_{i=1}^{K} \prod_{j=1}^{L} \{g_{i,k} g_{j|i,k} f(\mathbf{z}_k \mid \alpha_{j|i})\}^{d_{i,j,k}}. \quad (19)$$

The complete data likelihood, $l_c(\Theta;\mathcal{C})$, is obtained by taking the logarithm of the previous equation, yielding

$$l_c(\Theta;\mathcal{C}) = \sum_{k=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{L} d_{ij,k} \ln\{g_{i,k} g_{j|i,k} f(\mathbf{z}_k \mid \alpha_{j|i})\}$$

$$= \sum_{k=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{L} d_{ij,k}\{\ln g_{i,k} + \ln g_{j|i,k} + \ln f(\mathbf{z}_k \mid \alpha_{j|i})\}$$

$$(20)$$

where $\mathcal{C}$ denotes the "complete data" (composed of $\mathcal{O}$ and $\mathcal{M}$).

The E step of the EM algorithm is obtained by taking the expectation of the complete data likelihood and making use of the following facts:

$$f(\mathbf{z}_k \mid \Theta_p) P(d_{ij,k} = 1 \mid \mathbf{z}_k, \Theta_p)$$

$$= f(\mathbf{z}_k \mid d_{ij,k} = 1, \Theta_p) P(d_{ij,k} = 1 \mid \Theta_p)$$

and

$$E[d_{ij,k} \mid \mathcal{O}] = P(d_{ij,k} = 1 \mid \mathbf{z}_k, \Theta_p)$$

$$= \frac{f(\mathbf{z}_k \mid d_{ij,k} = 1, \Theta_p) P(d_{ij,k} = 1 \mid \Theta_p)}{f(\mathbf{z}_k \mid \Theta_p)}$$

$$= \frac{f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i,k} g_{i,k}}{\sum_{i=1}^{K} g_{i,k} \sum_{j=1}^{L} f(\mathbf{z}_k \mid \alpha_{j|i}) g_{j|i,k}}$$

$$= h_{ij,k}.$$

The expectation of $l_c(\Theta;\mathcal{C})$ is a deterministic function, $D(\Theta,\Theta_p)$, given by

$$D(\Theta,\Theta_p) = \sum_{k=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{L} h_{ij,k}\{\ln g_{i,k} + \ln g_{j|i,k} + \ln f(\mathbf{z}_k \mid \alpha_{j|i})\}$$

$$(21)$$

where Θ_p is the value of the parameters at iteration p.

The M step maximizes $D(\Theta, \Theta_p)$ with respect to the gating network parameters. Equation (21) indicates that the gating network parameters effect D only through the terms $h_{ij,k} \ln g_{i,k}$ and $h_{ij,k} \ln g_{j|i,k}$, which simplifies the M step into two maximization problems, given by

$$\mathbf{a}_{i,p+1} = \max_{\mathbf{a}_i} \sum_{k=1}^{N} \sum_{l=1}^{K} h_{l,k} \ln g_{l,k} \qquad (22)$$

and

$$\mathbf{a}_{j|i,p+1} = \max_{\mathbf{a}_{j|i}} \sum_{k=1}^{N} \sum_{l=1}^{K} h_{l,k} \sum_{m=1}^{L} h_{m|l,k} \ln g_{m|l,k}. \qquad (23)$$

The maximization in (22) can be solved efficiently by using iteratively reweighted least-squares (IRLS) [24]. The same is true for (23) which is a weighted maximum likelihood problem.

During an EM iteration, the effect of choosing a parameter value which increases the value of D is an increase in the incomplete likelihood [11]:

$$l(\Theta_{p+1}; \mathcal{O}) \geq l(\Theta_p; \mathcal{O}).$$

In other words, the likelihood l increases monotonically along the sequence of parameter estimates generated by an EM algorithm, implying convergence to a local maximum [11].

Note that the iterative process ends when no significant change is observed in the parameter values. In other words, the EM algorithm stops if the absolute difference (i.e., L_1 norm) between two consecutive iterations yields a set of parameters that differs by a prespecified convergence tolerance ϵ.

Algorithm Summary:

1) Using the current values of the gating networks parameters (i.e., $\mathbf{a}_i$s and $\mathbf{a}_{j|i}$s), process the measurement set and compute the posterior probabilities $h_{i,k}$ and $h_{j|i,k}$.

2) Solve an IRLS problem given the observations $\{(\mathbf{z}_k, h_{l,k})\}_{k=1}^{N}$ for the top-level gating network.

3) Solve a weighted IRLS problem given the observations $\{(\mathbf{z}_k, h_{m|l,k})\}_{k=1}^{N}$ and observation weights $\{h_{l,k}\}_{k=1}^{N}$ for each low-level gating network.

4) Iterate using the updated parameter values of the gating networks (i.e., $\mathbf{a}_i$s and $\mathbf{a}_{j|i}$s).

V. APPLICATION

Interplanetary Orbit Determination

We consider the interplanetary spacecraft tracking problem [13]. The system state vector for this problem has 33 components including the spacecraft (perturbed) position and velocity from the heliocentric conic and various error sources

which are perturbations relative to a nominal value. In this application, the processed measurements are radiometric (Doppler) collected from Deep Space Stations (DSSs) 15 (Goldstone, California), 45 (Canberra, Australia), and 65 (Madrid, Spain). The error sources include solar pressure, nongravitational accelerations (such as a thruster misfiring or gas leaks), atmosphere (formed by dry troposphere, day and night time ionosphere for each station), stations positions (for each station), and Earth platform. In other words, the 33rd-order system state vector is defined as

$$\mathbf{x} = (\mathbf{x}_{sc} \quad \dot{\mathbf{x}}_{sc} \quad \mathbf{x}_{srp} \quad \mathbf{x}_{nga} \quad \mathbf{x}_{atm} \quad \mathbf{x}_{sta} \quad \mathbf{x}_{eor})^T \qquad (24)$$

where $\mathbf{x}_{sc}$ and $\dot{\mathbf{x}}_{sc}$ are the spacecraft perturbed position and velocity from the heliocentric conic. The error sources form the remainder of the state vector: $\mathbf{x}_{srp}$ (solar radiation pressure) and $\mathbf{x}_{nga}$ (nongravitational accelerations) are error sources which directly affect the spacecraft; and $\mathbf{x}_{atm}$ (atmosphere formed by dry troposphere, day and night time ionosphere), $\mathbf{x}_{sta}$ (stations locations), and $\mathbf{x}_{eor}$ (Earth platform) are errors which corrupt the measurements. As in [14], the wet troposphere was not included because 90% of the total refractivity is due to the dry component of the troposphere in the first 7 km. Also, beyond this altitude the wet component diminishes and the dry component becomes the total refractivity. The inclusion of various filter parameters in the state vector follows the approach used in [29] where it was shown that improved navigation accuracy can be achieved when using this formulation (known as the "enhanced filter") rather than the standard filter which accounts for filter parameters as consider parameters but does not estimate them. A complete derivation of the problem, a description of $\mathbf{Q}_k$, $\mathbf{R}_k$, $\mathbf{\Phi}_k$, $\mathbf{H}_k$, and initial values used to initialize the filter equations can be found in [13, 14].

Station location errors are modeled as random biases while the remaining error sources are modeled as exponentially correlated random variables ECRVs [25]. An ECRV can be described by

$$x_{k+1} = \exp\left(\frac{-\Delta t}{\tau}\right) x_k + w_k \qquad (25)$$

where x represents the random process, Δt is the time interval from time k to $k + 1$, τ is the time constant of the process, and w_k is a Gaussian white noise process with covariance

$$E[w_k^2] = q_k = \sigma^2 \left(1 - \exp\left(\frac{-2\Delta t}{\tau}\right)\right) \qquad (26)$$

where σ is the steady-state deviation of the Gauss–Markov process x.

Measurement collections start with DSS 15 on day seven, DSS 45 on day ten, and DSS 65 on

day twelve. At a particular site, a measurement is collected every 10 min if the elevation angle is above a certain minimum. This scheme is continued until a specified final time. This tracking scenario is devised to match the Mars Pathfinder mission requirements [14]. In effect, this will allow the simulation of a single 4 hours pass per week at each station. The Doppler formulation used here is known as two-way differenced-range. A measurement is simply twice the one-way measurement which is obtained by taking the difference of two range measurements divided by the time between them. A complete description of the processing of differenced-range Doppler data for the Mars Pathfinder mission can be found in [26].

Error budget calculations can be used to determine the contribution of various error sources to the overall state estimation uncertainties [25]. For the Mars Pathfinder mission interplanetary orbit determination problem, the major contributing error source (when processing differenced-range Doppler measurements) is found to be nongravitational accelerations [27, 28]. Other major contributors are measurement noise and solar pressure. Hence, it is desirable to accurately find the parameters of these error sources. In this study, we concentrate on the adaptation of the nongravitational accelerations and measurement noise parameters. This will be accomplished utilizing the hierarchy of adaptive Kalman filter banks described earlier. The proposed hierarchy is shown to be a useful tool in accurately finding the parameters of the major error sources. The adaptive Kalman filter bank hierarchy is used to study the selections made by the gating networks as measurements are processed. In particular, we examine the effect on the selections made by the gating networks at different levels of hierarchy when the nongravitational accelerations level is changed half-way through the interplanetary trajectory. The components of the nongravitational accelerations perturbations vector are assumed independent. Each of the nongravitational accelerations perturbations vector components is modeled as an ECRV. The vector is composed of three ECRVs with a time constant of τ_{nga} and steady-state deviations $\sigma_{nga,r}$ (radial), and $\sigma_{nga,x}$, $\sigma_{nga,y}$ (transverse). The steady-state deviations are assumed to be equal. That is,

$$\sigma_{nga} = \sigma_{nga,r} = \sigma_{nga,x} = \sigma_{nga,y}.$$

The nominal nongravitational acceleration steady-state deviation is $\sigma_{nga1}^{\star} = 0.7 \times 10^{-12}$ km/s^2 for the first part of the trajectory and $\sigma_{nga2}^{\star} = 7.0 \times 10^{-12}$ km/s^2 for the second part of the trajectory. The time constant, $\tau_{nga}^{\star} = 7.0$ days, is nominal for the entire interplanetary trajectory. Also, the nominal measurement noise is $\sigma_n^{\star} = 0.01$ mm/s for the entire interplanetary trajectory. The abrupt change in the steady-state deviation of the nominal nongravitational acceleration is a test condition to show the power of the proposed method.

TABLE I
Values of Unknown Parameter Vector in Filter Models

Filter bank	Filter number	τ_{nga} (days)	σ_{nga} ($\times 10^{-12}$ km/sec^2)	σ_n (mm/sec)
1	1|1	1.0	0.07	0.01
	2|1	7.0	0.07	0.01
	3|1	30.0	0.07	0.01
2	1|2	1.0	0.7	0.01
	2|2	7.0	0.7	0.01
	3|2	30.0	0.7	0.01
3	1|3	1.0	7.0	0.01
	2|3	7.0	7.0	0.01
	3|3	30.0	7.0	0.01
4	1|4	7.0	0.7	0.001
	2|4	7.0	0.7	0.05
	3|4	7.0	0.7	0.10

Note: KF 2 | 2 optimal over 1st half of trajectory and KF 2 | 3 best over second half of trajectory.

In reality, we hope that such a large change doesn't actually occur.

Table I indicates that a hierarchical architecture consisting of four Kalman filter banks is used. Before discussing the hypothesized parameter vectors, it is important to point out that the optimal Kalman filter realization for the entire interplanetary cruise is a single filter which uses the nominal parameters of the first half of the trajectory and then the second set of nominal parameters corresponding to the last part of the trajectory. Hence, none of the filters are optimal for the entire interplanetary trajectory. However, we note that the KF 2 | 2 is the optimal Kalman filter for the first part of the trajectory. Also, KF 2 | 3 is the best Kalman filter realization for the second part of the trajectory since its parameters correspond to the second part of the trajectory. In the first bank of the hierarchy, all filters are modeled with the same σ_{nga} and σ_n. The filters differ in the selected values for the time constant τ_{nga}. A similar variation in the filter parameters was done in the second and third filter banks. The filters in the fourth bank of the hierarchy are modeled with the same parameters for the nongravitational accelerations, τ_{nga} and σ_{nga}, while differing in the measurement noise deviation, σ_n. Finally, it is assumed that the implemented (single) filter realization by the navigation team is KF 2 | 2. In other words, following current practices, only one filter parameter set is used for the entire trajectory, and in our case that filter is KF 2 | 2.

The following experiment illustrates the multi-level hypothesis testing capability of the hierarchy and the advantages of implementing a hierarchical parallel architecture of Kalman filters instead of a single filter. Furthermore, an assessment of the performance of various learning rules is presented.

Experiment:

In this experiment, we first investigate the behavior of the gating network when using an instantaneous gradient ascent learning algorithm. All gating
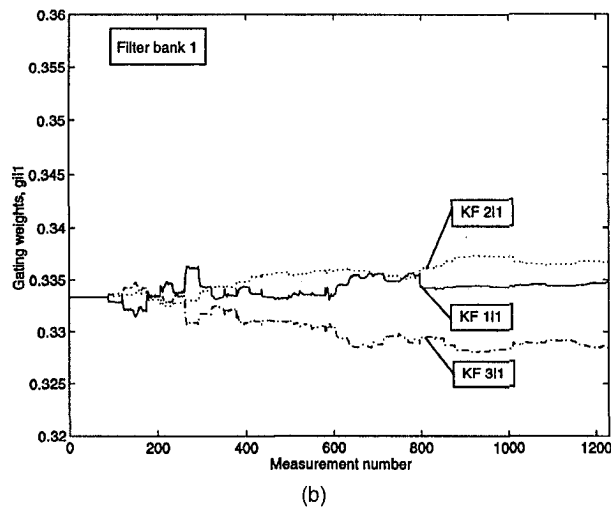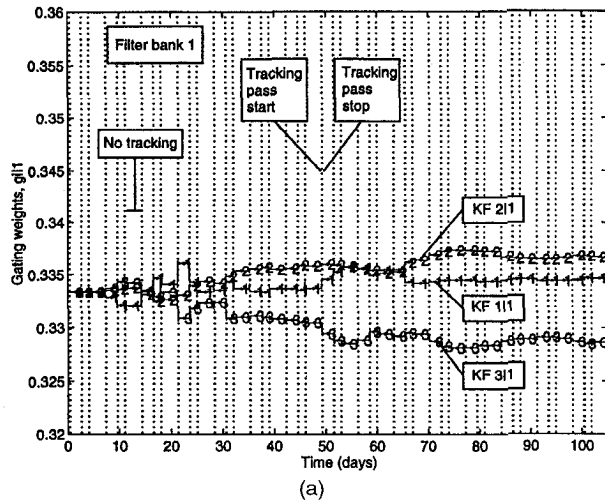
(a)

(b)

Fig. 5. Gating weight history using gradient ascent (Kalman filter bank 1), where (b) illustrates correspondence between time and measurement number. Same translation scale can be applied to Figs. 6–18 to convert time into measurement number on x-axis.

networks in the hierarchical architecture use a unity learning-rate parameter (i.e., $\eta = 1.0$). Also, all weight vectors $\mathbf{a}_{j|i}$ were initialized to the same value so as not to favor any filter over another (initially). Similarly, all weight vectors $\mathbf{a}_i$ were initialized to the same value filter bank initially. The gating weight histories, which indicate the various selections made in the hierarchical structure, are depicted in a plot versus time. The horizontal lines will be used to indicate a constant weight between tracking passes when no measurements are available. Also, vertical lines will be used to specify the beginning and end of the tracking passes when weights are computed. Tracking passes are so short that they appear as vertical lines. A total of 1230 observations are collected over the interplanetary trajectory.

Fig. 5(a) reveals that in the first Kalman filter bank, no filter is preferred over the others. The fact that σ_{nga} is an order of magnitude smaller than the optimal σ_{nga}^{*} dominates. Fig. 5(b) is used to plot the
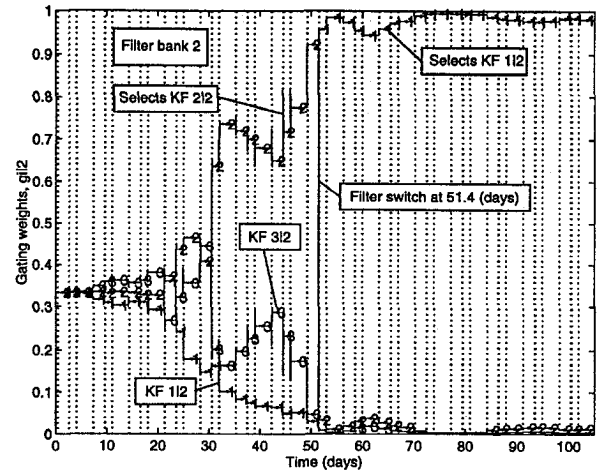

Fig. 6. Gating weight history using gradient ascent (Kalman filter bank 2).

gating weights versus the measurement number so as to clarify the translation from time to measurement.

The gating weight history of the second filter bank is illustrated in Fig. 6. It indicates that initially KF 2 | 2 was selected by the gating network, since $g_{2|2}$ is close to unity. This was expected since KF 2 | 2 is the optimal filter realization for the first part of the trajectory. This selection was later modified when the nongravitational accelerations steady-state deviation was increased. The increase may reflect a thruster misfiring or gas leak, for example. In fact, the gating weight of KF 1 | 2, is close to unity through the second half of the trajectory. This indicates a parameter change. This need not be a change in the time constant τ_{nga}. Rather, it is just a change in a parameter for which the gating network is attempting to find another best filter realization. In our case, since all filters in the second Kalman filter bank are modeled with the same σ_{nga} and σ_n, the gating network selects the filter with the best time constant τ_{nga} for the second part of the trajectory, i.e, the time constant that will result in the smallest measurement residual. The selection is justified by the fact that since all filters have the same σ_{nga} and σ_n, the gating network selects KF 1 | 2 which has a smaller τ_{nga}. The effect of this selection is to increase the noise as shown by (26), which indicates that for a constant steady-state deviation, measurement noise increases as the time constant decreases. This has the same effect of increasing the measurement noise by fixing the time constant and increasing the steady-state deviation. In fact, that was the happening when σ_{nga} was increased over the second half of the trajectory. Note that simulation results indicate that the explanation applicable here is not applicable to the first filter bank because its σ_{nga} is much smaller than that of the second filter bank.

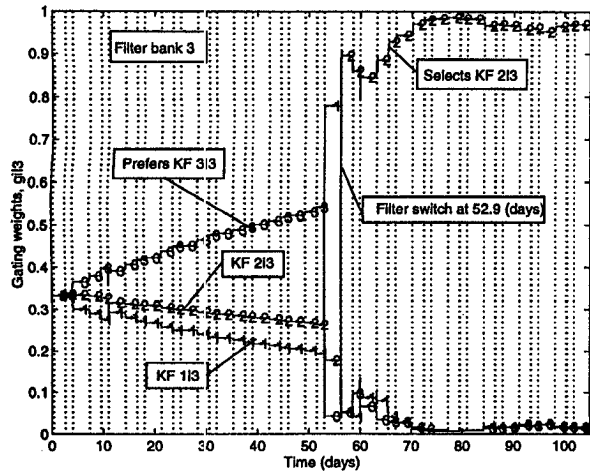Examining Fig. 7 reveals a similar switch phenomena in the third Kalman filter bank. Initially,

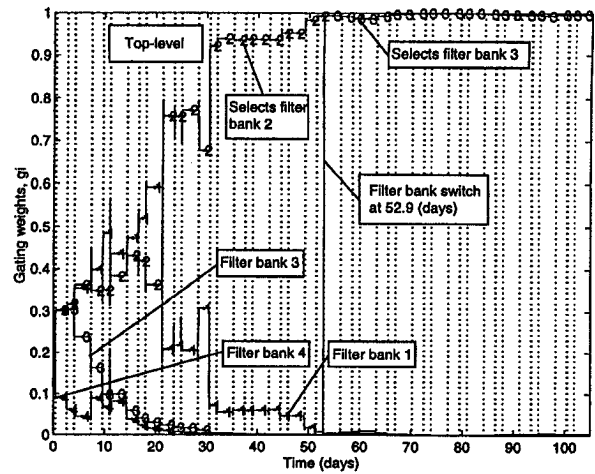Fig. 7. Gating weight history using gradient ascent (Kalman filter bank 3).

Fig. 8. Gating weight history using gradient ascent (Kalman filter bank 4).

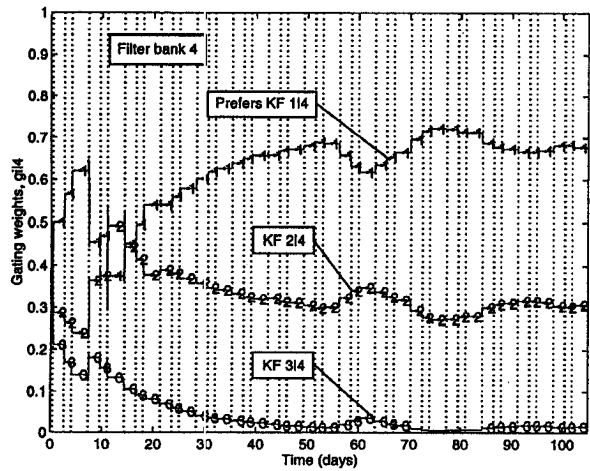Fig. 9. Gating weight history using gradient ascent (top-level).

the gating network regulating the third Kalman filter bank favors KF 3 | 3 slightly over the other filters. In fact, none of the filters is optimal for the first part of the trajectory. On the other hand, when the abrupt change in the nongravitational accelerations steady-state deviation occurs, the gating network rapidly selects the best available realization KF 2 | 3.

The gating weight history plot of the fourth bank is shown in Fig. 8. It indicates that the gating network favors KF 1 | 4 and KF 2 | 4. In this bank none of the filters are optimal for any part of the trajectory. The gating network is favoring those filters whose measurements residuals are smaller.

At the top-level, the gating network selects the second Kalman filter bank over the first part of the interplanetary trajectory as indicated by Fig. 9. This is expected since KF 2 | 2 is indeed optimal over this first regime and is located in the second filter bank. The top-level gating network modifies its selection to the third Kalman filter bank since KF 2 | 3 is best over

the second regime and is located in the third filter bank.

The selections made in the filter banks when combined with the top-level selection can shed light on what has really occurred. The fact that the top-level gating network switched from the second to the third Kalman filter bank indicates that KF 2 | 3 is preferred over both KFs 2 | 2 and 1 | 2 in the second half of the trajectory. A comparison of the parameters of KFs 2 | 2, 1 | 2, and 2 | 3, indicates that a change in the steady-state deviation σ_{nga} has occurred and not τ_{nga}. Note that such conclusions are not so readily obtained when regulating the filters with a single gating network.

In a realistic setup, the truth is unknown and hence this technique will yield the most suitable filter realization (which is not necessarily optimal). However, this approach will help identify the changing parameters. To adapt the parameters which most contribute to the estimation error, a number of techniques (described in [6]) can be used to accurately find an "optimal" realization of these parameters.

We now examine the estimation errors of the position components of the spacecraft when using the nominal filter estimate (i.e., the realization KF 2 | 2 implemented by the navigation team) and the hierarchical structure estimate. As shown in Fig. 3, the optimal estimate provided by the hierarchy is given by

$$\hat{x}^+_{opt,k} = \sum_{i=1}^{K} g_i \sum_{j=1}^{L} \hat{x}^+_{ji,k} g_{j|i}.$$

The estimation errors plots are shown in Figs. 10–12. Each of these figures is composed of two plots. These are plots of the estimation errors (before and after the state estimate update) versus time and measurement number, respectively. The dotted lines are the plots of the state estimate error of the spacecraft when using the nominal filter KF 2 | 2. The state estimate error of the spacecraft obtained when
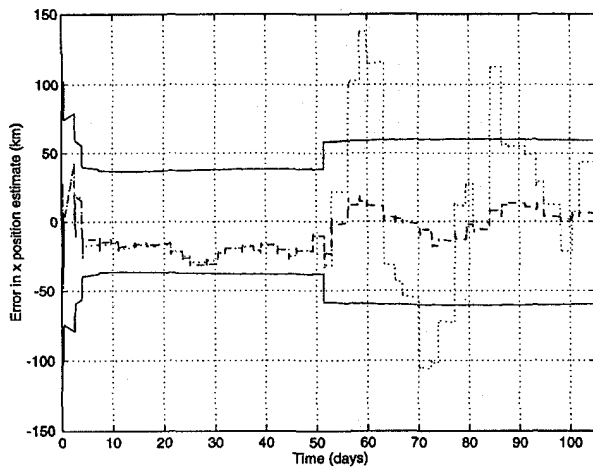
Fig. 10. Spacecraft position estimate error (x component) for hierarchy (dashed) and nominal filter (dotted).
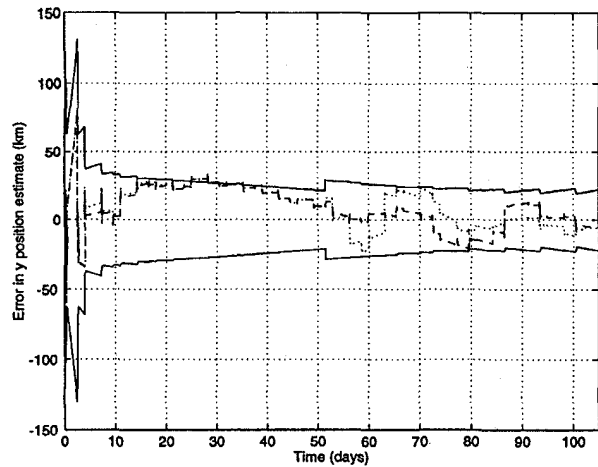

Fig. 11. Spacecraft position estimate error (y component) for hierarchy (dashed) and nominal filter (dotted).

utilizing the hierarchy is plotted using the dashed lines. It is the estimation error that occurs when using a weighted sum of the various filters and filter banks. The state estimation error plots indicate the error before and after measurement incorporation. The solid lines on the plots represent the plot of (+ and −) the square root of the corresponding diagonal elements of the covariance matrix (before and after the update). The covariance matrices are those of KF 2 | 2 (since it is the optimal) for the first part of the trajectory and KF 2 | 3 (since it is the best available) for the second part of the trajectory. Figs. 10–12 reveal that the estimation error of the hierarchy is very close to the optimal estimation error over the first half of the trajectory. In fact, a small difference is noted in the first part of the trajectory. This slight difference appears in the early part of the first trajectory as the gating networks have not yet selected (the optimal) KF 2 | 2, as shown in Figs. 6 and 9. However, once the selection is made, there is no difference between the optimal estimate error given by KF 2 | 2 and that of the hierarchy (since the selection is the same KF 2 | 2). In contrast, when the switch happens the estimation errors of the spacecraft provided by the nominal filter KF 2 | 2 clearly increase as shown by Figs. 10–12. However, since the hierarchical structure contains a diversity of models, a switch to an appropriate filter can be made and thus a better estimation error can still be attained. In this example, the switch within the hierarchy now favors KF 2 | 3 as shown earlier in the gating weights histories of Figs. 6 and 9. The combined estimation error for the on-line estimates (i.e., hierarchy's estimates) is illustrated in Fig. 13 which indicates a satisfactory and stable position error even after the abrupt change in nongravitational accelerations.

We conclude this experiment by illustrating how by using both learning techniques (i.e., gradient ascent and EM) we can better assess the performance
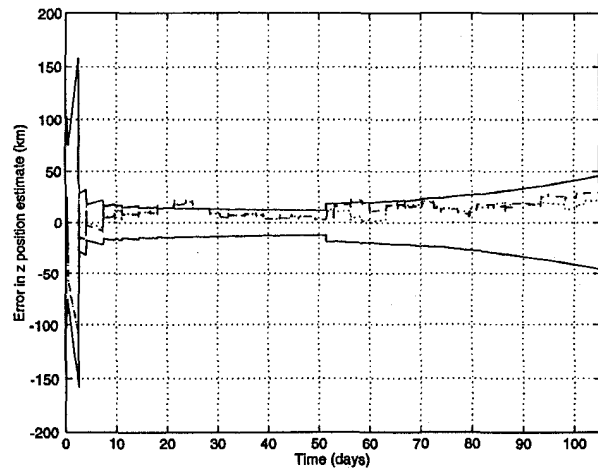

Fig. 12. Spacecraft position estimate error (z component) for hierarchy (dashed) and nominal filter (dotted).
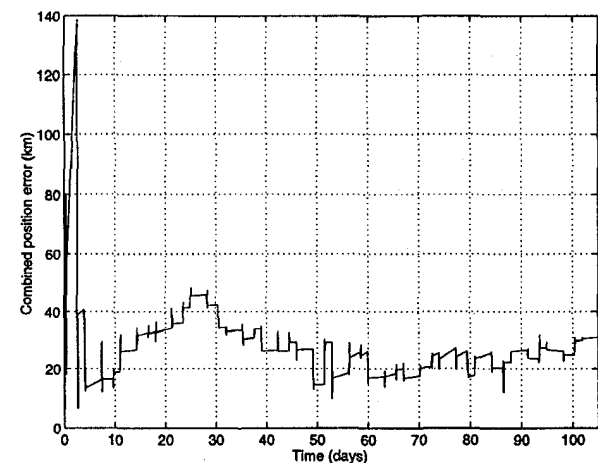

Fig. 13. Spacecraft combined position estimate error for hierarchy.

of the various Kalman filters by obtaining an "optimal" set of weights. This is done by processing the measurement set once using gradient ascent.

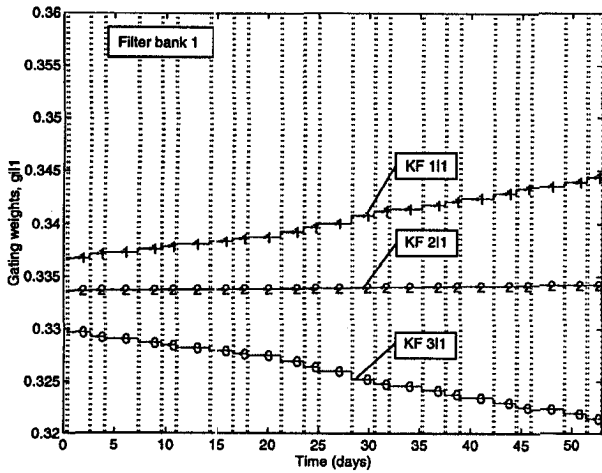Fig. 14. Gating weight history using EM for first measurement sequence (Kalman filter bank 1).

Fig. 15. Gating weight history using EM for first measurement sequence (Kalman filter bank 2).

The results, shown in Figs. 6, 7, and 9, reveal the occurrence of a switch in both the second and third Kalman filter banks and at the top-level of the hierarchy. Using this finding, the data is divided into two subsets consisting of the measurements processed before and after the measurements processed after the switching point, respectively. Now, EM which is a batch mode approach is used to solve for the gating network parameters corresponding to each set of measurements. That is, EM is solving the maximum likelihood problems posed by (22) and (23), and finding the "optimal" a_is and $a_{j|i}$s for each measurement subset. The obtained a_is and $a_{j|i}$s yield an "optimal" set of weights g_is and $g_{j|i}$s.

Unlike the gradient ascent algorithm which can be used in real-time and post-processed data applications, the iterative nature of EM makes the algorithm only suitable for post-processed data applications. The EM algorithm is not computationally intensive in terms of using the Kalman filters composing the hierarchy. In fact, the weighted measurements residuals, saved when using the gradient ascent algorithm, are the only needed information by the EM algorithm when solving for the gating network parameters. That is, the Kalman filters composing the hierarchy process the measurement set only once.

The returned solution for the first measurement set is shown in Figs. 14–18. This result was attained after 500 iterations. The gating weight plots in Figs. 14–18 show a similarity with the attained weights using gradient ascent and shown in Figs. 5–9. The justifications behind these results were discussed earlier in this experiment when examining the gating weight histories of Figs. 5–9. Hence, it is apparent that the gradient ascent algorithm is capable of yielding a good result in a single pass. However, gradient ascent goes through an initial learning phase as shown in Fig. 6 for example, where the algorithm is undecided in the very early stage on whether KF
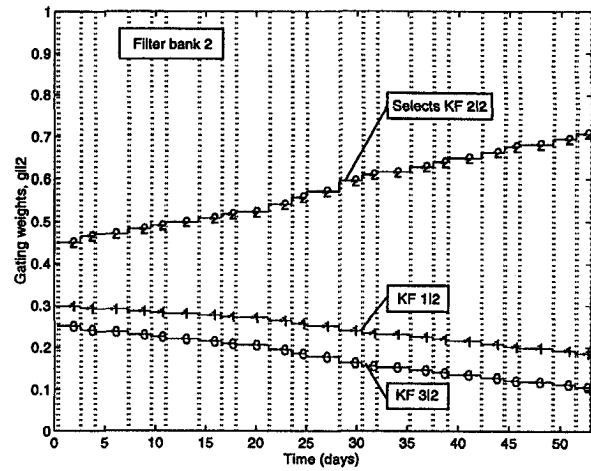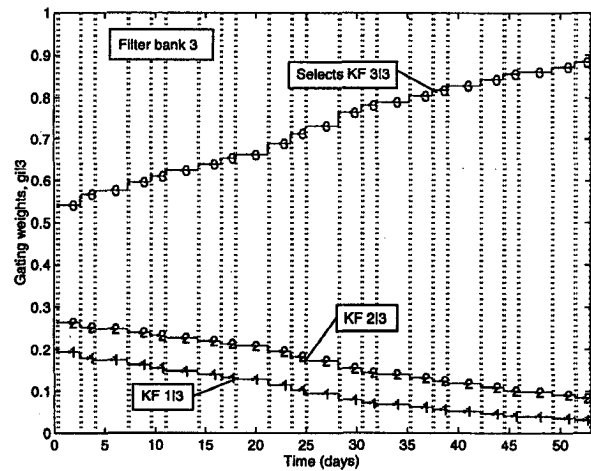

Fig. 16. Gating weight history using EM for first measurement sequence (Kalman filter bank 3).

2 | 2 is the optimal realization or not. To address that relatively brief initial stage, the EM algorithm may be used to clearly identify whether KF 2 | 2 is to be selected initially. This is indeed true as shown in Fig. 15, i.e., KF 2 | 2 is the realization to be used. It is important to note that the gating network parameters obtained with EM and plotted in Figs. 14–18 are not necessarily the global maximum of the maximization problems posed by (22) and (23). Rather, EM only guarantees that the returned a_is (g_is) and $a_{j|i}$s ($g_{j|i}$s) form a local maximum [11]. Hence, it is sufficient to use EM to clarify initial learning ambiguities by iterating until no significant change in the weight histories is observed. Finally, a similar approach can be used to find the best set of weights for the second measurement sequence.

VI. CONCLUSIONS

The proposed hierarchy of adaptive Kalman filter banks can be used as a multi-level hypothesis
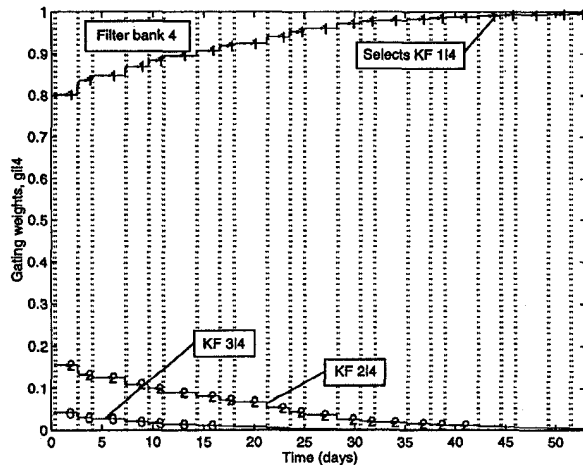
Fig. 17. Gating weight history using EM for first measurement sequence (Kalman filter bank 4).
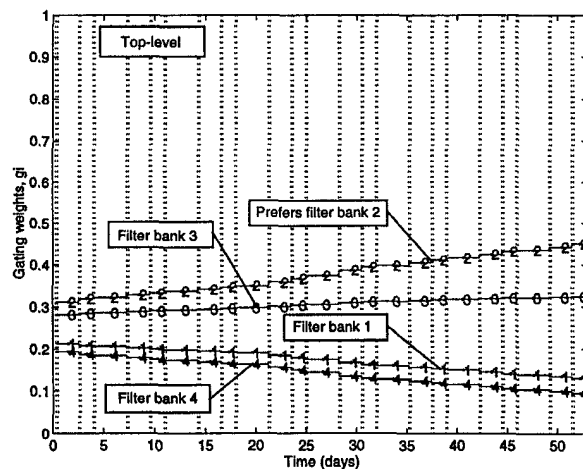

Fig. 18. Gating weight history using EM for first measurement sequence (top-level).

testing device. It allows the examination of a parameter change on an individual basis as well as in combination with other parameters. This is achieved by using a gradient ascent learning rule which has the ability of detecting abrupt changes due to a different operating regime. The use of EM allows us to reach a better solution for the gating weights since the returned solution seeks the best gating parameters by considering an entire sequence of measurements. Also, the solution returned by the EM algorithm is a local maximum [11]. However, since EM is a batch algorithm, its results are only meaningful if the underlying statistics do not vary much. In a real life problem such as interplanetary orbit determination, there may be occasional abrupt changes due to a different regime. Such a change can be detected by an on-line algorithm, which can thus partition the measurements into relatively homogeneous segments each of which can be addressed by applying EM. EM gives a clearer assessment of the initial learning phase experienced using gradient ascent.

The obtained results in the interplanetary orbit determination problem indicate that the proposed scheme can be used as an assistant tool in the process of selecting the parameters of the navigational filter. This is achieved by modeling each bank of the hierarchy with a different set of values of a particular parameter. Therefore, the gating network in a given bank indicates which filter is performing best when considering variations of a single parameter. Similarly, the top-level gating network indicates which bank is performing best. The presented simulation results show the ability of the scheme to identify parameter changes and hence modify its filter and bank selections. This information (i.e., the change in parameter) can be used to appropriately modify the navigational filter parameters and attain a better match to the actual parameters.

The proposed hierarchical architecture can be used with extended Kalman filters as the individual experts and may be applicable in the data fusion framework. Furthermore, recent results show that using a localized model for the gating network provides a methodical approach to the model selection problem, often leading to further performance improvements [30]. The application of such localized models to the interplanetary orbit determination problem is worth investigating further.

ACKNOWLEDGMENTS

REFERENCES

[1] Kalman, R. E., and Bucy, R. S. (1961)
     New results in linear filtering and prediction theory.
     Transactions of ASME, Journal of Basic Engineering, ser.
     D, **83** (Mar. 1961), 95–108.

[2] Kalman, R. E. (1961)
     New methods and results in linear prediction and filtering theory.
     Proceedings of the Symposium on Engineering Applications of Random Function Theory and Probability.
     New York: Wiley, 1961.

[3] Price, C. F. (1968)
     An analysis of the divergence problem in the Kalman filter.
     IEEE Transactions on Automatic Control, **AC-13** (Dec. 1968), 699–702.

[4] Fitzgerald, R. J. (1971)
     Divergence of the Kalman filter.
     IEEE Transactions on Automatic Control, **AC-16** (Dec. 1971), 736–747.

[5] Magill, D. T. (1965)
     Optimal adaptive estimation of sampled stochastic processes.
     IEEE Transactions on Automatic Control, **AC-10** (Oct. 1965), 434–439.

[6] Chaer, W. S., Bishop, R. H., and Ghosh, J. (1997)
A mixture-of-experts framework for adaptive Kalman filtering.
IEEE Transactions on Systems, Man, Cybernetics, **27**, Pt. B (June 1997), 452–464.

[7] Jacobs, R. A., and Jordan, M. I. (1991)
A competitive modular connectionist architecture.
In R. P. Lippman, et al. (Eds.), *Advances in Neural Information Processing Systems 3*, Morgan Kaufmann, 1991, 767–773.

[8] Jordan, M. I., and Jacobs, R. A. (1994)
Hierarchical mixtures of experts and the EM algorithm.
Neural Computation, **6** (1994), 181–214.

[9] Haykin, S. (1994)
Neural Networks.
New York: MacMillan, 1994.

[10] Waterhouse, S. R., and Robinson, A. J. (1994)
Classification using hierarchical mixture of experts.
Neural Networks for Signal Processing IV: Proceedings of the IEEE Workshop, 1994, 177–186.

[11] Dempster, A. P., Laird, N. M., and Rudin, D. B. (1977)
Maximum likelihood from incomplete data via the EM algorithm.
Journal of the Royal Statistics Society, Series B-39 (1977), 1–38.

[12] Redner, R. A., and Walker, H. F. (1984)
Mixture densities, maximum likelihood and the EM algorithm.
SIAM Review, **26** (Apr. 1984), 195–239.

[13] Burkhart, P. D., and Bishop, R. H. (1996)
Adaptive orbit determination for interplanetary spacecraft.
AIAA Journal of Guidance, Control, and Dynamics, **19** (May–June 1996), 693–701.

[14] Kallemeyn, P. (1994)
Pathfinder project navigation plan—Critical design review version.
JPL Document D-11349, Jet Propulsion Laboratory, Pasadena, CA, July 1994.

[15] Bridle, J. S. (1990)
Probabilistic interpretation of feedforward classification network outputs, with relationship to statistical pattern recognition.
In F. Fougelman-Soulie and J. Hérault (Eds.), *Neurocomputing: Algorithms, Architectures, and Applications*.
New York: Springer-Verlag, 1990, 227–236.

[16] Bridle, J. S. (1990)
Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters.
In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2*.
Morgan Kaufmann, 1990, 211–217.

[17] Jordan, M. I. (1996)
A statistical approach to decision tree modeling.
In *Proceedings of COLT 94*, 1996.

[18] Blom, H. A. P. (1984)
An efficient filter for abruptly changing systems.
In *Proceedings of the 23rd IEEE Conference on Decision and Control*, Dec. 1984, 656–658.

[19] Blom, H. A. P., and Bar-Shalom, Y. (1988)
The interacting multiple model algorithm for systems with Markovian switching coefficients.
IEEE Transactions on Automatic Control, **33** (Aug. 1988), 780–783.

[20] Lerro, D., and Bar-Shalom, Y. (1993)
Interacting multiple model tracking with target amplitude feature.
IEEE Transactions on Aerospace and Electronic Systems, **29** (Apr. 1993), 494–509.

[21] Li, X. R., and Bar-Shalom, Y. (1993)
Performance prediction of the interacting multiple model algorithm.
IEEE Transactions on Aerospace and Electronic Systems, **29** (July 1993), 755–771.

[22] Li, X. R., and Bar-Shalom, Y. (1994)
A recursive model approach to noise identification.
IEEE Transactions on Aerospace and Electronic Systems, **30** (July 1994), 671–684.

[23] Daeipour, E., and Bar-Shalom, Y. (1995)
An interacting multiple model approach for target tracking with glint noise.
IEEE Transactions on Aerospace and Electronic Systems, **31** (Apr. 1995), 706–715.

[24] McCullagh, P., and Nelder, J. A. (1983)
Generalized Linear Models.
London: Chapman and Hall, 1983.

[25] Gelb, A. (1974)
Applied Optimal Estimation.
Cambridge, MA: The M.I.T. Press, 1974.

[26] Bishop, R. H., and Chaer, W. S. (1997)
Adaptive interplanetary navigation using differenced range observables.
Advances in the Astronautical Sciences, series ed. H. Jacobs, 1997.

[27] Estefan, J. A., and Burkhart, P. D. (1994)
Sensitivity analysis of the enhanced orbit determination filter—Error budget development.
Advances in the Astronautical Sciences, J. E. Cochran, Jr., et al. (Eds.), **87** (1994), 1199–1214.

[28] Burkhart, P. D., Bishop, R. H., and Estefan, J. A. (1995)
Covariance analysis of mars pathfinder interplanetary cruise.
Advances in the Astronautical Sciences, series ed. H. Jacobs, **89** (1995).

[29] Estefan, J. A., Pollmeier, V. M., and Thurman, S. W. (1993)
Precision X-Band Doppler and ranging navigation for current and future Mars exploration missions.
Advances in the Astronautical Sciences, J. Teles and M. V. Samii (Eds.), **84** (1993), 3–16.

[30] Ramamurti, V., and Ghosh, J. (1996)
Structural adaptation in mixture of experts.
In *Proceedings of ICPR 96*, track D, 1996, 704–708.

Wassim S. Chaer (M'93) received the B.E. degree in computer and communication engineering from the American University of Beirut, in 1989, the M.S. degree in systems and control engineering from the University of West Florida, in 1991, and the Ph.D. degree in aerospace engineering from The University of Texas at Austin, in 1996.

Since October 1996, he has been a Consultant with SDT's Research Group. His research interests include adaptive estimation and control, data fusion, optimization, neural networks, fuzzy logic and control, and search algorithms.

Robert H. Bishop (M'92) received his B.S. and M.S. in aerospace engineering from Texas A&M University, College Station, and Ph.D. in electrical and computer engineering from Rice University, Houston, TX.

He currently holds the Myron L. Begeman Fellowship in Engineering in the Department of Aerospace and Engineering and Engineering Mechanics at the University of Texas at Austin. He spent ten years as a practicing engineer with The Charles Stark Draper Laboratory, including six years as an on-site resident at NASA Johnson Space Center. He was twice a Faculty Fellow of the NASA Jet Propulsion Laboratory and a Welliver Fellow of the Boeing Company.

Dr. Bishop is a co-author (with R. C. Dorf) of *Modern Control Systems* (Addison-Wesley, 1998) and author of *Modern Control Systems Analysis and Design Using MATLAB and SIMULINK* (Addison-Wesley, 1997). In 1991, he co-authored the *Best Conference Paper* at the *AIAA Guidance, Navigation and Control Conference* on the topic of aircraft tracking. He serves on the AAS Spaceflight Mechanics Technical Committee and is an associate editor of the *Journal of the Astronautical Sciences*. He is an Associate Fellow of AIAA.
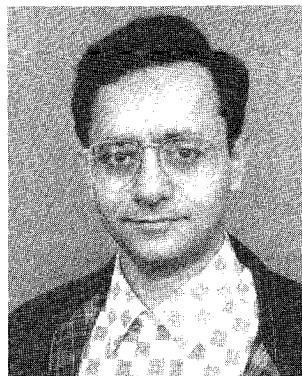
Joydeep Ghosh (M'88) was educated at the Indian Institute of Technology, Kanpur, India (B. Tech, 1983) and The University of Southern California, Los Angeles (M.S., Ph.D., 1988).

He is currently an Associate Professor with the Department of Electrical and Computer Engineering at The University of Texas at Austin, where he holds the Endowed Engineering Foundation Fellowship. He directs the Laboratory for Artificial neural Systems (LANS), where his research group is studying the theory and applications of adaptive pattern recognition and hybrid systems.

Dr. Ghosh has published more than 100 refereed papers and edited 6 books. He served as the general chairman for the SPIE/SPSE Conference on Image Processing Architectures, Santa Clara, Feb. 1990, as Conference Co-Chair of Artificial Neural Networks in Engineering (ANNIE)'93 through ANNIE'96, and in the program committee of several conferences on neural networks and parallel processing. Dr. Ghosh received the 1992 Darlington Award for the Best Paper in the areas of CAS/CAD, besides "best conference paper" citations for four neural network papers. He is an associate editor of *Pattern Recognition, IEEE Transactions on Neural Networks, Neural Computing Surveys* and *International Journal of Smart Engineering Design*.