Marginalized Particle Filtering for Blind System Identification

MARGINALIZED PARTICLE FILTERING FOR BLIND SYSTEM IDENTIFICATION

BY

MICHAEL DALY

OCTOBER 2004

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2004) McMaster University
(Electrical & Computer Engineering) Hamilton, Ontario

TITLE:	Marginalized Particle Filtering for Blind System Identification
AUTHOR:	Michael Daly B.Sc. (Mathematics and Engineering) Queen's University, Kingston, ON, Canada
SUPERVISOR:	Dr. James P. Reilly
NUMBER OF PAGES:	xiii, 94

To my parents and sisters

Abstract

This thesis develops a marginalized particle filtering algorithm for the blind system identification problem. The blind system identification problem arises in many fields, including speech processing, communications, biomedical signal processing, sonar and seismology. The state space model under consideration uses a time-varying autoregressive (AR) model for the sources, and a time-varying finite impulse response (FIR) model for the channel. The multi-sensor measurements result from the convolution of the sources with the channels in the presence of additive noise. A numerical approximation to the optimal Bayesian solution for the sequential state estimation problem is implemented using the particle filter. Estimates of the sources are recovered directly by marginalizing the AR and FIR coefficients out of the posterior distribution for the unknown system parameters. The resulting marginalized particle filtering algorithm allows efficient identification of the system. Simulation results are given to verify the performance of the proposed method. The block sequential importance sampling (BSIS) formulation of the particle filter is also introduced to exploit the structure inherent in the convolution state space model.

Acknowledgements

I would first like to thank Dr. Reilly for his tremendous support and enthusiasm. His encouragement has been invaluable throughout my research. I also thank Dr. Reilly for providing the opportunity to join him at Melbourne University while he was on research sabbatical. The trip to Australia was both productive and enjoyable. I was fortunate to have the opportunity to collaborate with Dr. Jonathon Manton and Dr. Mark Morelande while visiting. I also learned a great deal about particle filters from Mark during his visit to McMaster this summer. Dr. James Hopgood from the University of Cambridge provided helpful suggestions early in the research while visiting McMaster. I thank them all for their insightful comments and discussions on my research.

The members of my research group Amin, Nazanin and Derek have created a relaxed environment in which to work. I had the chance to collaborate most often with Derek, and I have gained a lot from our many conversations on research. Around the Department, Cheryl Gies deserves special mention for her extraordinary help.

I was also lucky enough to meet Cristina while at McMaster, and she has made my time all the more enjoyable. Gracias mi querido. Finally, thanks to my parents and sisters for all their love and support.

Notation and Acronyms

Symbol and Definition	
x	Scalar
$\mathbf{x}$	Vector
$\mathbf{x}[i]$	i^{th} element of vector $\mathbf{x}$
$\mathbf{X}$	Matrix
$\mathbf{X}^{\mathrm{T}}$	Matrix transpose
$\mathbf{X}^{\mathrm{H}}$	Hermitian transpose
$\lvert\mathbf{X}\rvert$	Determinant of $\mathbf{X}$
$\mathrm{diag}(\mathbf{x})$	Diagonal matrix formed from vector $\mathbf{x}$
$\mathbb{E}\{\cdot\}$	Expectation Operator
$\mathcal{N}(\mu, \Sigma)$	Normal distribution with mean μ, covariance Σ
$\mathcal{IG}(\upsilon, \gamma)$	Inverse Gamma distribution
$\lvert\lvert\mathbf{x}\rvert\rvert_2$	Euclidean norm of vector $\mathbf{x}$
$\otimes$	Kronecker product
$\delta(\cdot)$	Dirac-delta function
i.i.d.	identically and independently distributed
$\mathbf{I}_n$	Identify matrix of dimension n
$\mathbf{0}_{n,m}$	Matrix of zeros of dimension $n \times m$
Bayesian Filtering	
N	Number of sources
J	Number of sensors

L	Order of FIR channel
P	Order of source AR model
M	Memory of particle filter: max(P,L)
B	Block length for BSIS implementation
$\mathbf{s}_k$	Source vector
$\mathbf{s}_{L,k}$	Concatenation of most recent L source vectors
$\mathbf{y}_k$	Measurement vector
$\mathbf{h}_k$	FIR channel coefficients vector
$\mathbf{H}_k$	FIR channel coefficients matrix
$\mathbf{a}_k$	AR coefficients vector
$\mathbf{A}_k$	AR coefficients matrix
$\mathbf{v}_{k-1}$	Source noise vector
$\sigma^2_{v,n}$	Noise variance of n^{th} source
$\mathbf{\Sigma}_v$	Source noise covariance
$\mathbf{w}_k$	Measurement noise vector
$\sigma^2_{w,j}$	Noise variance of j^{th} sensor
$\mathbf{\Sigma}_w$	Measurement noise covariance
$\text{SNR}_{s,n}$	n^{th} source Signal-to-Noise ratio
$\text{SNR}_{y,j}$	j^{th} sensor Signal-to-Noise ratio
N_p	Number of particles
w^i_k	Importance weight of i^{th} particle
$\widehat{N_{\text{eff}}}$	Approximate effective sample size
N_{thresh}	Threshold for dynamic resampling
N_t	Number of Monte Carlo trials
Acronyms	
AIR	Acoustic Impulse Response
ASIC	Application-Specific Integrated Circuit

AR	Autoregressive
ARMA	Autoregressive Moving Average
BIDS	Blind Identification via Decorrelating Subchannels
BSIS	Block Sequential Importance Sampling
CRB	Cramér-Rao Bound
FIR	Finite Impulse Response
HOS	Higher-Order Statistics
IS	Importance Sampling
KF	Kalman Filter
MAP	Maximum A Posteriori
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MCMC	Markov Chain Monte Carlo
MH	Metropolis Hastings
MIMO	Multiple-Input/Multi-Output
NVR	Noise Variance Ratio
PCRB	Posterior Cramér-Rao Bound
PF	Particle Filter
RBPF	Rao-Blackwellised Particle Filter
SIS	Sequential Importance Sampling
SISO	Single-Input Single-Output
SIMO	Single-Input Multi-Output
SMC	Sequential Monte Carlo
SNR	Signal-to-Noise Ratio
SOS	Second-Order Statistics
TVAR	Time-Varying Autoregressive
VLSI	Very Large-Scale Integration

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

The blind system identification problem arises in many fields, including speech processing, communications, biomedical signal processing, sonar and seismology. An overview of the problem is shown in Figure 1.1, where the objective is to identify the N-dimensional input signal $\mathbf{x}$ or the channel $\mathbf{h}$ given only the J-dimensional output signal $\mathbf{y}$.

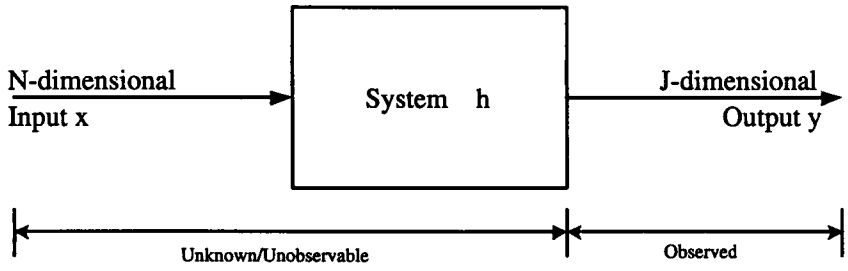


Figure 1.1: Blind system identification problem

The problem is described as blind since *both* the input and the system are unknown. In communications systems design the classical approach to the problem is to periodically transmit a known training sequence, and rely on non-blind methods

to identify the channel. The channel estimate can then be used to equalize the remaining unknown signal outside of the training sequence. Blind signal processing approaches to this problem are attractive since they can increase the available information throughput of a channel by removing the need for training sequence transmission. A more critical problem with the training sequence approach is that in many physical problems, including speech processing, it is not feasible to produce known training signals from the source to perform non-blind methods.

Referring to the n^{th} input signal as $\mathbf{s}_n$, the j^{th} output signal as $\mathbf{y}_j$, and the finite impulse response (FIR) channel from the n^{th} input to j^{th} output as $\mathbf{h}_{j,n}$, the blind system identification problem is modelled as

$$\mathbf{y}_j = \sum_{n=1}^{N} \mathbf{s}_n * \mathbf{h}_{j,n} + \mathbf{w}_j, \tag{1.1}$$

where $*$ denotes the convolution operation and $\mathbf{w}_j$ is the additive noise at the j^{th} output. Blind techniques are classified based on the system dimension and estimation objective. *Blind deconvolution* deals with recovering a single input signal $\mathbf{s}_1$ from either a single-input single-output (SISO) system or a single-input multiple-output (SIMO) system. The recovery of the input signals $\mathbf{s}_n$ from a multiple-input multiple-output (MIMO) system is addressed using *blind source separation* algorithms. In *blind channel identification* methods the goal is to estimate the unknown channels $\mathbf{h}_{j,n}$.

1.2 Application to the Dereverberation of Speech Signals

An important application of blind signal processing techniques is the dereverberation of speech signals in an audio environment. The reverberation of speech signals can

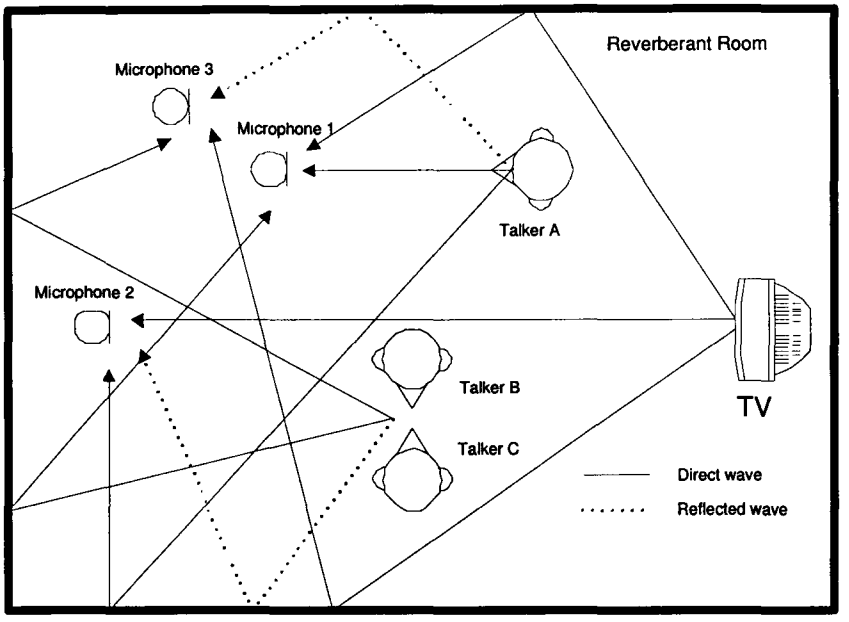


Figure 1.2: Graphical depiction of reverberation

cause significant degradation in the perceptual quality of speech for hands-free telephony, digital hearing aids, music recording, teleconferencing and other audio applications. Reverberation is the effect caused by multiple reflections of sound off the walls, floor and ceiling of the enclosure from the sound sources to the microphones. The resulting annoying acoustical problem is the so-called *barrel effect*, since the speaker sounds distant, like at the bottom of a barrel. The situation is further complicated by the *cocktail party phenomenon* [12] when multiple speech sources are present simultaneously, along with other background noise such as moving fans, television and street traffic. The undesired speech sources and background noise act as interference on the desired speech signal. The reverberation problem is shown in Figure 1.2. Mathematically, the received signal is the convolution of the actual speech sources with the acoustic impulse response (AIR) of the room in additive noise as described in (1.1). An informative introduction to the problem of reverberation cancellation in acoustic environments is presented in [42].

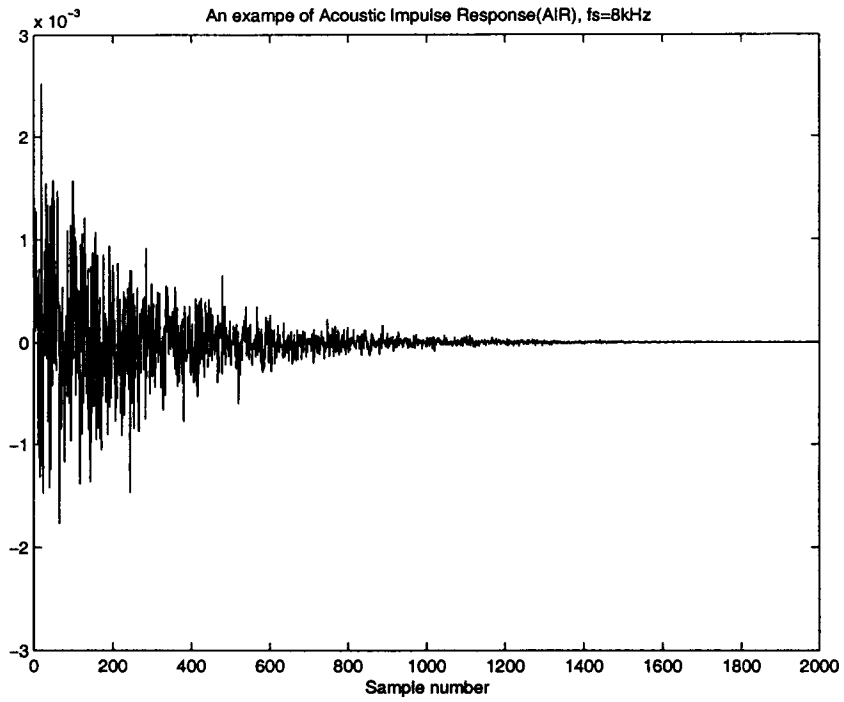


Figure 1.3: Typical AIR in an audio enclosure

As shown in Figure 1.3, typical AIR's have coefficients that decay smoothly towards zero, making the blind channel identification problem ill-conditioned [92]. The AIR of Figure 1.3 was measured using the method of maximal-length sequence identification [53]. The proposed marginalized particle filtering algorithm is developed with an eye towards future application to the dereverberation problem, and as such has the potential to directly recover the source. This approach offers a more computationally stable method for recovery of the source signal. A filter bank implementation of the proposed algorithm for application to the speech dereverberation problem is discussed in Section 5.3. It is noted that the proposed marginalized particle filtering algorithm is not limited to audio applications.

1.3 Blind System Identification Literature Survey

A summary of methods for blind channel identification, deconvolution, and source separation is now presented. An overview of applications and algorithms for the blind system identification problem is available in the review paper [3].

The approach taken by blind channel identification methods is to estimate the unknown channel from the measured signals. The review paper [82] provides a summary of multichannel identification methods. For SISO systems, the use of high-order statistics (HOS) is required to handle the case of a non-minimum phase channel [32]. The use of HOS requires the assumptions that the input signal is non-Gaussian, and that the channel statistics are stationary over the large sample size needed to compute the high-order moments of the output signal [75]. The use of multiple sensors in a SIMO system allows for channel identification, up to the inherent complex scaling ambiguity, based on second-order statistics (SOS) [83][84] by exploiting the diversity in the channels [82]. SOS methods have the advantage that in general they require a smaller sample size to converge compared to HOS, but may require stronger conditions on the channel or source [75]. Subspace methods for the SIMO [1] and the MIMO [37] cases perform channel identification based on subspace structure present in the SOS of the output signals. An alternative approach using the SOS of oversampled output signals is based on the method of linear prediction [2][80]. The problem can also be formulated using the maximum likelihood approach, including the two-step maximum likelihood method appearing in [45]. If an estimate of the input source is required using channel identification methods, an inverse of channel estimate can be applied to the measured signals to recover an estimate of the original source. This is a result of the Bezout identity described in [82] for the FIR inverse of a SIMO system under the condition that the channels are coprime (do not share common zeros). A well-known approach for the MIMO case is the blind identification via decorrelating subchannels (BIDS) method [46][47].

Blind deconvolution methods are constructed to directly recover an estimate of

the source signal. Blind deconvolution algorithms for many communication systems make use of the finite alphabet nature of the input signal [16][57]. Methods based on subspace approaches also appear in [85][59]. For the general blind system identification problem, for example in the case of continuous speech sources, the assumption of finite alphabet no longer holds. A Bayesian approach for blind deconvolution of SIMO [44] and MIMO [43] systems by Hopgood is applied to the audio reverberation problem. The algorithm operates by exploiting the differences between the assumed nonstationary AR sources and stationary all-pole channels.

Blind source separation methods address the case of estimating multiple input signals from multiple outputs. A review of blind source separation methods is available in [14]. One of the earliest approaches to the problem for instantaneous mixing models was the Independent Component Analysis (ICA) method presented in [20]. Algorithms based on information theory appear in [8]. A maximum likelihood formulation can be implemented using joint diagonalization procedures [69]. Recent work has focused on convolutive mixing models which arise in practice for mixtures of speech signals in reverberant audio environments [68][74][70]. A source separation approach based on particle filtering appeared in [34],[6].

1.4 Proposed Solution

The proposed algorithm uses Sequential Monte Carlo (SMC) methods (otherwise known as particle filtering) to solve the blind system identification problem. Particle filters provide a numerical approximation to the optimal Bayesian solution for the nonlinear/non-Gaussian sequential state estimation problem. Tutorials on SMC methods are presented in [7] [24], while the state-of-the-art is available in [28][26][61]. The use of particle filters in signal processing was in large part prompted by the introduction of the resampling step into the sequential importance sampling (SIS) procedure [36]. Along with recent advances in computational power, active research

in particle filtering methods for nonlinear/non-Gaussian problems has led to application in a wide variety of technical fields, including communication systems [64][40][31], target tracking [65][9][13] and speech processing [30][88],[6]. A particle filtering approach can result in significant computational complexity; however, particle filters lend themselves well to a parallel implementation. A VLSI ASIC implementation of a real-time particle filter architecture appeared recently in [19], and methods for real-time implementation are discussed in [55].

For the specific blind system identification problem presented in this thesis, the state space model uses a time-varying autoregressive (AR) model for the sources, and a time-varying finite impulse response (FIR) model for the channel. The multi-sensor measurements result from the convolution of the sources with the channels in the presence of additive noise. A Bayesian framework for solving the sequential state estimation problem for the dynamical model is developed using the posterior distribution of the unknown sources, AR coefficients, FIR coefficients and noise variances.

A particle filter approach is used to recursively update the posterior distribution of the nonlinear state space model. The particle filter is efficiently implemented by marginalizing out the AR and FIR coefficients from the posterior distribution using the Rao-Blackwellisation procedure. This reduces the particle filter problem to that of estimating the sources. The AR and FIR coefficients can then be estimated using the optimal Kalman filter. Analytical expressions for maximum a posteriori (MAP) estimates of the noise variances are also developed.

In addition to the proposed particle filtering approach that uses the classical SIS method, a novel formulation called block sequential importance sampling (BSIS) is introduced to exploit the structure of the convolution state space model.

1.5 Outline of Thesis

- Chapter 1 introduces the blind system identification problem. The application of the dereverberation of speech signals in an audio environment is then presented highlighting the design challenges that must be addressed in a dereverberation algorithm. A brief literature survey of blind deconvolution, channel identification, and source separation methods is also included. The key features of the proposed approach are then outlined.

- Chapter 2 provides background material on Bayesian filtering methods. It outlines Bayesian theory, and its application to the sequential state estimation problem. The fundamentals of the Kalman filter for linear-Gaussian state space models, and the Monte Carlo-based particle filter for nonlinear/non-Gaussian models are shown. The use of the Posterior Cramér-Rao bound (PCRB) as a theoretical performance benchmark for nonlinear filtering problems is also introduced.

- Chapter 3 describes the proposed marginalized particle filtering approach to the blind system identification problem. The state space model and statistical assumptions are first introduced. The structure of marginalized particle filter is then developed. The remaining sections develop the particle filtering algorithm for nonlinear estimation of the sources, the Kalman filtering algorithm for estimation of the AR and FIR coefficients, and the MAP expressions for estimating the noise variances.

- Chapter 4 illustrates the performance of the algorithm using simulation results. This begins with the derivation of the PCRB based on the proposed state space model. The definitions of the parameter settings and performance measures used in the simulations follows. The proposed algorithm is tested to demonstrate the performance. Comparison of the results to the PCRB is also included.

- Chapter 5 summarizes with conclusions and contributions to the scientific literature. Suggestions for future research are also provided, including a detailed introduction to the block sequential importance sampling formulation of the particle filter.

Chapter 2

Bayesian Filtering Background

Relevant background material on Bayesian filtering is summarized in this chapter. The sequential state estimation problem is introduced, and a Bayesian approach to the problem is developed. The optimal Kalman filter for linear-Gaussian models, and the Monte Carlo-based particle filter for nonlinear/non-Gaussian models are described. The Posterior Cramér-Rao bound (PCRB) is introduced to provide a theoretical performance benchmark for the Bayesian sequential state estimation problem.

2.1 Bayesian Theory

Bayes' theorem for the random parameter $\mathbf{x}$ and measurement $\mathbf{y}$ may be stated as

$$p(\mathbf{x},\mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}), \tag{2.1}$$

where $p(\mathbf{x})$ and $p(\mathbf{y})$ are prior distributions, $p(\mathbf{y}|\mathbf{x})$ is the likelihood, and $p(\mathbf{x}|\mathbf{y})$ is the posterior distribution. The posterior can also be written as

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{2.2}$$

where the normalizing constant $p(\mathbf{y})$ is

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}. \tag{2.3}$$

Using a Bayesian framework, the posterior distribution captures all statistical information about the random parameter $\mathbf{x}$ from the measurements and prior knowledge. As such, the posterior distribution can be used to generate estimates of the unknown parameter. For example, the maximum a posteriori (MAP) estimate is the parameter value that maximizes the posterior distribution:

$$\hat{\mathbf{x}}^{\text{MAP}} = \arg\max p(\mathbf{x}|\mathbf{y}). \tag{2.4}$$

The minimum mean square error (MMSE) estimates of the parameter may be also be computed:

$$\begin{aligned} \hat{\mathbf{x}}^{\text{MMSE}} &= \mathbb{E}_{p(\mathbf{x}|\mathbf{y})}\{\mathbf{x}\} \\ &= \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}. \end{aligned} \tag{2.5}$$

In addition, a measure of belief in the estimates can be determined from the posterior distribution, for example in the form of confidence intervals.

An important advantage of using a Bayesian formulation is that parameters can be marginalized out of the posterior distribution. This can lead to more efficient and better performing estimation algorithms. The use of marginalization plays a key role in the proposed blind identification approach.

A Bayesian approach requires the use of prior information for the parameters. This need not be a restriction, in that non-informative priors may be used when reasonable previous prior knowledge of the parameter values is not available [10].

2.2 Bayesian Sequential State Estimation

The sequential state estimation problem is to recursively determine the state sequence $\mathbf{x}_k$ which is assumed to evolve according to the first-order Markov process equation

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}), \tag{2.6}$$

where $\mathbf{f}_k$ is the state transition function and $\mathbf{v}_{k-1}$ is the process noise sequence. The state is estimated from the measurement (observation) sequence $\mathbf{y}_k$ given by

$$\mathbf{y}_k = \mathbf{g}_k(\mathbf{x}_k, \mathbf{w}_k), \tag{2.7}$$

where $\mathbf{g}_k$ is the measurement function and $\mathbf{w}_k$ is the measurement noise. The current measurement $\mathbf{y}_k$ is assumed dependent on only the current state $\mathbf{x}_k$ and measurement noise $\mathbf{w}_k$ at time k.

In a Bayesian approach to sequential state estimation, the filtered posterior distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ of the current state conditioned on all the measurements is recursively computed from $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$. First, the predicted posterior distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ is determined from the Chapman-Kolmogorov equation [7]:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1}, \tag{2.8}$$

where the transition prior $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is determined from the process equation (2.6) and the statistics of $\mathbf{v}_{k-1}$. The predicted distribution is then updated using the new measurement $\mathbf{y}_k$ by applying Bayes' theorem:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})}, \tag{2.9}$$

where the likelihood distribution $p(\mathbf{y}_k|\mathbf{x}_k)$ follows from the measurement model (2.7) and the statistics of $\mathbf{w}_k$. The normalizing constant is given by

$$p(\mathbf{y}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k. \tag{2.10}$$

An optimal Bayesian filter is one for which the required integrals for the prediction and update steps can be computed exactly. In general, however, these integrations are analytically intractable. The next section presents the Kalman filter, which is optimal when the state space model is restricted to be linear with independent Gaussian noise and initial state processes [4]. This is followed by the introduction of the particle filter, a Monte Carlo-based approach to numerically approximate the optimal solution for more general nonlinear and non-Gaussian state space models [26].

2.3 Kalman Filtering

Under the conditions that the process noise $\mathbf{v}_{k-1}$, measurement noise $\mathbf{w}_k$ and initial state $\mathbf{x}_0$ are independent Gaussian variables, and the state transition $\mathbf{f}_k(\cdot,\cdot)$ and measurement $\mathbf{g}_k(\cdot,\cdot)$ functions are linear, the state space model is

$$\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{v}_{k-1}, \tag{2.11}$$

$$\mathbf{y}_k = \mathbf{G}_k\mathbf{x}_k + \mathbf{w}_k, \tag{2.12}$$

where the matrices $\mathbf{F}_k$, $\mathbf{G}_k$ are known, along with the state noise covariance $\mathbf{\Sigma}_{v,k-1}$ and measurement noise covariance $\mathbf{\Sigma}_{w,k}$.

For this linear-Gaussian system, the optimal Bayesian solution for the sequential state estimation problem can be computed analytically using the Kalman filter. The resulting posterior distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is Gaussian, and can therefore be represented completely by its mean and covariance. Starting with the filtered distribution at time $k-1$ with mean $\hat{\mathbf{x}}_{k-1|k-1}$ and covariance $\mathbf{\Phi}_{k-1|k-1}$ given by

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{\Phi}_{k-1|k-1}), \tag{2.13}$$

the integral for the prediction step of the optimal Bayesian solution in (2.8) reduces to computing the mean $\hat{\mathbf{x}}_{k|k-1}$ and covariance $\mathbf{\Phi}_{k|k-1}$ of the predicted posterior distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ given by [58]

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{\Phi}_{k|k-1}), \tag{2.14}$$

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}, \tag{2.15}$$

$$\mathbf{\Phi}_{k|k-1} = \mathbf{F}_k\mathbf{\Phi}_{k-1|k-1}\mathbf{F}_k^{\mathrm{T}} + \mathbf{\Sigma}_{v,k-1}. \tag{2.16}$$

The update step in (2.9) produces the mean $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{\Phi}_{k|k}$ of the filtered posterior at time k:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{\Phi}_{k|k}), \tag{2.17}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k(\mathbf{y}_k - \mathbf{G}_k\hat{\mathbf{x}}_{k|k-1}), \tag{2.18}$$

$$\mathbf{\Phi}_{k|k} = \mathbf{\Phi}_{k|k-1} - \mathbf{W}_k\mathbf{G}_k\mathbf{\Phi}_{k|k-1}, \tag{2.19}$$

where the Kalman gain is defined as

$$\mathbf{W}_k = \mathbf{\Phi}_{k|k-1}\mathbf{G}_k^{\mathrm{T}}\mathbf{Q}_k^{-1}, \tag{2.20}$$

and, for the innovation defined as $\mathbf{y}_k - \mathbf{G}_k\hat{\mathbf{x}}_{k|k-1}$, the innovation covariance is

$$\mathbf{Q}_k = \mathbf{G}_k\mathbf{\Phi}_{k|k-1}\mathbf{G}_k^{\mathrm{T}} + \mathbf{\Sigma}_{w,k}. \tag{2.21}$$

2.4 Particle Filtering

We now consider the case when the prediction (2.8) and update (2.9) steps of the optimal Bayesian filter cannot be computed analytically. For a nonlinear and/or non-Gaussian state space model, as in the case of the blind system identification problem, suboptimal approaches are required. We now introduce the particle filter which uses a Monte-Carlo based approach to numerically approximate the posterior distribution.

2.4.1 Monte Carlo Integration

The approach to approximating the intractable integrations is based on the method of Monte Carlo integration [77]. Consider the expectation of an arbitrary function $f(\cdot)$ with respect to the posterior distribution:

$$\mathbb{E}_{p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})}(f) = \int f(\mathbf{x}_{1:k})p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})d\mathbf{x}_{1:k}. \tag{2.22}$$

If N_p identically and independently distributed (*i.i.d.*) samples $\mathbf{x}_{1:k}^i$ can be drawn from the distribution $p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})$ then a discrete approximation to the posterior distribution is

$$p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) \approx \frac{1}{N_p}\sum_{i=1}^{N_p}\delta(\mathbf{x}_{1:k} - \mathbf{x}_{1:k}^i). \tag{2.23}$$

Using this representation, the Monte Carlo method approximates the integral in (2.22) with the summation

$$\hat{\mathbb{E}}_{p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})}(f) = \frac{1}{N_p}\sum_{i=1}^{N_p}f(\mathbf{x}_{1:k}^i). \tag{2.24}$$

Results presented in [25] show that this numerical approximation demonstrates almost sure convergence to the true value as $N_p \rightarrow \infty$.

2.4.2 Importance Sampling

The Monte Carlo sampling method cannot be used on its own to approximate the optimal Bayesian filter integrals since, in general for nonstandard multidimensional distributions, the posterior is not easy to draw *i.i.d.* samples from. An importance function $q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})$ that is easy to sample from is introduced into (2.22) as shown:

$$\mathbb{E}_{p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})}(f) = \int f(\mathbf{x}_{1:k}) \frac{p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})}{q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})} q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) d\mathbf{x}_{1:k}. \tag{2.25}$$

By selecting N_p *i.i.d.* samples from $q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})$, the discrete weighted sum approximation to the posterior distribution is

$$p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) \approx \frac{1}{N_p} \sum_{i=1}^{N_p} w_k^{*(i)} \delta(\mathbf{x}_{1:k} - \mathbf{x}_{1:k}^i), \tag{2.26}$$

where the set of samples $\{\mathbf{x}_{1:k}^i, i = 1, \ldots, N_p\}$ are referred to as *particles*. The associated true *importance weights* $\{w_k^{*(i)}, i = 1, \ldots, N_p\}$ are defined as

$$\begin{aligned} w_k^{*(i)} &= \frac{p(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})}{q(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})} \\ &= \frac{p(\mathbf{y}_{1:k}|\mathbf{x}_{1:k}^i) p(\mathbf{x}_{1:k}^i)}{p(\mathbf{y}_{1:k}) q(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})}. \end{aligned} \tag{2.27}$$

In general, it is not possible to evaluate the true importance weights since the normalizing constant $p(\mathbf{y}_{1:k})$ can be difficult to compute. Instead, the importance weights are first evaluated as

$$\begin{aligned} w_k^{(i)} &\propto \frac{p(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})}{q(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})} \\ &= \frac{p(\mathbf{y}_{1:k}|\mathbf{x}_{1:k}^i) p(\mathbf{x}_{1:k}^i)}{q(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})}, \end{aligned} \tag{2.28}$$

and then normalized to ensure the posterior distribution integrates to one using the following:

$$w_k^i = \frac{w_k^i}{\sum_{j=1}^{N_p} w_k^j}. \tag{2.29}$$

As shown in [25], the resulting discrete representation of the posterior distribution in terms of the normalized importance weights in (2.29) is

$$p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) \approx \sum_{i=1}^{N_p} w_k^i \delta(\mathbf{x}_{1:k} - \mathbf{x}_{1:k}^i), \tag{2.30}$$

where the true importance weights $w_k^{*(i)}$ in (2.26) are replaced by $N_p w_k^i$. Using importance sampling (IS), the Monte Carlo approximation to the integral in (2.25) is then given by

$$\hat{\mathbb{E}}_{p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})}(f) = \sum_{i=1}^{N_p} f(\mathbf{x}_{1:k}^i) w_k^i. \tag{2.31}$$

For example, the filtered MMSE estimate of the state $\mathbf{x}_k$ given the measurements $\mathbf{y}_{1:k}$ can be approximated from (2.5) using the set of particles and importance weights:

$$\begin{aligned} \hat{\mathbf{x}}_k^{\text{MMSE}} &= \mathbb{E}_{p(\mathbf{x}_k|\mathbf{y}_{1:k})}\{\mathbf{x}_k\} \\ &\approx \sum_{i=1}^{N_p} \mathbf{x}_k^i \cdot w_k^i. \end{aligned} \tag{2.32}$$

In the ideal case the importance function is exactly equal to the posterior distribution, which for the true importance weights in (2.27) results in

$$w_k^{*(i)} = 1, \tag{2.33}$$

$$\text{var}_{q(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})}(w_k^{*(i)}) = 0. \tag{2.34}$$

The variance of the importance weights can therefore be used as a measure of optimality of the importance sampling method, and in subsequent sections is used in selecting the best choice of the importance function and for dynamic scheduling of resampling steps.

2.4.3 Sequential Importance Sampling

For the sequential state estimation problem, the method of sequential importance sampling (SIS) is used to recursively generate particles $\mathbf{x}_k^i$ and update the importance

weights w_k^i. First, a recursion for the posterior distribution is found by applying Bayes' theorem:

$$\begin{aligned} p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) &= \frac{p(\mathbf{y}_k|\mathbf{x}_{1:k}, \mathbf{y}_{1:k-1})p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_{1:k}, \mathbf{y}_{1:k-1})p(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} p(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1}) \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} p(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1}). \end{aligned} \tag{2.35}$$

The last line follows since the measurement $\mathbf{y}_k$ does not depend on the past states $\mathbf{x}_{1:k-1}$ or measurements $\mathbf{y}_{1:k-1}$, and the state is first-order Markov depending on only $\mathbf{x}_{k-1}$.

Applying Bayes' theorem to the importance function results in

$$q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) = q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k})q(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k}). \tag{2.36}$$

An importance function containing the factor $q(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k})$ implies that the particles for the past states $\mathbf{x}_{1:k-1}$ are regenerated at time k based on the new measurement $\mathbf{y}_k$. This introduces a significant computational burden since it requires storage of the entire particle history $\mathbf{x}_{1:k-1}^i$ for each particle. To avoid this problem, the importance function is restricted to satisfy the form [25]

$$q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) = q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k})q(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1}). \tag{2.37}$$

The approximation is that the *smoothed* importance distribution $q(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k})$ for the past states $\mathbf{x}_{1:k-1}$ incorporating the new measurement is replaced by the *filtered* distribution $q(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1})$. As a result, the past particles $\mathbf{x}_{1:k-1}^i$ do not have to be modified since they are only dependent on the measurements up to time $k-1$. The factorization also specifies that particles $\mathbf{x}_k^i$ are drawn from the distribution $q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k})$. The new particles $\mathbf{x}_k^i$ are then appended to the history of particles $\mathbf{x}_{1:k-1}^i$ to form the particle trajectories $\mathbf{x}_{1:k}^i = \{\mathbf{x}_{1:k-1}^i, \mathbf{x}_k^i\}$. Typically, as is shown in Section 2.4.5, the importance function $q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k})$ is selected to be only dependent on the most recent state $\mathbf{x}_{k-1}$, and possibly the current measurement $\mathbf{y}_k$. This

simplifies the implementation by requiring only storage of $\mathbf{x}_{k-1}^i$ instead of the complete particle history. It is noted that the proposed block sequential importance sampling method uses an alternative factorization property of the importance function.

Using the recursion for the posterior distribution in (2.35) and the factorization of the importance distribution in (2.37), the recursive update for the importance weights is

$$\begin{aligned} w_k^i &\propto \frac{p(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})}{q(\mathbf{x}_{1:k}^i|\mathbf{y}_{1:k})} \\ &\propto \frac{p(\mathbf{y}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{1:k-1}^i,\mathbf{y}_{1:k})} \frac{p(\mathbf{x}_{1:k-1}^i|\mathbf{y}_{1:k-1})}{q(\mathbf{x}_{1:k-1}^i|\mathbf{y}_{1:k-1})} \\ &\propto w_{k-1}^i \frac{p(\mathbf{y}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{1:k-1}^i,\mathbf{y}_{1:k})}. \end{aligned} \tag{2.38}$$

The importance weights $\{w_0^i, i = 1, 2, \ldots, N_p\}$ are initialized to value 1. At each time step, the importance weights are normalized using (2.29).

2.4.4 Resampling

SIS particle filtering algorithms suffer from the problem of importance weight degeneracy, in which after a few iterations of the recursion only one particle has a significant normalized importance weighting and the rest are close to zero. In [25], it is shown that the degeneracy problem cannot be avoided since the variance of the importance weights can only increase over time. The introduction of the resampling step in [36] made SIS methods practical by effectively reducing the weight degeneracy problem. This is accomplished by performing a resampling step after the weight update in which particles with large importance weights are duplicated while particles with small weights are removed. This leads to a more efficient particle filter since computational resources are not wasted on updating particles that contribute little to the numerical approximation of the posterior distribution. A graphical presentation of the resampling step for a one-dimensional state x_k and $N_p = 10$ particles is shown in Figure 2.1, based partly on [22]. The resampling step removes the particles x_k^i

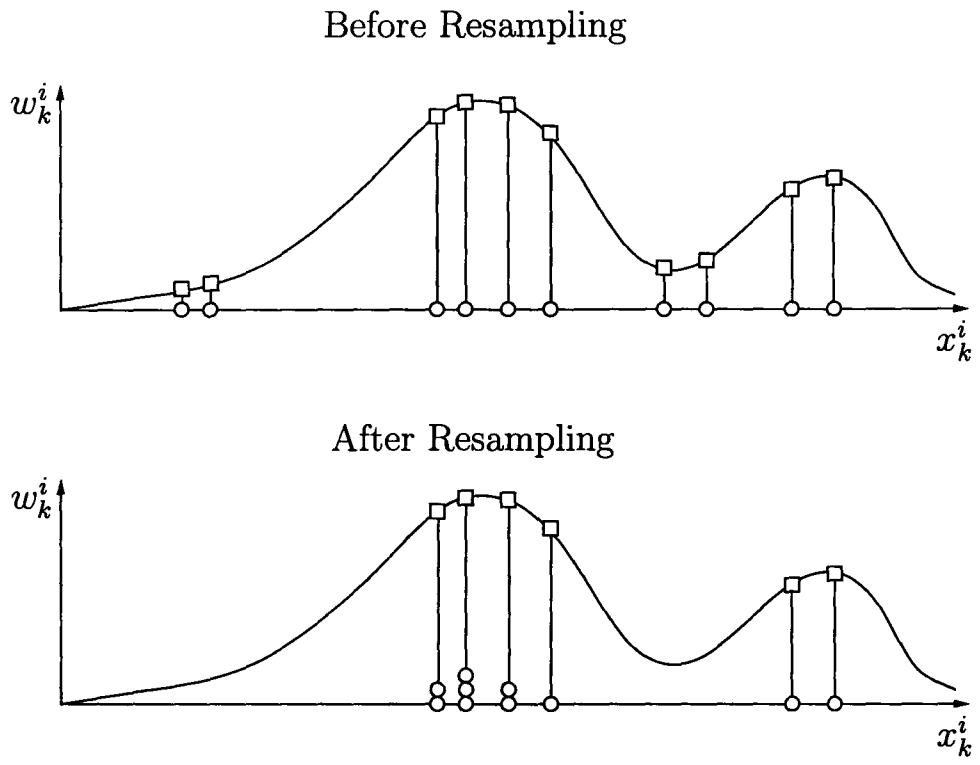


Figure 2.1: Graphical presentation of the resampling step

(circles) associated with small importance weights w_k^i (squares), and duplicates particles with large importance weights. The particle filter representation consisting of the random measure $\{x_k^i, w_k^i, i = 1, 2, \ldots, N_p\}$ is seen to be a discrete approximation to the true continuous posterior distribution curve shown.

Resampling algorithms provide ways of drawing *i.i.d.* samples $\{\tilde{\mathbf{x}}_{1:k}^j, j = 1, 2, \ldots, N_p\}$ from the discrete approximation to the posterior distribution in (2.30). As a result, the new particles are selected with probability

$$Pr(\tilde{\mathbf{x}}_{1:k}^j = \mathbf{x}_{1:k}^i) = w_k^i. \tag{2.39}$$

The original particles are replaced with the resampled particles using

$$\mathbf{x}_{1:k}^i = \tilde{\mathbf{x}}_{1:k}^i, \tag{2.40}$$

and the importance weights are all set equal

$$w_k^i = \frac{1}{N_p}. \tag{2.41}$$

A variety of resampling schemes are available in the literature, including multinomial [36], residual [61], and systematic resampling [54]. A comparison of the complexity and performance of these resampling algorithms is presented in [18].

A measure of the degeneracy of the SIS algorithm is the effective sample size N_{eff} [61] given by

$$N_{\text{eff}} = \frac{N_p}{1 + \text{var}(w_k^{*(i)})}. \tag{2.42}$$

The variance of the importance weights is non-negative so that $N_{\text{eff}} \leq N_p$, and also from (2.34) in the ideal case the variance is zero which means that small values of the effective sample size indicate algorithm degeneracy. From the discussion in Section 2.4.2, in general the true importance weights $w_k^{*(i)}$ cannot be evaluated, and so the following approximation to the effective sample size is used:

$$\widehat{N_{\text{eff}}} = \frac{1}{\sum_{i=1}^{N_p} (w_k^i)^2}. \tag{2.43}$$

The use of the approximate effective sample size allows resampling to occur only when the algorithm displays degeneracy. This is accomplished by resampling whenever $\widehat{N_{\text{eff}}}$ is below a fixed threshold N_{thresh}.

2.4.5 Selection of the Importance Function

The selection of the importance function is an important factor in determining the performance and efficiency of the SIS particle filter. In practice, there are an unlimited number of choices for the importance distribution. Three of the most common are briefly introduced here. The first importance function presented is the transition prior given by

$$q(\mathbf{x}_k | \mathbf{x}_{1:k-1}^i, \mathbf{y}_{1:k}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}^i). \tag{2.44}$$

The resulting weight update equation simplifies to multiplication by the likelihood

$$w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_k^i). \tag{2.45}$$

A second popular choice is the optimal importance function. For the filtered posterior distribution, it is defined as the function which minimizes the variance of the true importance weights conditional upon $\mathbf{x}_{k-1}^i$ and $\mathbf{y}_k$. In [28], the optimal importance function which results in a variance of zero is shown to be

$$\begin{aligned} q(\mathbf{x}_k|\mathbf{x}_{1:k-1}^i, \mathbf{y}_{1:k}) &= p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k) \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}^i)}{p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)}. \end{aligned} \tag{2.46}$$

The corresponding weight update becomes

$$w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i). \tag{2.47}$$

From the recursion for the posterior distribution in (2.35), it is noted that the optimal importance function is equal to the update factor on the previous posterior distribution $p(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1})$. It can be seen that the benefit of the optimal importance function is that it incorporates information from the current measurement $\mathbf{y}_k$ in generating the particles. This is at the expense of additional computation compared to the use of the transition prior as the importance function. In many cases, the optimal importance function cannot be derived analytically or is difficult to sample from, and instead approximations are used. For example, the method of local linearisation can be used to approximate the optimal importance function with a Gaussian distribution [25].

The third approach is the auxilliary particle filter [71] which uses the transition prior to generate particles, but also incorporates the current measurement $\mathbf{y}_k$ into the resampling step at time $k-1$. For more details see the description in [41].

A pictorial overview of the particle filtering algorithm introduced in this section is shown in Figure 2.2, based partly on [24].

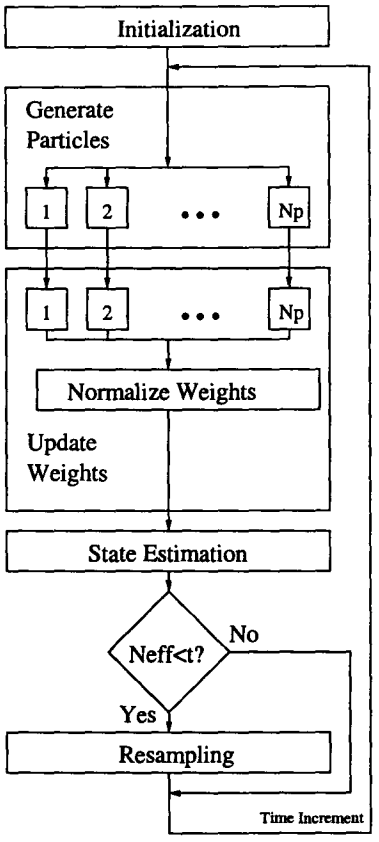


Figure 2.2: Particle Filtering Algorithm Structure

2.5 Posterior Cramér-Rao bound

For analyzing the performance of estimation algorithms, lower bounds based on information limits are developed to indicate theoretical performance limitations. For the estimation of a set of *nonrandom* parameters $\mathbf{x}_{1:k}$ given the set of measurements $\mathbf{y}_{1:k}$, the Cramér-Rao bound (CRB) provides the lower bound on the covariance of an *unbiased* estimator $\hat{\mathbf{x}}_{1:k}(\mathbf{y}_{1:k})$ [78]:

$$\mathbf{M} = \mathbb{E}\left\{[\hat{\mathbf{x}}_{1:k}(\mathbf{y}_{1:k}) - \mathbf{x}_{1:k}]^{\mathrm{T}}[\hat{\mathbf{x}}_{1:k}(\mathbf{y}_{1:k}) - \mathbf{x}_{1:k}]\right\} \succeq \mathbf{J}_{1:k}^{-1}. \tag{2.48}$$

The matrix inequality implies that the matrix $\mathbf{M} - \mathbf{J}_{1:k}^{-1}$ must be positive semidefinite. This also specifies that the mean square error (MSE) of the n^{th} component of $\mathbf{x}_{1:k}$ is

bounded by the n^{th} diagonal element of lower bound:

$$\mathrm{MSE}(\hat{\mathbf{x}}_{1:k}[n]) = \mathbb{E}\left\{[\hat{\mathbf{x}}_{1:k}[n](\mathbf{y}_{1:k}) - \mathbf{x}_{1:k}[n]]^{\mathrm{T}}[\hat{\mathbf{x}}_{1:k}[n](\mathbf{y}_{1:k}) - \mathbf{x}_{1:k}[n]]\right\} \geq \mathbf{J}_{1:k}^{-1}[n,n]$$

The lower bound is given by the inverse of the Fisher information matrix $\mathbf{J}_{1:k}$, which is determined from the likelihood $p(\mathbf{y}_{1:k}|\mathbf{x}_{1:k})$:

$$\mathbf{J}_{1:k} = -\mathbb{E}\left\{\Delta_{\mathbf{x}_{1:k}}^{\mathbf{x}_{1:k}} \log p(\mathbf{y}_{1:k}|\mathbf{x}_{1:k})\right\}. \tag{2.49}$$

The expectation is taken over the measurements $\mathbf{y}_{1:k}$. The first-order (∇) and second-order (Δ) partial derivative operators are defined for the vector $\boldsymbol{\alpha}$ of dimension n_α as

$$\nabla_{\boldsymbol{\alpha}} = \left[\frac{\partial}{\partial \boldsymbol{\alpha}[1]}, \frac{\partial}{\partial \boldsymbol{\alpha}[2]} \cdots, \frac{\partial}{\partial \boldsymbol{\alpha}[n_\alpha]}\right]^{\mathrm{T}}, \tag{2.50}$$

$$\Delta_{\boldsymbol{\beta}}^{\boldsymbol{\alpha}} = \nabla_{\boldsymbol{\beta}} \nabla_{\boldsymbol{\alpha}}^{\mathrm{T}}. \tag{2.51}$$

The Posterior Cramér-Rao bound (PCRB) introduced in [86] provides the lower bound when the parameter $\mathbf{x}_{1:k}$ is a *random* variable. In this case, the Fisher information matrix is determined with respect to the joint distribution $p(\mathbf{x}_{1:k}, \mathbf{y}_{1:k})$:

$$\mathbf{J}_{1:k} = -\mathbb{E}\left\{\Delta_{\mathbf{x}_{1:k}}^{\mathbf{x}_{1:k}} \log p(\mathbf{x}_{1:k}, \mathbf{y}_{1:k})\right\}, \tag{2.52}$$

and the expectation is taken over the measurements $\mathbf{y}_{1:k}$ and the random variable $\mathbf{x}_{1:k}$.

For the sequential state estimation problem it is of interest to recursively determine the Fisher information submatrix $\mathbf{J}_k$ which determines the lower bound on the estimation covariance of the N-dimensional state $\mathbf{x}_k$ given the measurements $\mathbf{y}_{1:k}$. For the discrete-time nonlinear filtering problem, [81] derives an efficient recursion for $\mathbf{J}_k$. However, alternative methods are required to handle the case when the transition prior $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ is singular. This is the case for the proposed state space model, and also frequently in tracking applications, in which the transition prior contains Dirac delta measures. In [81], the approach to the singular case is dealt with by introducing

regularization noise into the system. An alternative approach for the singular case is also presented in [89] for nonlinear systems with unknown constant parameters. The approach taken here follows from [9] (also summarized in Chapter 15 of [26]) in which a recursion for the inverse of the Fisher information, denoted by $\mathbf{P}_k = \mathbf{J}_k^{-1}$, is derived. This avoids the problem of requiring inversions of singular noise covariance matrices from the transition prior. The recursion is provided for the one-step-ahead predictor, that is, the estimation of the state $\mathbf{x}_k$ given the measurements $\mathbf{y}_{1:k-1}$. This method was found to provide the most straightforward derivation of the PCRB for the proposed state space model, as compared to [81],[89].

In the following it is assumed the discrete-time nonlinear state space model is of the form shown in (2.6) and (2.7). The lower bound for the one-step-ahead predictor is:

$$\mathbb{E}\left\{[\hat{\mathbf{x}}_k(\mathbf{y}_{1:k-1}) - \mathbf{x}_k]^{\mathrm{T}}[\hat{\mathbf{x}}_k(\mathbf{y}_{1:k-1}) - \mathbf{x}_k]\right\} \succeq \mathbf{P}_k. \tag{2.53}$$

The recursion for $\mathbf{P}_k$, under the assumption that the prior distribution tends to zero at infinity but with no requirement of the estimator being unbiased, is given by [26]

$$\mathbf{P}_{k+1} = \mathbf{F}_k(\mathbf{P}_k^{-1} + \mathbf{R}_k^{-1})^{-1}\mathbf{F}_k^{\mathrm{T}} + \mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^{\mathrm{T}}, \tag{2.54}$$

where the following matrices, all of dimension $N \times N$, are defined as:

$$\mathbf{F}_k^{\mathrm{T}} = \mathbb{E}\{\nabla_{\mathbf{x}_k}\mathbf{f}_k^{\mathrm{T}}(\mathbf{x}_k, \mathbf{v}_k)\}, \tag{2.55}$$

$$\mathbf{R}_k^{-1} = \mathbb{E}\{-\Delta_{\mathbf{x}_k}^{\mathbf{x}_k}\log p(\mathbf{y}_k|\mathbf{x}_k)\}, \tag{2.56}$$

$$\mathbf{G}_k^{\mathrm{T}} = \mathbb{E}\{\nabla_{\mathbf{v}_k}\mathbf{f}_k^{\mathrm{T}}(\mathbf{x}_k, \mathbf{v}_k)\}, \tag{2.57}$$

$$\mathbf{Q}_k^{-1} = \mathbb{E}\{-\Delta_{\mathbf{v}_k}^{\mathbf{v}_k}\log p(\mathbf{v}_k)\}, \tag{2.58}$$

$$\mathbf{P}_0^{-1} = \mathbb{E}\{-\Delta_{\mathbf{x}_0}^{\mathbf{x}_0}\log p(\mathbf{x}_0)\}. \tag{2.59}$$

In practice, it may not be possible to evaluate the expectations analytically. Monte Carlo techniques can be employed to produce numerical estimates to the expectations by averaging *i.i.d.* statistical realizations of the system model [26].

Chapter 3

Marginalized Particle Filtering for Blind System Identification

The state space model and statistical assumptions used to develop the joint posterior distribution for the blind system identification problem are introduced. An efficient Bayesian solution to the sequential state estimation problem is then developed using a marginalized particle filtering algorithm. The resulting nonlinear estimation problem for the sources is implemented using a particle filter, the FIR and AR coefficients are estimated using a Kalman filter after being marginalized out of the posterior distribution, and analytical MAP estimates for the unknown noise variances are developed. A discussion of sufficient identifiability conditions is also provided.

3.1 Bayesian Formulation

3.1.1 State Space Model

The state space model under consideration for the blind system identification problem is shown graphically in Figure 3.1. We now develop the dynamical equations and variable descriptions for the state space model.

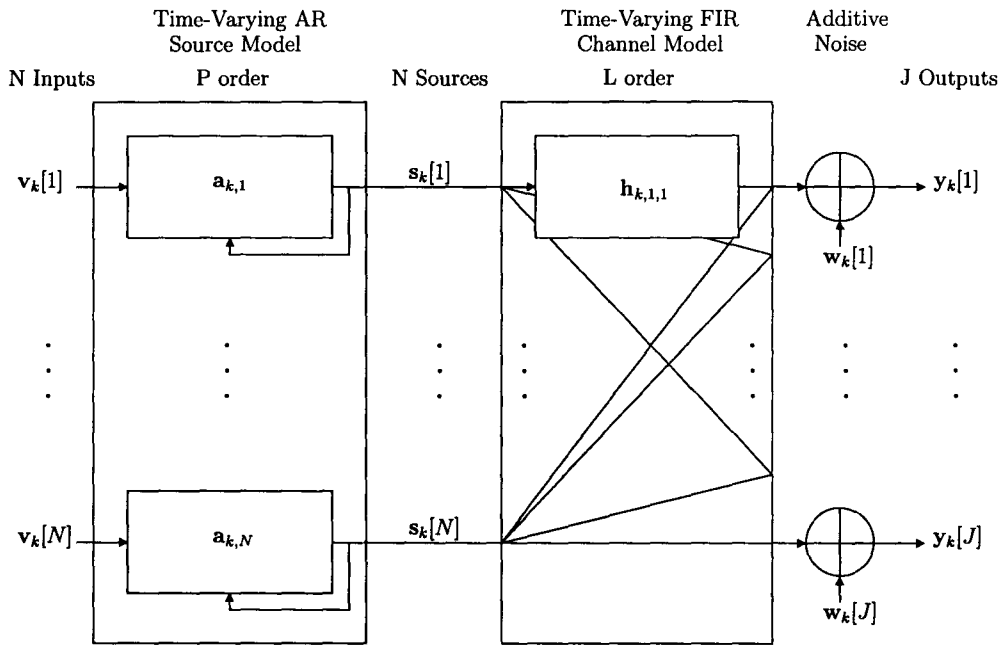


Figure 3.1: Blind system identification state space model

The n^{th} source $\mathbf{s}_k[n]$ is assumed to evolve according to the following P-order time-varying autoregressive (TVAR) model:

$$\mathbf{s}_k[n] = \mathbf{a}_{k,n}^{\mathrm{T}} \mathbf{s}_{P,k-1,n} + \mathbf{v}_{k-1}[n]. \tag{3.1}$$

The source noise $\mathbf{v}_{k-1}[n] \in \mathbb{R}$ is assumed to be Gaussian distributed with mean zero and unknown variance $\sigma_{v,n}^2$. The noise variances are assumed to be independent between sources. The source vector $\mathbf{s}_{P,k-1,n} \in \mathbb{R}^{P\times 1}$ is composed of the most recent P samples at time $k-1$ for the n^{th} source:

$$\mathbf{s}_{P,k-1,n} = [\mathbf{s}_{k-P}[n], \mathbf{s}_{k-P+1}[n], \ldots, \mathbf{s}_{k-1}[n]]^{\mathrm{T}}. \tag{3.2}$$

The AR coefficient vector $\mathbf{a}_{k,n} \in \mathbb{R}^{P\times 1}$ corresponding to the n^{th} source is given by

$$\mathbf{a}_{k,n} = [\mathbf{a}_{k,n}[P], \mathbf{a}_{k,n}[P-1] \ldots, \mathbf{a}_{k,n}[1]]^{\mathrm{T}}. \tag{3.3}$$

A TVAR process equation is frequently used to model the dynamics of speech sources [88][6], and is used here in view of potential applications to the dereverberation of

speech signals. The following two matrix representations for the dynamics of the N sources are used in the development of the algorithm:

$$\begin{aligned} \mathbf{s}_k &= \mathbf{A}_k \mathbf{s}_{P,k-1} + \mathbf{v}_{k-1} && (3.4) \\ &= \mathbf{S}_{k-1} \mathbf{a}_k + \mathbf{v}_{k-1}. && (3.5) \end{aligned}$$

The quantities in the above equation are defined in the following paragraphs. The N sources are collected in the vector $\mathbf{s}_k \in \mathbb{R}^{N\times 1}$, and similarly the noise variables are collected in $\mathbf{v}_{k-1} \in \mathbb{R}^{N\times 1}$:

$$\begin{aligned} \mathbf{s}_k &= [\mathbf{s}_k[1], \mathbf{s}_k[2], \ldots, \mathbf{s}_k[N]]^{\mathrm{T}}, && (3.6) \\ \mathbf{v}_{k-1} &= [\mathbf{v}_{k-1}[1], \mathbf{v}_{k-1}[2], \ldots, \mathbf{v}_{k-1}[N]]^{\mathrm{T}}. && (3.7) \end{aligned}$$

The vector $\mathbf{v}_{k-1}$ is assumed Gaussian distributed with mean zero and diagonal covariance $\mathbf{\Sigma}_v$ given by

$$\mathbf{\Sigma}_v = \mathrm{diag}([\sigma_{v,1}^2, \sigma_{v,2}^2, \ldots, \sigma_{v,N}^2]). \tag{3.8}$$

The operator diag($\mathbf{x}$) denotes the diagonal matrix formed from the elements of the vector $\mathbf{x}$. In (3.4), $\mathbf{s}_{P,k-1} \in \mathbb{R}^{NP\times 1}$ is the concatenation of the most recent P source vectors at time $k-1$:

$$\mathbf{s}_{P,k-1} = [\mathbf{s}_{k-P}^{\mathrm{T}}, \mathbf{s}_{k-P+1}^{\mathrm{T}}, \ldots, \mathbf{s}_{k-1}^{\mathrm{T}}]^{\mathrm{T}}, \tag{3.9}$$

and the AR coefficient matrix $\mathbf{A}_k \in \mathbb{R}^{N\times NP}$ is given by

$$\mathbf{A}_k = [\mathbf{A}_{k,P-1}, \mathbf{A}_{k,P-2}, \ldots, \mathbf{A}_{k,0}], \tag{3.10}$$

where $\mathbf{A}_{k,p} \in \mathbb{R}^{N\times P}$, $p = 0, 1, \ldots, P-1$ is defined as

$$\mathbf{A}_{k,p} = \mathrm{diag}([\mathbf{a}_{k,1}[p], \mathbf{a}_{k,2}[p], \ldots, \mathbf{a}_{k,N}[p]]). \tag{3.11}$$

In (3.5), the matrix $\mathbf{S}_{k-1} \in \mathbb{R}^{N\times NP}$ is block diagonal as given by

$$\mathbf{S}_{k-1} = \begin{bmatrix} \mathbf{s}_{P,k-1,1}^{\mathrm{T}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{s}_{P,k-1,2}^{\mathrm{T}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{s}_{P,k-1,N}^{\mathrm{T}} \end{bmatrix}, \tag{3.12}$$

and the vector $\mathbf{a}_k \in \mathbb{R}^{NP\times 1}$ is the concatenation of the N vectors $\mathbf{a}_{k,n}$ of length P corresponding to each source:

$$\mathbf{a}_k = [\mathbf{a}_{k,1}^{\mathrm{T}}, \mathbf{a}_{k,2}^{\mathrm{T}}, \ldots, \mathbf{a}_{k,N}^{\mathrm{T}}]^{\mathrm{T}}. \tag{3.13}$$

The time-varying $\mathbf{a}_k$ is itself assumed to evolve according to a first-order AR model as follows:

$$\mathbf{a}_k = a_a \mathbf{a}_{k-1} + \mathbf{v}_{a,k-1}, \tag{3.14}$$

with $0 < a_a < 1$ assumed known and the process noise vector $\mathbf{v}_{a,k-1}$ Gaussian distributed with mean zero and known covariance $\mathbf{\Sigma}_a$. In practice, the AR coefficients are constrained to be stable with all poles inside the unit circle.

We now consider the measurement equation for the j^{th} sensor which is assumed to evolve according to the convolution of the sources with time-varying FIR channels in the presence of additive noise as follows:

$$\mathbf{y}_k[j] = \mathbf{h}_{k,j}^{\mathrm{T}} \mathbf{s}_{L,k} + \mathbf{w}_k[j]. \tag{3.15}$$

The measurement noise $\mathbf{w}_k[j] \in \mathbb{R}$ is assumed to be Gaussian distributed with mean zero and unknown variance $\sigma_{w,j}^2$. The noise variances are assumed to be independent between sensors. The source vector $\mathbf{s}_{L,k} \in \mathbb{R}^{NL\times 1}$ is the concatenation of the most recent L source vectors at time k:

$$\mathbf{s}_{L,k} = [\mathbf{s}_{k-L+1}^{\mathrm{T}}, \mathbf{s}_{k-L+2}^{\mathrm{T}}, \ldots, \mathbf{s}_k^{\mathrm{T}}]^{\mathrm{T}}. \tag{3.16}$$

The channel vector $\mathbf{h}_{k,j} \in \mathbb{R}^{NL\times 1}$ corresponding to the j^{th} sensor is formed from the FIR filters $\mathbf{h}_{k,j,n}$ of length L from the n^{th} source to the j^{th} sensor:

$$\begin{aligned} \mathbf{h}_{k,j} &= [\mathbf{h}_{k,j,L-1}^{\mathrm{T}}, \mathbf{h}_{k,j,L-2}^{\mathrm{T}}, \ldots, \mathbf{h}_{k,j,0}^{\mathrm{T}}]^{\mathrm{T}}, \\ \mathbf{h}_{k,j,\ell} &= [\mathbf{h}_{k,j,1}[\ell], \mathbf{h}_{k,j,2}[\ell], \ldots, \mathbf{h}_{k,j,N}[\ell]]^{\mathrm{T}} \end{aligned} \tag{3.17}$$

where $\ell = 0, 1, \ldots, L-1$ is the lag from the current time. The two matrix representations used for the measurements at the J sensors are:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s}_{L,k} + \mathbf{w}_k \tag{3.18}$$

$$= \mathbf{T}_k \mathbf{h}_k + \mathbf{w}_k. \tag{3.19}$$

The measurement vector $\mathbf{y}_k \in \mathbb{R}^{J\times 1}$ and the measurement noise vector $\mathbf{w}_k \in \mathbb{R}^{J\times 1}$ are defined as

$$\mathbf{y}_k = [\mathbf{y}_k[1], \mathbf{y}_k[2], \ldots, \mathbf{y}_k[J]]^{\mathrm{T}}, \tag{3.20}$$

$$\mathbf{w}_{k-1} = [\mathbf{w}_{k-1}[1], \mathbf{w}_{k-1}[2], \ldots, \mathbf{w}_{k-1}[J]]^{\mathrm{T}}. \tag{3.21}$$

The vector $\mathbf{w}_k$ is assumed Gaussian distributed with mean zero and diagonal covariance $\mathbf{\Sigma}_w$ given by

$$\mathbf{\Sigma}_w = \mathrm{diag}([\sigma^2_{w,1}, \sigma^2_{w,2}, \ldots, \sigma^2_{w,J}]). \tag{3.22}$$

In (3.18), the channel matrix $\mathbf{H}_k \in \mathbb{R}^{J\times NL}$ is given by

$$\mathbf{H}_k = [\mathbf{H}_{k,L-1}, \mathbf{H}_{k,L-2}, \ldots, \mathbf{H}_{k,0}], \tag{3.23}$$

where $\mathbf{H}_{k,\ell} \in \mathbb{R}^{J\times N}$, $\ell = 0, 1, \ldots, L-1$ is defined as

$$\mathbf{H}_{k,\ell} = \begin{bmatrix} \mathbf{h}^{\mathrm{T}}_{k,1,\ell} \\ \mathbf{h}^{\mathrm{T}}_{k,2,\ell} \\ \vdots \\ \mathbf{h}^{\mathrm{T}}_{k,J,\ell} \end{bmatrix}. \tag{3.24}$$

In (3.19), the matrix $\mathbf{T}_k \in \mathbb{R}^{J\times JNL}$ is block diagonal with structure

$$\begin{aligned} \mathbf{T}_k &= \mathbf{I}_J \otimes \mathbf{s}^{\mathrm{T}}_{L,k} \\ &= \begin{bmatrix} \mathbf{s}^{\mathrm{T}}_{L,k} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{s}^{\mathrm{T}}_{L,k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{s}^{\mathrm{T}}_{L,k} \end{bmatrix}, \end{aligned} \tag{3.25}$$

and the channel vector $\mathbf{h}_k \in \mathbb{R}^{JNL\times 1}$ is the concatenation of the J $\mathbf{h}_{k,j}$

$$\mathbf{h}_k = [\mathbf{h}^{\mathrm{T}}_{k,1}, \mathbf{h}^{\mathrm{T}}_{k,2}, \ldots, \mathbf{h}^{\mathrm{T}}_{k,J}]^{\mathrm{T}}. \tag{3.26}$$

The time-varying $\mathbf{h}_k$ is assumed to evolve according to a first-order AR model as follows:

$$\mathbf{h}_k = a_h \mathbf{h}_{k-1} + \mathbf{v}_{h,k-1}, \tag{3.27}$$

with $0 < a_h < 1$ assumed known and the process noise vector $\mathbf{v}_{h,k-1}$ Gaussian distributed with mean zero and known covariance $\boldsymbol{\Sigma}_h$.

The scope of the thesis has been limited to consider only additive Gaussian source and measurement noise. This is to allow for a clearer presentation of the particle filter approach for the nonlinear aspects of the state space model for blind system identification. It is also noted that in many practical applications the assumption of Gaussian noise is reasonable. A particle filtering approach to the problem does allow the current algorithm to be extended to the case of non-Gaussian noise [72].

The state space model shown in Figure 3.1 can also be interpreted as a time-varying autoregressive moving average (ARMA) system operating on the source noise inputs $\mathbf{v}_k[n]$, $n = 1, 2, \ldots, N$ to produce the outputs $\mathbf{y}_k[j]$, $j = 1, 2, \ldots, J$.

3.1.2 Joint Posterior Distribution

The unknown states are collected in the composite state $\boldsymbol{\theta}$ from time 1 to time K:

$$\boldsymbol{\theta}_{1:K} = \{\mathbf{s}_{1:K}, \mathbf{h}_{1:K}, \mathbf{a}_{1:K}, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_w\}. \tag{3.28}$$

Using Bayes' theorem and the dependencies in the state space model the posterior distribution of the unknown parameters given the set of measurements is

$$\begin{aligned} p(\boldsymbol{\theta}_{1:K}|\mathbf{y}_{1:K}) \propto\ & p(\mathbf{y}_{1:K}|\mathbf{s}_{1:K}, \mathbf{h}_{1:K}, \boldsymbol{\Sigma}_w) \\ & \times p(\mathbf{s}_{1:K}|\mathbf{a}_{1:K}, \boldsymbol{\Sigma}_v) p(\mathbf{h}_{1:K}) p(\mathbf{a}_{1:K}) p(\boldsymbol{\Sigma}_v) p(\boldsymbol{\Sigma}_w). \end{aligned} \tag{3.29}$$

The likelihood term follows from the measurement model in (3.19):

$$p(\mathbf{y}_{1:K}|\mathbf{s}_{1:K}, \mathbf{h}_{1:K}, \boldsymbol{\Sigma}_w) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{T}_k \mathbf{h}_k, \boldsymbol{\Sigma}_w). \tag{3.30}$$

The remaining terms in the posterior distribution are the assumed prior distributions for the unknown system variables. From the source model in (3.5), the prior for the source is

$$p(\mathbf{s}_{1:K}|\mathbf{a}_{1:K}, \boldsymbol{\Sigma}_v) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{S}_{k-1} \mathbf{a}_k, \boldsymbol{\Sigma}_v). \tag{3.31}$$

The priors for the AR coefficients and channel follow from the first-order AR models in (3.14) and (3.27):

$$p(\mathbf{a}_{1:K}) = \prod_{k=1}^{K} \mathcal{N}(a_a \mathbf{a}_{k-1}, \boldsymbol{\Sigma}_a), \tag{3.32}$$

$$p(\mathbf{h}_{1:K}) = \prod_{k=1}^{K} \mathcal{N}(a_h \mathbf{h}_{k-1}, \boldsymbol{\Sigma}_h). \tag{3.33}$$

The AR model order P and convolution length L are assumed known. The case of source- and channel-specific model orders P_n and L_{nj}, respectively, is not considered here to simplify the presentation. In the case of unknown model orders, reversible jump Markov Chain Monte Carlo (RJMCMC) methods could be investigated for model order detection [38][5][87].

Inverse Gamma distributions, the conjugate prior for the normal likelihood distribution [10], are assumed for the priors of the source and measurement noise variances:

$$p(\boldsymbol{\Sigma}_v) = \prod_{n=1}^{N} \mathcal{IG}\left(\frac{\upsilon_{v,n}}{2}, \frac{\gamma_{v,n}}{2}\right), \tag{3.34}$$

$$p(\boldsymbol{\Sigma}_w) = \prod_{j=1}^{J} \mathcal{IG}\left(\frac{\upsilon_{w,j}}{2}, \frac{\gamma_{w,j}}{2}\right), \tag{3.35}$$

assuming the noise variances $\sigma^2_{v,n}, n = 1, 2, \ldots, N$ and $\sigma^2_{w,j}, j = 1, 2, \ldots, J$ are independent. The Inverse Gamma distribution for a random variable σ^2 is of the form

$$\mathcal{IG}(\upsilon, \gamma) = (\sigma^2)^{-\upsilon-1} \exp\left[-\frac{\gamma}{\sigma^2}\right]. \tag{3.36}$$

Using the above likelihood and prior distributions, the posterior distribution is then

$$\begin{aligned} p(\boldsymbol{\theta}_{1:K}|\mathbf{y}_{1:K}) \propto & \prod_{k=1}^{K} \frac{1}{(2\pi)^{\frac{J}{2}}|\boldsymbol{\Sigma}_w|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)\right] \\ & \times \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{\Sigma}_v|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{s}_k - \mathbf{S}_{k-1}\mathbf{a}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_v^{-1}(\mathbf{s}_k - \mathbf{S}_{k-1}\mathbf{a}_k)\right] \\ & \times \frac{1}{(2\pi)^{\frac{NJL}{2}}|\boldsymbol{\Sigma}_h|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{h}_k - a_h\mathbf{h}_{k-1})^{\mathrm{T}}\boldsymbol{\Sigma}_h^{-1}(\mathbf{h}_k - a_h\mathbf{h}_{k-1})\right] \\ & \times \frac{1}{(2\pi)^{\frac{NP}{2}}|\boldsymbol{\Sigma}_a|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{a}_k - a_a\mathbf{a}_{k-1})^{\mathrm{T}}\boldsymbol{\Sigma}_a^{-1}(\mathbf{a}_k - a_a\mathbf{a}_{k-1})\right] \\ & \times \prod_{n=1}^{N}(\sigma_{v,n}^2)^{-\frac{\nu_{v,n}}{2}-1} \exp\left[-\frac{\gamma_{v,n}}{2\sigma_{v,n}^2}\right] \\ & \times \prod_{j=1}^{J}(\sigma_{w,j}^2)^{-\frac{\nu_{w,j}}{2}-1} \exp\left[-\frac{\gamma_{w,j}}{2\sigma_{w,j}^2}\right]. \end{aligned} \tag{3.37}$$

3.2 Marginalized Particle Filtering

The estimation problem for the given state space model is nonlinear since *both* the source vector $\mathbf{s}_{L,k}$ and channel $\mathbf{h}_k$ are unknown. As a result of the discussion in Section 2.2, the optimal Bayesian recursion for the joint posterior distribution in (3.37) cannot be evaluated analytically. An efficient marginalized particle filtering algorithm [79], also commonly known as Rao-Blackwellised particle filters (RBPF) [22][27] or mixture Kalman filters [17], is developed by exploiting the conditionally linear-Gaussian substructure of the state space model.

3.2.1 Introduction to the Rao-Blackwellisation Procedure

The fundamentals of the marginalized particle filter technique are introduced first with an illustrative example. The general strategy is then applied to the more complicated state space model from Section 3.1.1. The example we now consider is a

nonlinear state space model whose transition equations have the specific structure

$$\mathbf{x}_k^1 = \mathbf{f}_k(\mathbf{x}_{k-1}^1, \mathbf{v}_{k-1}^1), \tag{3.38}$$

$$\mathbf{x}_k^2 = \mathbf{F}_k \mathbf{x}_{k-1}^2 + \mathbf{v}_{k-1}^2, \tag{3.39}$$

and the measurement equation is given by

$$\mathbf{y}_k = \mathbf{G}(\mathbf{x}_k^1)\mathbf{x}_k^2 + \mathbf{w}_k. \tag{3.40}$$

The noise vectors $\mathbf{v}_{k-1}^1$,$\mathbf{v}_{k-1}^2$, and $\mathbf{w}_k$ are all assumed to be Gaussian distributed. The state variable $\mathbf{x}_k^1$ enters the dynamics through the *nonlinear* process function $\mathbf{f}_k$, while the other unknown state variable $\mathbf{x}_k^2$ evolves from the *linear* process function given by the known matrix $\mathbf{F}_k$. The measurement equation is nonlinear in the two variables, where $\mathbf{G}(\mathbf{x}_k^1)$ is a matrix function of the vector $\mathbf{x}_k^1$. For this nonlinear state space model, it is possible to implement a Bayesian solution to the sequential state estimation problem using the particle filter for the composite state vector $\mathbf{x}_k = [(\mathbf{x}_k^1)^{\mathrm{T}}, (\mathbf{x}_k^2)^{\mathrm{T}}]^{\mathrm{T}}$. However, the particle filtering algorithm performance can be improved by considering the specific structure of (3.38)-(3.40).

We now examine how the Rao-Blackwellisation procedure [15] uses the inherent linear-Gaussian substructure in the model to improve the performance of the estimation algorithm. It is seen from the measurement equation that conditional on $\mathbf{x}_k^1$ (implying the quantity $\mathbf{G}(\mathbf{x}_k^1)$ is a constant matrix) the measurement equation is linear-Gaussian in the variable $\mathbf{x}_k^2$. We observe that since the process equation for $\mathbf{x}_k^2$ is also linear-Gaussian, the problem of estimating $\mathbf{x}_k^2$ conditional on $\mathbf{x}_k^1$ can be solved using the optimal Kalman filter. This realization motivates the following factorization of the joint posterior distribution using Bayes' rule:

$$p(\mathbf{x}_{1:k}^1, \mathbf{x}_{1:k}^2 | \mathbf{y}_{1:k}) = p(\mathbf{x}_{1:k}^2 | \mathbf{x}_{1:k}^1, \mathbf{y}_{1:k}) p(\mathbf{x}_{1:k}^1 | \mathbf{y}_{1:k}). \tag{3.41}$$

This equation is pivotal in the Rao-Blackwellisation strategy. Each term in (3.41) is developed is the subsequent sub-sections:

The *marginalized* posterior distribution $p(\mathbf{x}^1_{1:k}|\mathbf{y}_{1:k})$**:** Since this distribution must be evaluated recursively, a particle filtering approach is used to estimate $\mathbf{x}^1_k$ [22]. From (2.35), the marginalized posterior distribution satisfies the recursion

$$p(\mathbf{x}^1_{1:k}|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}^1_{1:k}, \mathbf{y}_{1:k-1})p(\mathbf{x}^1_k|\mathbf{x}^1_{1:k-1}, \mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} p(\mathbf{x}^1_{1:k-1}|\mathbf{y}_{1:k-1}). \tag{3.42}$$

This recursion is now in terms of a *marginalized* likelihood $p(\mathbf{y}_k|\mathbf{x}^1_{1:k}, \mathbf{y}_{1:k-1})$ and prior $p(\mathbf{x}^1_k|\mathbf{x}^1_{1:k-1}, \mathbf{y}_{1:k-1})$. The marginalized likelihood is found using the Rao-Blackwellisation strategy to marginalize the state $\mathbf{x}^2_k$ out of the conditional likelihood $p(\mathbf{y}_k|\mathbf{x}^1_{1:k}, \mathbf{x}^2_{1:k})$:

$$p(\mathbf{y}_k|\mathbf{x}^1_{1:k}, \mathbf{y}_{1:k-1}) = \int p(\mathbf{y}_k|\mathbf{x}^1_{1:k}, \mathbf{x}^2_k) p(\mathbf{x}^2_k|\mathbf{x}^1_{1:k}, \mathbf{y}_{1:k-1}) d\mathbf{x}^2_k \tag{3.43}$$

The conditional likelihood (which is independent of the measurement history $\mathbf{y}_{1:k-1}$) follows directly from the measurement equation (3.40):

$$p(\mathbf{y}_k|\mathbf{x}^1_{1:k}, \mathbf{x}^2_{1:k}) = \mathcal{N}(\mathbf{G}(\mathbf{x}^1_k)\mathbf{x}^2_k, \mathbf{\Sigma}_w), \tag{3.44}$$

where it assumed that the Gaussian noise $\mathbf{w}_k$ is zero-mean with covariance $\mathbf{\Sigma}_w$.
To implement the marginalization procedure, the distribution $p(\mathbf{x}^2_k|\mathbf{x}^1_{1:k}, \mathbf{y}_{1:k})$ in (3.43) is also required. From the discussion above on the linear-Gaussian substructure of $\mathbf{x}^2_k$ conditioned on $\mathbf{x}^1_k$, this distribution can be recursively computed using the Kalman filter. Since the particle filter for the state $\mathbf{x}^1_k$ discussed above produces the set of particles $\{\mathbf{x}^{1(i)}_{1:k}, i = 1, 2, \ldots, N_p\}$, a set of Kalman filters conditioned on each of the particle values is required as shown:

$$p(\mathbf{x}^2_k|\mathbf{x}^{1(i)}_{1:k}, \mathbf{y}_{1:k}) = \mathcal{N}(\hat{\mathbf{x}}^{2(i)}_k, \mathbf{\Phi}^{2(i)}_k), \tag{3.45}$$

where $\hat{\mathbf{x}}^{2(i)}_k$ and $\mathbf{\Phi}^{2(i)}_k$ are the mean and covariance of the i^{th} Kalman filter which follow the recursions described in Section 2.3. The integration in (3.43) can then be computed using the distributions in (3.44) and (3.45). The details of this integration procedure are presented in a later section for the specific state space model structure introduced in Section 3.1.1. This completes the discussion on the evaluation of the term $p(\mathbf{y}_k|\mathbf{x}^1_{1:k}, \mathbf{y}_{1:k-1})$ in (3.42).

We now look at the marginalized prior $p(\mathbf{x}_k^1|\mathbf{x}_{1:k-1}^1, \mathbf{y}_{1:k-1})$ in (3.42). Evaluation of this distribution follows directly from the state equation in (3.38) and the noise statistics of $\mathbf{v}_k^1$, which are both not dependent on $\mathbf{x}_k^2$. From Section 2.4.3, it is seen that recursion of the posterior distribution in (3.42) is the basis for the particle filter recursion. The denominator in (3.42) is only a normalization term and need not be evaluated. Neither does the term $p(\mathbf{x}_{1:k-1}^1|\mathbf{y}_{1:k})$, since this is available from the previous time instant. Thus, from the development of this section, all terms in (3.42) are available, and the particle filter recursion for evaluating the nonlinear state $\mathbf{x}_{1:k}^1$ can proceed independent of $\mathbf{x}_{1:k}^2$.

The linear-Gaussian distribution $p(\mathbf{x}_{1:k}^2|\mathbf{x}_{1:k}^1, \mathbf{y}_{1:k})$**:** This distribution in (3.41) has already been obtained from the Kalman filter as shown in (3.45). □

The benefit of using the marginalized particle filter is that the dimension of the state estimated using the particle filter is reduced. It is known that the performance of the particle filter degrades as the dimension of the state variable increases [26]. The Rao-Blackwellisation strategy simplifies the particle filtering problem for the composite state vector $\mathbf{x}_k$ to a problem in terms of only $\mathbf{x}_k^1$. This method is shown in [29] to reduce the variance in the importance weights of the particle filter, which translates into improved estimation performance. The intuition behind this result is that the particle filter, a Monte Carlo approximation to the optimal Bayesian filter, is now only used to estimate the truly nonlinear/non-Gaussian states, while the remaining conditional linear-Gaussian states can be estimated analytically using the optimal Kalman filter [79].

3.2.2 Application of Rao-Blackwellisation to the Blind System Identification Problem

We now apply the Rao-Blackwellisation strategy to the Bayesian formulation of the proposed state space model in Section 3.1.1. It can be seen from the state space model that conditional on the sources $\mathbf{s}_{1:k}$ (which form $\mathbf{T}_k$) and the measurement

noise covariance $\mathbf{\Sigma}_w$, equations (3.27) and (3.19) for the FIR coefficients $\mathbf{h}_k$ form a linear-Gaussian subsystem:

$$\mathbf{h}_k = a_h \mathbf{h}_{k-1} + \mathbf{v}_{h,k-1} \tag{3.46}$$

$$\mathbf{y}_k = \mathbf{T}_k \mathbf{h}_k + \mathbf{w}_k \tag{3.47}$$

Similarly, the pair of equations (3.14) and (3.5) for the AR coefficients $\mathbf{a}_k$ conditioned on the sources and the source noise covariance $\mathbf{\Sigma}_v$ also form a linear-Gaussian subsystem:

$$\mathbf{a}_k = a_a \mathbf{a}_{k-1} + \mathbf{v}_{a,k-1} \tag{3.48}$$

$$\mathbf{s}_k = \mathbf{S}_{k-1} \mathbf{a}_k + \mathbf{v}_{k-1} \tag{3.49}$$

The joint posterior distribution in (3.37) is then factorized using Bayes' rule to exploit this conditionally linear-Gaussian substructure:

$$p(\mathbf{s}_{1:k}, \mathbf{a}_{1:k}, \mathbf{h}_{1:k} | \mathbf{y}_{1:k}) = p(\mathbf{a}_{1:k} | \mathbf{s}_{1:k}, \mathbf{y}_{1:k}) p(\mathbf{h}_{1:k} | \mathbf{s}_{1:k}, \mathbf{y}_{1:k}) p(\mathbf{s}_{1:k} | \mathbf{y}_{1:k}). \tag{3.50}$$

MAP estimates of the noise covariances $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_w$ are developed separately, as shown in Section 3.5, and the dependence on these terms is not shown explicitly in the subsequent derivations of the particle and Kalman filter algorithms. In the proposed procedure, the marginalized posterior distribution $p(\mathbf{s}_{1:k} | \mathbf{y}_{1:k})$ is obtained using the Rao-Blackwellisation strategy for marginalizing out the conditionally linear-Gaussian AR and FIR coefficients. The sources $\mathbf{s}_{1:k}$ are then estimated numerically using the particle filter for the resulting nonlinear problem. The filtered distributions $p(\mathbf{a}_k | \mathbf{s}_{1:k}, \mathbf{y}_{1:k})$ and $p(\mathbf{h}_k | \mathbf{s}_{1:k}, \mathbf{y}_{1:k})$ are computed recursively for the conditionally linear-Gaussian problems using the Kalman filter. It is noted that the two linear-Gaussian subsystems for $\mathbf{h}_k$ and $\mathbf{a}_k$ are decoupled from each other after conditioning on the sources. This allows the set of Kalman filters for the AR coefficients to operate in parallel with the set of Kalman filters for the FIR coefficients.

The plan for the remaining portion of this chapter is as follows. Section 3.3 details the particle filter algorithm for nonlinear estimation of the sources, Section

3.4 describes the Kalman filtering procedure for estimating the conditionally linear-Gaussian FIR and AR coefficients, and the analytical expressions for the MAP noise variance estimates are derived in Section 3.5. A pictorial overview of the marginalized particle filtering algorithm structure is presented in Figure 3.2.

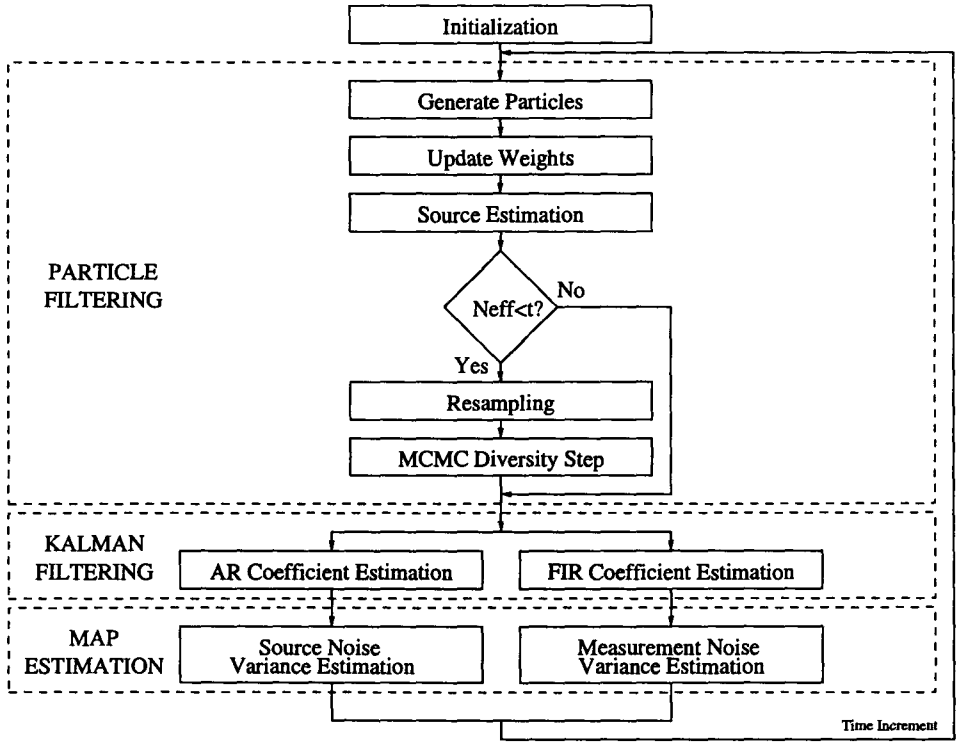


Figure 3.2: Marginalized Particle Filtering Algorithm Structure

3.3 Particle Filtering for Source Estimation

From the discussion of the Rao-Blackwellised particle filter in the previous section, sequential Monte Carlo methods are used to recursively estimate the marginalized posterior distribution $p(\mathbf{s}_{1:k}|\mathbf{y}_{1:k})$ of the sources in (3.50). In Section 2.4, the particle filter was introduced under the assumption that the state is first-order Markov. It is seen from the state equation in (3.4) that the current source vector $\mathbf{s}_k$ for the

AR model is dependent on the past P source vectors contained in the vector $\mathbf{s}_{P,k-1}$. Therefore, the current formulation of the state space model is not suitable for implementation of the SIS particle filter since the state is not first-order Markov. The state space model for the sources is now reformulated in terms of a new state variable that follows a first-order Markov process equation that considers the past P values of the sources. This does not change any of the underlying statistical assumptions on the dynamical model introduced in Section 3.1.1. The state variable $\mathbf{s}_{M,k} \in \mathbb{R}^{MN\times 1}$ is introduced, where $M = max(P, L)$ is the maximum of the orders of the AR and FIR models:

$$\begin{aligned}\mathbf{s}_{M,k} &= [\mathbf{s}_{k-M+1}^{\mathrm{T}}, \mathbf{s}_{k-M+2}^{\mathrm{T}}, \ldots, \mathbf{s}_{k}^{\mathrm{T}}]^{\mathrm{T}} \\ &= [\mathbf{s}_{M-1,k-1}^{\mathrm{T}}, \mathbf{s}_{k}^{\mathrm{T}}]^{\mathrm{T}}.\end{aligned} \tag{3.51}$$

The state equation for $\mathbf{s}_{M,k}$ is derived by considering the autoregressive property of the sources. At time k, the first $M-1$ blocks $\mathbf{s}_{k-M+i}$ of $\mathbf{s}_{M,k}$ are a shifted version of the blocks at time $k-1$ as reflected in the partitioning in (3.51). In combination with (3.4), for the case of $M = P$ as an example, we may write the source state update as

$$\begin{bmatrix} \mathbf{s}_{k-M+1} \\ \mathbf{s}_{k-M+2} \\ \vdots \\ \mathbf{s}_{k-1} \\ \mathbf{s}_{k} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_N & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_N & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_N \\ \mathbf{A}_{k,P-1} & \mathbf{A}_{k,P-2} & \ldots & \ldots & \mathbf{A}_{k,1} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{k-M} \\ \mathbf{s}_{k-M+1} \\ \vdots \\ \mathbf{s}_{k-2} \\ \mathbf{s}_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mathbf{v}_{k-1} \end{bmatrix}.$$

In general, this can be written compactly as

$$\mathbf{s}_{M,k} = \tilde{\mathbf{A}}_k \mathbf{s}_{M,k-1} + \tilde{\mathbf{v}}_{k-1}. \tag{3.52}$$

To describe the variables in this process equation, we first define the quantity $\mathbf{0}_{a,b}$ as a matrix of zeros of dimension $a \times b$ if $a, b > 0$ and empty otherwise. The notation $\mathbf{I}_a$ denotes the $a \times a$ identity matrix. The variable definitions then follow as

$$\tilde{\mathbf{A}}_k = \begin{bmatrix} [\mathbf{0}_{(M-1)N,N}, \mathbf{I}_{(M-1)N}] \\ [\mathbf{0}_{N,(L-P)N}, \mathbf{A}_k] \end{bmatrix}, \tag{3.53}$$

$$\tilde{\mathbf{v}}_{k-1} = [\mathbf{0}_{1,(M-1)N}, \mathbf{v}_{k-1}^{\mathrm{T}}]^{\mathrm{T}}. \tag{3.54}$$

The measurement equation in (3.18) is rewritten in terms of $\mathbf{s}_{M,k}$ to produce

$$\mathbf{y}_k = \tilde{\mathbf{H}}_k \mathbf{s}_{M,k} + \mathbf{w}_k, \tag{3.55}$$

where

$$\tilde{\mathbf{H}}_k = [\mathbf{0}_{J,(P-L)N}, \mathbf{H}_k]. \tag{3.56}$$

The second form of the measurement equation for $\mathbf{s}_{M,k}$ remains unchanged from (3.19). It is noted that in (3.55), the current measurement is only dependent on the current state and noise vector which is also a required assumption from the particle filter development in Section 2.4. This was previously not satisfied using the formulation in (3.18).

The state space representation in (3.52) and (3.55) is now suitable for developing a SIS particle filter implementation. From (2.35), the recursion for the marginalized posterior distribution in terms of $\mathbf{s}_{M,k}$ is

$$p(\mathbf{s}_{M,1:k}|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1}) p(\mathbf{s}_{M,k}|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} p(\mathbf{s}_{M,1:k-1}|\mathbf{y}_{1:k-1}). \tag{3.57}$$

The *marginalized prior* distribution $p(\mathbf{s}_{M,k}|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1})$ and *marginalized likelihood* distribution $p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1})$ are now developed. It is first observed from the partition of $\mathbf{s}_{M,k}$ in (3.51) that the marginalized prior can be factorized as

$$p(\mathbf{s}_{M,k}|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1}) = p(\mathbf{s}_k|\mathbf{s}_{M,k-1}, \mathbf{y}_{1:k-1}) \prod_{i=1}^{(M-1)N} \delta(\mathbf{s}_{M,k}[i] - \mathbf{s}_{M,k-1}[i+N]). \tag{3.58}$$

This equation follows from the fact that only block $\mathbf{s}_k$ of $\mathbf{s}_{M,k}$ is a random variable, as indicated by (3.5). The remaining blocks of $\mathbf{s}_{M,k}$ are deterministic shifts of the blocks from time $k-1$. Thus, instead of the distribution $p(\mathbf{s}_{M,k}|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1})$ in (3.57), it suffices to consider only the distribution $p(\mathbf{s}_k|\mathbf{s}_{M,k-1}, \mathbf{y}_{1:k-1})$. The marginalized prior $p(\mathbf{s}_k|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1})$ is now determined using the Rao-Blackwellisation strategy by marginalizing the conditional prior $p(\mathbf{s}_k|\mathbf{a}_k, \mathbf{s}_{P,k-1})$ over the AR coefficients $\mathbf{a}_k$:

$$p(\mathbf{s}_k|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1}) = \int p(\mathbf{s}_k|\mathbf{a}_k, \mathbf{s}_{P,k-1}) p(\mathbf{a}_k|\mathbf{s}_{1:k-1}, \mathbf{y}_{1:k-1}) d\mathbf{a}_k. \tag{3.59}$$

Similarly, the marginalized likelihood $p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1})$ is found by marginalizing the conditional likelihood $p(\mathbf{y}_k|\mathbf{h}_k, \mathbf{s}_{L,k})$ over the FIR coefficients $\mathbf{h}_k$:

$$p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1}) = \int p(\mathbf{y}_k|\mathbf{h}_k, \mathbf{s}_{L,k}) p(\mathbf{h}_k|\mathbf{s}_{1:k}, \mathbf{y}_{1:k-1}) d\mathbf{h}_k. \tag{3.60}$$

The predicted distributions for the AR and FIR coefficients are available from the Kalman filter as described in Section 3.4:

$$p(\mathbf{a}_k|\mathbf{s}_{1:k-1}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{a}}_{k|k-1}, \mathbf{\Phi}_{a,k|k-1}), \tag{3.61}$$

$$p(\mathbf{h}_k|\mathbf{s}_{1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{h}}_{k|k-1}, \mathbf{\Phi}_{h,k|k-1}). \tag{3.62}$$

Recursions for the quantities $\hat{\mathbf{a}}_k$,$\mathbf{\Phi}_{a,k|k-1}$,$\hat{\mathbf{h}}_k$ and $\mathbf{\Phi}_{h,k|k-1}$ are given in Section 3.4.

The prior for $\mathbf{s}_k$ conditional on the AR coefficients $\mathbf{a}_k$ in (3.59) is obtained from (3.5):

$$\begin{aligned} p(\mathbf{s}_k|\mathbf{a}_k, \mathbf{s}_{M,k-1}) &= p(\mathbf{s}_k|\mathbf{a}_k, \mathbf{s}_{P,k-1}) \\ &= \mathcal{N}(\mathbf{S}_{k-1}\mathbf{a}_k, \mathbf{\Sigma}_v), \end{aligned} \tag{3.63}$$

using the fact that $\mathbf{s}_k$ is dependent on only the past P source vectors at time $k-1$.

The likelihood conditional on the FIR coefficients $\mathbf{h}_k$ in (3.60) follows from (3.19):

$$\begin{aligned} p(\mathbf{y}_k|\mathbf{h}_k, \mathbf{s}_{M,k}) &= p(\mathbf{y}_k|\mathbf{h}_k, \mathbf{s}_{L,k}) \\ &= \mathcal{N}(\mathbf{T}_k\mathbf{h}_k, \mathbf{\Sigma}_w), \end{aligned} \tag{3.64}$$

which uses the fact that $\mathbf{y}_k$ is dependent on only the past L source vectors at time k.

We now substitute (3.61),(3.62) and (3.63),(3.64) into (3.59) and (3.60) respectively. The integrations of (3.59) and (3.60) may then performed. The results provided in Appendix A are:

$$p(\mathbf{s}_k|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{S}_{k-1}\hat{\mathbf{a}}_{k|k-1}, \mathbf{R}_k), \tag{3.65}$$

$$p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{T}_k\hat{\mathbf{h}}_{k|k-1}, \mathbf{Q}_k), \tag{3.66}$$

where the following two definitions have been introduced

$$\mathbf{R}_k = \mathbf{S}_{k-1}\mathbf{\Phi}_{a,k|k-1}\mathbf{S}_{k-1}^{\mathrm{T}} + \mathbf{\Sigma}_v, \tag{3.67}$$

$$\mathbf{Q}_k = \mathbf{T}_k\mathbf{\Phi}_{h,k|k-1}\mathbf{T}_k^{\mathrm{T}} + \mathbf{\Sigma}_w. \tag{3.68}$$

Finally, (3.65) and (3.66) are substituted into the recursion for the marginalized posterior distribution in (3.57). This facilitates the importance weight update process which is described later.

Comparing the conditional prior (3.63) and likelihood (3.64) with the corresponding marginalized expressions in (3.65) and (3.66), it can be seen that the Kalman predicted means $\hat{\mathbf{a}}_{k|k-1}$ and $\hat{\mathbf{h}}_{k|k-1}$ are used as estimates of the true state variables $\mathbf{a}_k$ and $\mathbf{h}_k$, while the uncertainty in the Kalman filter estimates is reflected in the particle filter through the inclusion of the covariance terms $\mathbf{\Phi}_{a,k|k-1}$ and $\mathbf{\Phi}_{h,k|k-1}$.

For notational convenience, we now write the recursion for the marginalized posterior distribution from (3.57) in the standard form shown in the last line of (2.35):

$$p(\mathbf{s}_{M,1:k}|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{s}_{M,k})p(\mathbf{s}_{M,k}|\mathbf{s}_{M,k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} p(\mathbf{s}_{M,1:k-1}|\mathbf{y}_{1:k-1}). \tag{3.69}$$

This hides the dependence on the state $\mathbf{s}_{M,1:k-2}$ and measurement $\mathbf{y}_{1:k-1}$ histories that were used in the marginalization procedure. In practice, these histories do not have to be stored since they are captured by the sufficient statistics of the mean and covariance of the Kalman filter recursions. The steps of the SIS particle filtering algorithm for source estimation are now presented based on the state space formulation and marginalized distributions developed in this section.

3.3.1 Generation of the Particles

From Section 2.4.5, the optimal importance function is

$$\begin{aligned} q(\mathbf{s}_{M,k}|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k}) &= p(\mathbf{s}_{M,k}|\mathbf{s}_{M,k-1}, \mathbf{y}_k) \\ &= \frac{p(\mathbf{y}_k|\mathbf{s}_{M,k})p(\mathbf{s}_{M,k}|\mathbf{s}_{M,k-1})}{p(\mathbf{y}_k|\mathbf{s}_{M,k-1})}. \end{aligned} \tag{3.70}$$

The optimal importance function is a function of $\mathbf{s}_{M,k}$ which is of dimension MN. From the form of the marginalized prior in (3.58), the quantities $\mathbf{s}_{k-M+1:k-1}$ in the state vector $\mathbf{s}_{M,k}$ are deterministic functions of the state at the previous time step. Therefore, in the proposed particle filter algorithm, the particle values for $\mathbf{s}_{k-M+1:k-1}$

in $\mathbf{s}_{M,k}$ are reused from previous time steps. This reduces the problem to generating particles from the optimal importance function for the component $\mathbf{s}_k$ of dimension N given by

$$\begin{aligned} p(\mathbf{s}_k|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k}) &= \frac{p(\mathbf{y}_k|\mathbf{s}_{L,k})p(\mathbf{s}_k|\mathbf{s}_{P,k-1})}{p(\mathbf{y}_k|\mathbf{s}_{M,k-1})} \\ &= \frac{p(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}, \mathbf{s}_k)p(\mathbf{s}_k|\mathbf{s}_{P,k-1})}{p(\mathbf{y}_k|\mathbf{s}_{M,k-1})}, \end{aligned} \tag{3.71}$$

where a partition of the vector $\mathbf{s}_{L,k}$ is applied in order to isolate the variable $\mathbf{s}_k$. To determine an expression for the marginalized likelihood in terms of $\mathbf{s}_k$, rather than (3.66) which is in terms of the matrix $\mathbf{T}_k$, the equivalence of the measurement representations (3.18) and (3.19) is used along with the definitions for $\mathbf{s}_{L,k}$ in (3.16) and $\mathbf{H}_k$ in (3.23):

$$\begin{aligned} \mathbf{T}_k\hat{\mathbf{h}}_{k|k-1} &= \hat{\mathbf{H}}_{k|k-1}\mathbf{s}_{L,k} \\ &= \sum_{\ell=0}^{L-1} \hat{\mathbf{H}}_{k|k-1,\ell}\ \mathbf{s}_{k-\ell} \\ &= \hat{\mathbf{H}}_{k|k-1,0}\mathbf{s}_k + \hat{\mathbf{y}}_{k|k-1}, \end{aligned} \tag{3.72}$$

where the matrices $\hat{\mathbf{H}}_{k|k-1,\ell}$ are formed from the Kalman filter estimate $\hat{\mathbf{h}}_{k|k-1}$, and the predicted measurement is defined as

$$\hat{\mathbf{y}}_{k|k-1} = \sum_{\ell=1}^{L-1} \hat{\mathbf{H}}_{k|k-1,\ell}\ \mathbf{s}_{k-\ell}. \tag{3.73}$$

Using this relation results in a marginalized likelihood equivalent to (3.66), but now written explicitly in terms of $\mathbf{s}_k$:

$$p(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}, \mathbf{s}_k) = \mathcal{N}(\hat{\mathbf{H}}_{k|k-1,0}\mathbf{s}_k + \hat{\mathbf{y}}_{k|k-1}, \mathbf{Q}_k). \tag{3.74}$$

Even though the optimal importance function for $\mathbf{s}_k$ in (3.71) is proportional to the product of the Gaussian distributions (3.65),(3.74), it is not Gaussian itself since the covariance term $\mathbf{Q}_k$ has a dependence on $\mathbf{T}_k$, and therefore on the variable of interest $\mathbf{s}_k$.

where

$$\mathbf{W}_k = \mathbf{R}_k \hat{\mathbf{H}}_{k|k-1,0}^{\mathrm{T}} [\hat{\mathbf{H}}_{k|k-1,0} \mathbf{R}_k \hat{\mathbf{H}}_{k|k-1,0}^{\mathrm{T}} + \hat{\mathbf{Q}}_k]^{-1}. \tag{3.85}$$

It is assumed in constructing $\mathbf{s}_{M,k}^i$ that the initial sources $\mathbf{s}_0, \mathbf{s}_{-1}, \mathbf{s}_{-2}, \ldots = \mathbf{0}$. Through comparison with (2.15) and (2.16), the approximation to the optimal importance function can be interpreted as a Kalman filter update on the prediction $\hat{\mathbf{s}}_{k|k-1}$ from the transition prior. The current measurement is incorporated into the generation of particles after multiplication by the Kalman gain factor $\mathbf{W}_k$. The generation of particles is reduced to a linear-Gaussian problem by conditioning on the predicted estimates $\hat{\mathbf{h}}_{k|k-1}$,$\hat{\mathbf{a}}_{k|k-1}$, and approximating the marginalized measurement covariance $\mathbf{Q}_k$ using $\hat{\mathbf{Q}}_k$.

3.3.2 Update of the Importance Weights

With use of the given approximation to the optimal importance function, the recursive update for the importance weights reduces to

$$\begin{aligned} w_k^i &\propto w_{k-1}^i \frac{p(\mathbf{y}_k|\mathbf{s}_{M,k}^i) p(\mathbf{s}_{M,k}^i|\mathbf{s}_{M,k-1}^i)}{q(\mathbf{s}_{M,k}^i|\mathbf{s}_{M,k-1}^i, \mathbf{y}_{1:k})} \\ &\propto w_{k-1} \frac{p(\mathbf{y}_k|\mathbf{s}_{L,k}^i)}{\hat{p}(\mathbf{y}_k|\mathbf{s}_{L,k}^i)} \hat{p}(\mathbf{y}_k|\mathbf{s}_{M,k-1}^i), \end{aligned} \tag{3.86}$$

where $p(\mathbf{y}_k|\mathbf{s}_{L,k}^i)$ is evaluated using (3.74), $\hat{p}(\mathbf{y}_k|\mathbf{s}_{L,k}^i)$ is evaluated using (3.79), and

$$\begin{aligned} \hat{p}(\mathbf{y}_k|\mathbf{s}_{M,k-1}^i) &= \int \hat{p}(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}^i, \mathbf{s}_k) p(\mathbf{s}_k|\mathbf{s}_{P,k-1}^i) d\mathbf{s}_k \\ &= \mathcal{N}(\hat{\mathbf{H}}_{k|k-1,0}^i \mathbf{S}_{k-1}^i \hat{\mathbf{a}}_{k|k-1}^i + \hat{\mathbf{y}}_{k|k-1}^i, \hat{\mathbf{Q}}_k^i + \hat{\mathbf{H}}_{k|k-1,0}^i (\mathbf{R}_k^i) \hat{\mathbf{H}}_{k|k-1,0}^{(i)\mathrm{T}}). \end{aligned} \tag{3.87}$$

The identity (B.24) in Appendix B has been used to rewrite the integrand as the product of the Gaussian importance function for $\mathbf{s}_k$ and a Gaussian distribution for $\mathbf{y}_k$ that is independent of $\mathbf{s}_k$. Integrating over $\mathbf{s}_k$ integrates the importance function for $\mathbf{s}_k$ to one, and results in the remaining Gaussian distribution. After the weight update, the weights are normalized using (2.29).

3.3.3 Source Estimation

Statistical inference on the sources can be determined from the set of particles $\mathbf{s}_{M,k}^i$ and corresponding importance weights w_k^i. Source estimation is performed before the subsequent resampling step since resampling introduces random variation in the current sample [61]. The filtered MMSE estimate of $\mathbf{s}_k$ given $\mathbf{y}_{1:k}$ is computed based on the Monte Carlo approximation in (2.32):

$$\hat{\mathbf{s}}_k^{\text{MMSE}} \approx \sum_{i=1}^{N_p} \mathbf{s}_k^i \cdot w_k^i. \tag{3.88}$$

Since the past $M-1$ source vectors are also being stored in $\mathbf{s}_{M,k}^i$, smoothed estimates of the past states $\mathbf{s}_{k-\ell}$ at a lag $\ell = 1, 2, \ldots, M-1$ from the current time can be computed using a fixed-lag Monte Carlo smoothing algorithm [25]. The smoothed MMSE estimate of the source vector $\mathbf{s}_{k-\ell}$ given the measurements up to the current time k is given by

$$\hat{\mathbf{s}}_{k-\ell|k}^{\text{MMSE}} \approx \sum_{i=1}^{N_p} \mathbf{s}_{k-\ell}^i \cdot w_k^i. \tag{3.89}$$

Smoothing methods can improve performance by incorporating future measurements into the state estimate. For real-time applications in which an estimate $\mathbf{s}_k$ must be produced at time k, then the filtered estimate in (3.88) must be used. The smoothing algorithm can only be considered when real-time estimates are not required. Other smoothing methods, including fixed-interval smoothing, could also be considered [35],[54].

For comparison with the PCRB, the MMSE estimate of the one-step-ahead predicted sources is computed using

$$\hat{\mathbf{s}}_{k|k-1}^{\text{MMSE}} \approx \sum_{i=1}^{N_p} \hat{\mathbf{s}}_{k|k-1}^i \cdot w_k^i, \tag{3.90}$$

where $\mathbf{s}_{k|k-1}$ is defined in (3.75).

3.3.4 Resampling

In order to prevent the problem of importance weight degeneracy as described in Section 2.4.4, a resampling step is used. The approximate effective sample size $\widehat{N_{\text{eff}}}$ in (2.43) is used as a measure of the degeneracy of the sampling scheme. A resampling step occurs whenever $\widehat{N_{\text{eff}}}$ is below a fixed threshold N_{thresh}.

The systematic resampling algorithm [54] is used for the simulations. This approach minimizes the Monte Carlo variation introduced into the resampling step. In addition to resampling the particles for the state variable of the particle filter $\mathbf{s}_{M,k}^{i}$, the corresponding Kalman filter means and covariances, and noise variance MAP estimates are also resampled according to the resampled indices.

3.3.5 MCMC Diversity Step

An undesired consequence of resampling is that the particles with high importance weights can be selected numerous times. This is shown in Figure 2.1, in which particles with high importance weights are duplicated after resampling. The resulting loss of statistical diversity in the particles is known as sample impoverishment [7], and can degrade the performance of the particle filter. It is instead preferable to have *distinct* particles values that are clustered in the regions of high importance weighting. One method of reintroducing diversity into the resampling procedure is the use of a Markov Chain Monte Carlo (MCMC) step after the resampling step [26]. The Metropolis-Hastings (MH) algorithm is used to implement the MCMC step by drawing candidates from a proposal distribution $q(\cdot)$ and using an accept-reject method to produce samples distributed from the desired invariant distribution $\pi(\cdot)$ of the chain [33]. The optimal importance function is used as the MH proposal distribution to generate the candidates s_k^*:

$$q(\mathbf{s}_k^*) = \frac{\hat{p}(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}, \mathbf{s}_k^*)p(\mathbf{s}_k^*|\mathbf{s}_{P,k-1})}{\hat{p}(\mathbf{y}_k|\mathbf{s}_{M,k-1})}. \tag{3.91}$$

We are interested in creating diversity by producing samples for the current source vector $\mathbf{s}_k$ from the posterior distribution. In this case the desired invariant distribution is the update factor for $\mathbf{s}_k$ in (2.35), which is equivalent to the optimal importance function as discussed in Section 2.4.5:

$$\pi(\mathbf{s}_k^*) = \frac{p(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}, \mathbf{s}_k^*)p(\mathbf{s}_k^*|\mathbf{s}_{P,k-1})}{p(\mathbf{y}_k|\mathbf{s}_{M,k-1})}. \tag{3.92}$$

The MH acceptance ratio r is then

$$\begin{aligned} r &= \frac{\pi(\mathbf{s}_k^*)q(\mathbf{s}_k^{(i)})}{\pi(\mathbf{s}_k^i)q(\mathbf{s}_k^*)} \\ &= \frac{p(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}^i, \mathbf{s}_k^*)\hat{p}(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}^i, \mathbf{s}_k^i)}{p(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}^i, \mathbf{s}_k^i)\hat{p}(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}^i, \mathbf{s}_k^*)}. \end{aligned} \tag{3.93}$$

The probability that the candidate is accepted, in which case $\mathbf{s}_k^i$ is replaced with the candidate $\mathbf{s}_k^*$ in the particle trajectory, is given by

$$\alpha = \min\{1, r\}. \tag{3.94}$$

3.4 Kalman Filtering for AR and FIR Coefficient Estimation

Kalman filtering is used to analytically compute the recursion for the Gaussian distributions $p(\mathbf{a}_k|\mathbf{s}_{1:k}^i, \mathbf{y}_{1:k})$ and $p(\mathbf{h}_k|\mathbf{s}_{1:k}^i, \mathbf{y}_{1:k})$ appearing in (3.50) using the source particles $\mathbf{s}_{1:k}^i$ from the particle filtering algorithm. The Kalman filter also produces the predicted distributions $p(\mathbf{a}_k|\mathbf{s}_{1:k}, \mathbf{y}_{1:k-1})$ and $p(\mathbf{h}_k|\mathbf{s}_{1:k}, \mathbf{y}_{1:k-1})$ used to develop the marginalized likelihood and prior in Section 3.3. This allows the decoupled pair of Kalman filters for each source particle to operate in parallel. MMSE estimates can

be determined from the particle filter importance weights and the Kalman state estimates, which are described in the next two subsections:

$$\hat{\mathbf{h}}_k^{\text{MMSE}} = \sum_{i=1}^{N_p} \hat{\mathbf{h}}_{k|k}^i \cdot w_k^i, \tag{3.95}$$

$$\hat{\mathbf{a}}_k^{\text{MMSE}} = \sum_{i=1}^{N_p} \hat{\mathbf{a}}_{k|k}^i \cdot w_k^i. \tag{3.96}$$

3.4.1 AR Coefficient Estimation

The Kalman filter updates the mean and covariance of the filtered Gaussian distribution for the AR coefficients $\mathbf{a}_k$:

$$p(\mathbf{a}_k|\mathbf{s}_{1:k}^i, \mathbf{y}_{1:k}) = \mathcal{N}(\hat{\mathbf{a}}_{k|k}^i, \mathbf{\Phi}_{a,k|k}^i), \tag{3.97}$$

and the predicted distribution

$$p(\mathbf{a}_k|\mathbf{s}_{1:k}^i, \mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{a}}_{k|k-1}^i, \mathbf{\Phi}_{a,k|k-1}^i). \tag{3.98}$$

The Kalman filtering algorithm for AR coefficient estimation is [58]:

$$\textit{Updated state estimate:} \quad \hat{\mathbf{a}}_{k|k}^i = \hat{\mathbf{a}}_{k|k-1}^i + \mathbf{W}_k^i(\mathbf{s}_k^i - \mathbf{S}_{k-1}^i\hat{\mathbf{a}}_{k|k-1}^i), \tag{3.99}$$

$$\textit{Updated state covariance:} \quad \mathbf{\Phi}_{a,k|k}^i = \mathbf{\Phi}_{a,k|k-1}^i + \mathbf{W}_k^i\mathbf{S}_{k-1}^i\mathbf{\Phi}_{a,k|k-1}^i, \tag{3.100}$$

$$\textit{Innovation covariance:} \quad \mathbf{R}_k^i = \mathbf{S}_{k-1}^i\mathbf{\Phi}_{a,k|k-1}^i\mathbf{S}_{k-1}^{(i)\text{T}} + \hat{\mathbf{\Sigma}}_v^i, \tag{3.101}$$

$$\textit{Kalman gain:} \quad \mathbf{W}_k^i = \mathbf{\Phi}_{a,k|k-1}^i\mathbf{S}_{k-1}^{(i)\text{T}}(\mathbf{R}_k^i)^{-1}, \tag{3.102}$$

$$\textit{Predicted state estimate:} \quad \hat{\mathbf{a}}_{k|k-1}^i = a_a\hat{\mathbf{a}}_{k-1|k-1}^i, \tag{3.103}$$

$$\textit{Predicted state covariance:} \quad \mathbf{\Phi}_{a,k|k-1}^i = a_a^2\mathbf{\Phi}_{a,k-1|k-1}^i + \mathbf{\Sigma}_a. \tag{3.104}$$

3.4.2 FIR Coefficient Estimation

Similarly, the filtered and predicted Gaussian distributions for the channel $\mathbf{h}_k$ are

$$p(\mathbf{h}_k|\mathbf{s}_{1:k}^i, \mathbf{y}_{1:k}) = \mathcal{N}(\hat{\mathbf{h}}_{k|k}^i, \mathbf{\Phi}_{h,k|k}^i), \tag{3.105}$$

$$p(\mathbf{h}_k|\mathbf{s}_{1:k}^i, \mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{h}}_{k|k-1}^i, \mathbf{\Phi}_{h,k|k-1}^i). \tag{3.106}$$

The Kalman filtering algorithm for FIR coefficient estimation is:

$$\begin{aligned}
&\textit{Updated state estimate:} && \hat{\mathbf{h}}^i_{k|k} = \hat{\mathbf{h}}^i_{k|k-1} + \mathbf{W}^i_k(\mathbf{y}_k - \mathbf{T}^i_k\hat{\mathbf{h}}^i_{k|k-1}), && (3.107)\\
&\textit{Updated state covariance:} && \mathbf{\Phi}^i_{h,k|k} = \mathbf{\Phi}^i_{h,k|k-1} + \mathbf{W}^i_k\mathbf{T}^i_k\mathbf{\Phi}^i_{h,k|k-1}, && (3.108)\\
&\textit{Innovation covariance:} && \mathbf{Q}^i_k = \mathbf{T}^i_k\mathbf{\Phi}_{h,k|k-1}\mathbf{T}^{(i)\mathrm{T}}_k + \hat{\mathbf{\Sigma}}^i_w, && (3.109)\\
&\textit{Kalman gain:} && \mathbf{W}^i_k = \mathbf{\Phi}^i_{h,k|k-1}\mathbf{T}^{(i)\mathrm{T}}_k(\mathbf{Q}^i_k)^{-1}, && (3.110)\\
&\textit{Predicted state estimate:} && \hat{\mathbf{h}}^i_{k|k-1} = a_h\hat{\mathbf{h}}^i_{k-1|k-1}, && (3.111)\\
&\textit{Predicted state covariance:} && \mathbf{\Phi}^i_{h,k|k-1} = a_h^2\mathbf{\Phi}^i_{h,k-1|k-1} + \mathbf{\Sigma}_h. && (3.112)
\end{aligned}$$

3.5 MAP Estimation of Noise Variances

MAP estimates of the noise covariance parameters $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_w$ are now developed. For each particle, an estimate $\hat{\mathbf{\Sigma}}^i_v$ of the source noise covariance is computed by conditioning on $\mathbf{s}^i_{1:k}$ from the particle filter, and $\hat{\mathbf{a}}^i_k$ from the Kalman filter. In parallel, the computation of the estimate $\hat{\mathbf{\Sigma}}^i_w$ can be performed conditioned on $\mathbf{s}^i_{1:k}$ and $\hat{\mathbf{h}}^i_k$. The MAP estimates for each particle can be combined into a single MMSE estimate of the variance terms using the importance weights from the particle filter:

$$(\hat{\sigma}^2_{k,v,n})^{\mathrm{MMSE}} = \sum_{i=1}^{N_p} \hat{\sigma}^{2(i)}_{k,v,n} \cdot w^i_k, \quad n = 1, 2, \ldots, N, \tag{3.113}$$

$$(\hat{\sigma}^2_{k,w,j})^{\mathrm{MMSE}} = \sum_{i=1}^{N_p} \hat{\sigma}^{2(i)}_{k,w,j} \cdot w^i_k, \quad j = 1, 2, \ldots, J. \tag{3.114}$$

The calculation of the terms $\hat{\sigma}^{2(i)}_{k,v,n}$ and $\hat{\sigma}^{2(i)}_{k,w,j}$ now follows.

3.5.1 Source Noise Variance Estimation

To calculate the MAP estimate of the diagonal source noise covariance $\mathbf{\Sigma}_v$, consider only the terms dependent on the noise variance $\sigma^2_{v,n}$ of the n^{th} source in the joint

posterior distribution (3.37):

$$\begin{aligned} p(\boldsymbol{\theta}_{1:k}|\mathbf{y}_{1:k}) &\propto (\sigma_{v,n}^2)^{-\frac{\upsilon_n}{2}-1} \exp\left[-\frac{\gamma_n}{2\sigma_{v,n}^2}\right] \\ &\prod_{\tilde{k}=1}^{k} \frac{1}{(\sigma_{v,n}^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma_{v,n}^2}(\mathbf{s}_{\tilde{k}}[n] - \mathbf{s}_{P,\tilde{k}-1,n}^{\mathrm{T}}\mathbf{a}_{\tilde{k},n})^2\right] \\ &\propto (\sigma_{v,n}^2)^{-(\frac{k}{2}+\frac{\upsilon_n}{2}+1)} \exp\left[-\frac{1}{\sigma_{v,n}^2}\left(\frac{\gamma_n}{2} + \frac{1}{2}\sum_{\tilde{k}=1}^{k}(\mathbf{s}_{\tilde{k}}[n] - \mathbf{s}_{P,\tilde{k}-1,n}^{\mathrm{T}}\mathbf{a}_{\tilde{k},n})^2\right)\right]. \end{aligned} \tag{3.115}$$

It is seen through comparison with (3.36) that the last line is an Inverse Gamma distribution $\mathcal{IG}(\upsilon, \gamma)$ with

$$\upsilon = \frac{k}{2} + \frac{\upsilon_n}{2}, \tag{3.116}$$

$$\gamma = \frac{\gamma_n}{2} + \frac{1}{2}\sum_{\tilde{k}=1}^{k}(\mathbf{s}_{\tilde{k}}[n]^i - \mathbf{s}_{P,\tilde{k}-1,n}^{(i)\mathrm{T}}\mathbf{a}_{\tilde{k},n}^i)^2. \tag{3.117}$$

It is readily verified that the mode of the Inverse Gamma distribution is $\frac{\gamma}{\upsilon+1}$. Therefore the MAP estimate corresponding to the i^{th} particle is

$$\hat{\sigma}_{k,v,n}^{2(i)} = \frac{\frac{\gamma_n}{2} + \frac{1}{2}\sum_{\tilde{k}=1}^{k}(\mathbf{s}_{\tilde{k}}[n]^i - \mathbf{s}_{P,\tilde{k}-1,n}^{(i)\mathrm{T}}\mathbf{a}_{\tilde{k},n}^i)^2}{\frac{k}{2} + \frac{\upsilon_n}{2} + 1}. \tag{3.118}$$

Assuming a non-informative prior using $\upsilon_n = 0, \gamma_n = 0$, this reduces to

$$\hat{\sigma}_{k,v,n}^{2(i)} = \frac{\frac{1}{2}\sum_{\tilde{k}=1}^{k}(\mathbf{s}_{\tilde{k}}[n]^i - \mathbf{s}_{P,\tilde{k}-1,n}^{(i)\mathrm{T}}\mathbf{a}_{\tilde{k},n}^i)^2}{\frac{k}{2} + 1}. \tag{3.119}$$

3.5.2 Measurement Noise Variance Estimation

The MAP estimation of the diagonal measurement noise covariance $\boldsymbol{\Sigma}_w$ follows from the method for the source covariance above, where assuming a non-informative prior $\upsilon_{w,j} = 0, \gamma_{w,j} = 0$ for the j^{th} sensor noise variance, the MAP estimate corresponding to the i^{th} particle is

$$\hat{\sigma}_{k,w,j}^{2(i)} = \frac{\frac{1}{2}\sum_{\tilde{k}=1}^{k}(\mathbf{y}_{\tilde{k}}[j] - \mathbf{s}_{L,\tilde{k}}^{(i)\mathrm{T}}\mathbf{h}_{\tilde{k},j}^i)^2}{\frac{k}{2} + 1}. \tag{3.120}$$

In implementation, both the source and measurement noise variances can also be recursively updated using an exponential weighting factor $0 \leq \lambda < 1$ as shown for only the source variance:

$$\xi^i_{k,v,n} = \lambda \xi^i_{k,v,n} + (\mathbf{s}_k[n]^i - \mathbf{s}^{(i)\mathrm{T}}_{P,k-1,n}\mathbf{a}^i_{k,n})^2, \tag{3.121}$$

$$\hat{\sigma}^{2(i)}_{k,v,n} = \frac{\frac{1}{2}\xi^i_{k,v,n}}{\frac{1-\lambda^k}{2(1-\lambda)} + 1}. \tag{3.122}$$

and $\xi^i_{0,v,n} = 0$. This allows the algorithm to follow small fluctuations in the noise variances.

3.6 Identifiability Conditions

The discussion of identifiability conditions for the proposed Bayesian formulation is based on [63]. It has been shown [48],[46] that sufficient conditions exist for blind identification of time-invariant MIMO systems with FIR channels and coloured inputs using second-order statistics (SOS). The proposed state space model consists of FIR channels and coloured inputs. Therefore we use the existing conditions in [46], based on results from [48], to first justify identifiability of the sources, channels and measurement noise variances. It is also noted that the posterior distribution using a Bayesian approach captures all statistical information about the system, including SOS of the outputs.

It is well-known that blind identification of the sources and/or channel for the blind system identification problem is subject to an inherent permutation and scaling ambiguity [14]. We now quote the sufficient identifiability conditions *directly* from [46]:

> *A FIR MIMO system is identifiable (i.e. the power spectral matrix of the outputs $\mathbf{S}_{yy}(z)$ implies the channel matrix $\mathbf{H}(z)$ up to a column-wise scaling and column-wise permutation) if we have the following:*
>
> *A1) $\mathbf{H}(z)$ is irreducible.*

A2) The input power spectral matrix $\mathbf{S}_{xx}(z)$ *is diagonal and of distinct (polynomial or rational) functions.*

A3) The noise power spectral matrix $\mathbf{S}_{ww}(z) = p(z)p(z^{-1})\mathbf{I}$ *and* $J(sensors) > N(sources)$.

In the first condition, an irreducible polynomial matrix has full column rank for all z except $z = 0$. The second condition ensures the sources are mutually uncorrelated with distinct power spectrum [48]. The third condition specifies that the measurement noise variance can also be identified provided the noise elements are mutually uncorrelated with equal power spectra. The condition of more sensors than sources ensures that the noise variance can be determined from the smallest real eigenvalue of $\mathbf{S}_{yy}(z)$ [46]. Refer to [50] for more detailed descriptions of irreducibility, and polynomial and rational functions. These identifiability conditions were used in the development of the existing BIDS algorithm referenced in Section 1.3.

Given that the source can be identified under these sufficient conditions, we can also identify the AR coefficients and source noise variance using previous identifiabilty results for autoregressive time series [11]. Therefore the source, channel, AR coefficients and noise variances are jointly identifiable, assuming the system is time-invariant. This implies that the joint posterior distribution $p(\boldsymbol{\theta}_{1:K}|\mathbf{y}_{1:k})$ has a set of unique maximum values which identify $\boldsymbol{\theta}_{1:K}$ up to the inherent permutation and scaling ambiguity. Therefore, if the problem has a solution using e.g. BIDS, then it is identifiable using our Bayesian approach. The case of a time-varying system is beyond the scope of the current thesis. Whether these sufficient identifiability conditions can be relaxed using the proposed algorithm is an open research question.

Chapter 4

Performance Evaluation

This chapter presents simulation results for the developed marginalized particle filtering algorithm. First, the PCRB for the given state space model is derived. Definitions of the parameter settings and performance measures used in the simulations are then described. The simulation results demonstrate the performance of the algorithm in estimating the unknown system.

4.1 PCRB for Blind System Identification

The Posterior Cramér-Rao bound for the discrete-time nonlinear filtering problem was summarized in Section 2.5. The random parameter we are estimating for the blind system identification problem is $\boldsymbol{\theta}_{1:K}$ from (3.28). The derivation of the PCRB is simplified by assuming the noise variances and AR coefficients are known. The reduced state variable under consideration for this section is then

$$\boldsymbol{\phi}_k = [\mathbf{h}_k^{\mathrm{T}}, \mathbf{s}_{M,k}^{\mathrm{T}}]^{\mathrm{T}}. \tag{4.1}$$

The dimension of $\boldsymbol{\phi}_k$ is denoted by r, which is equal to the sum of the dimension of $\mathbf{h}_k$ (JNL) and $\mathbf{s}_{M,k}$ (NM):

$$r = JNL + MN. \tag{4.2}$$

The state transition equation in terms of the state variable $\boldsymbol{\phi}_k$ is written in the form of (2.6) as

$$\begin{aligned} \boldsymbol{\phi}_{k+1} &= \mathbf{f}_k(\boldsymbol{\phi}_k, \mathbf{u}_k) \\ &= \tilde{\mathbf{F}}_k \boldsymbol{\phi}_k + \mathbf{u}_k. \end{aligned} \tag{4.3}$$

The definitions for the matrix $\tilde{\mathbf{F}}_k$ and the vector $\mathbf{u}_k$ follow from the state equations for $\mathbf{h}_k$ (3.27) and $\mathbf{s}_{M,k}$ (3.52):

$$\begin{aligned} \begin{bmatrix} \mathbf{h}_{k+1} \\ \mathbf{s}_{M,k+1} \end{bmatrix} &= \begin{bmatrix} a_h \mathbf{I}_{NJL} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{A}}_{k+1} \end{bmatrix} \begin{bmatrix} \mathbf{h}_k \\ \mathbf{s}_{M,k} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{h,k} \\ \tilde{\mathbf{v}}_k \end{bmatrix} \\ &\triangleq \tilde{\mathbf{F}}_k \boldsymbol{\phi}_k + \mathbf{u}_k \end{aligned} \tag{4.4}$$

The corresponding measurement equation in the form of (2.7) follows from the equivalent representations in (3.55) and (3.19):

$$\mathbf{g}_k(\boldsymbol{\phi}_k, \mathbf{w}_k) = \tilde{\mathbf{H}}_k \mathbf{s}_{M,k} + \mathbf{w}_k \tag{4.5}$$

$$= \mathbf{T}_k \mathbf{h}_k + \mathbf{w}_k. \tag{4.6}$$

The PCRB is given as a lower bound for the covariance of the one-step-ahead predictor in (2.53). MMSE one-step-ahead predicted estimates of $\boldsymbol{\phi}_k$ given the measurements $\mathbf{y}_{1:k-1}$ are obtained for the source $\mathbf{s}_k$ from (3.90), and for the channel $\mathbf{h}_k$ by substituting the predicted channel from the Kalman filter in (3.111) into (3.95).

The matrix expressions $\mathbf{F}_k$, $\mathbf{R}_k^{-1}$, $\mathbf{G}_k$, $\mathbf{Q}_k$ and $\mathbf{P}_0$ used in the recursion for $\mathbf{P}_{k+1}$ in (2.54) are derived in Appendix C.

4.2 Simulation Definitions

Definitions of parameter settings and performance measures used in the simulations are now developed. First, the ratio of the source noise variance to the measurement noise variance is defined as the Noise Variance Ratio (NVR):

$$\text{NVR} = 10 \log_{10} \left(\frac{\sigma_v^2}{\sigma_w^2} \right). \tag{4.7}$$

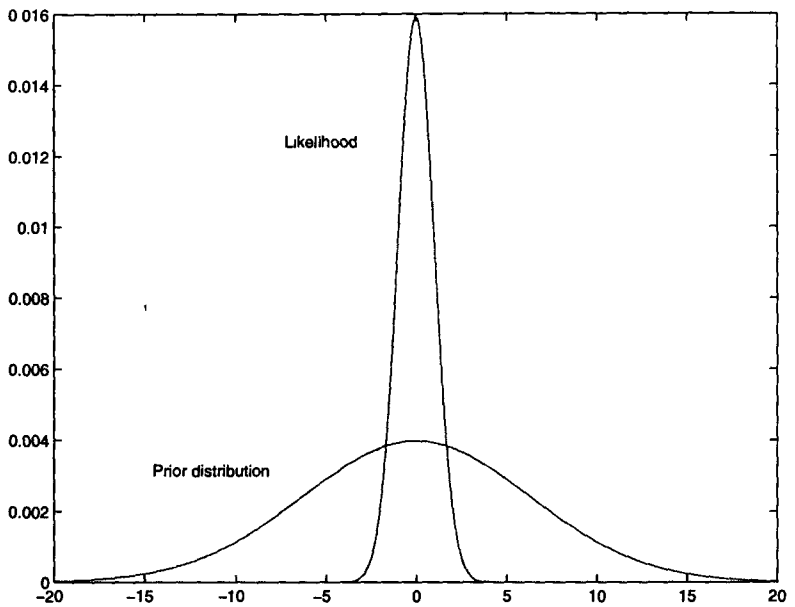


Figure 4.1: Example of dominant likelihood scenario

The ratio of the two noise variance parameters indicates a measure of the dominance of the prior over the likelihood. An example of a scenario with a dominant likelihood (small NVR), is illustrated in Figure 4.2.

The signal-to-noise ratio for the n^{th} source $\mathrm{SNR}_{s,n}$ is defined from (3.1) by identifying $\mathbf{a}_{k,n}^{\mathrm{T}}\mathbf{s}_{P,k-1,n}$ as the signal component, and $\mathbf{v}_{k-1}[n]$ as the noise component:

$$\mathrm{SNR}_{s,n} = 10\log_{10}\left(\frac{\mathbb{E}\{\mathbf{a}_{k,n}^{\mathrm{T}}\mathbf{s}_{P,k-1,n}\mathbf{s}_{P,k-1,n}^{\mathrm{T}}\mathbf{a}_{k,n}\}}{\sigma_{v,n}^2}\right). \tag{4.8}$$

For the case of a single source following a first-order AR model with time-invariant AR coefficient $\mathbf{a}_k = a$, the expression can be evaluated as

$$\mathrm{SNR}_{s,n} = 10\log_{10}\left(\frac{\sigma_v^2}{1-a^2}\right). \tag{4.9}$$

In general, $\mathrm{SNR}_{s,n}$ is not tractable in closed-form since it requires the computation of the covariance matrix for the AR process $\mathbf{s}_{k,n}$ [52]. The $\mathrm{SNR}_{s,n}$ is computed numerically based on the statistical realization in the simulation, except when it can be computed using (4.9). The $\mathrm{SNR}_{s,n}$ determines the predictability of the source

equation, which influences the effectiveness of the prior in generating particles and the performance of the Kalman filter in estimating $\mathbf{a}_{k,n}$.

The signal-to-noise ratio for the j^{th} measurement $\text{SNR}_{y,j}$ is defined as

$$\text{SNR}_{y,j} = 10\log_{10}\left(\frac{\mathbb{E}\{\mathbf{h}_{k,j}^{\text{T}}\mathbf{s}_{L,k}\mathbf{s}_{L,k}^{\text{T}}\mathbf{h}_{k,j}\}}{\sigma_{w,j}^2}\right), \tag{4.10}$$

where from (3.15) the term $\mathbf{h}_{k,j}^{\text{T}}\mathbf{s}_{L,k}$ is identified as the signal component and $\mathbf{w}_k[j]$ is the noise component. For a single source from a first-order AR model with time-invariant AR coefficient $\mathbf{a}_k = a$ and channel $\mathbf{h}_{k,j} = \mathbf{h}_j$, the expression simplifies to

$$\text{SNR}_{y,j} = 10\log_{10}\left(\frac{\mathbf{h}_j^{\text{T}}\mathbf{R}_s\mathbf{h}_j}{\sigma_{w,j}^2}\right). \tag{4.11}$$

The Toeplitz matrix $\mathbf{R}_s$ is formed from the correlation function $r_s(\tau)$ [52]:

$$\begin{aligned}
\mathbf{R}_s &= \mathbb{E}\{\mathbf{s}_{L,k}\mathbf{s}_{L,k}^{\text{T}}\} \\
&= \begin{bmatrix} r(0) & r(1) & r(2) & \dots & r(L-1) \\ r(1) & r(0) & r(1) & \dots & r(L-2) \\ r(2) & r(1) & r(0) & \dots & r(L-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(L-1) & r(L-2) & r(L-3) & \dots & r(0) \end{bmatrix} \\
r(\tau) &= \sigma_v^2\frac{a^\tau}{1-a^2}, \quad \tau = 0,1,\dots,L-1
\end{aligned} \tag{4.12}$$

In the general case, $\text{SNR}_{y,j}$ is computed numerically. The $\text{SNR}_{y,j}$ influences the contribution of the likelihood in generating particles, and the performance of the Kalman filter in estimating $\mathbf{h}_{k,j}$.

The performance of the marginalized particle filtering algorithm in estimating the general variable $\boldsymbol{\alpha}_k$ is measured using the mean square error (MSE)

$$\text{MSE}(\hat{\boldsymbol{\alpha}}_k) = 10\log_{10}\left(\frac{1}{N_t}\sum_{t=1}^{N_t}\|\boldsymbol{\alpha}_k^t - \hat{\boldsymbol{\alpha}}_k^t\|_2^2\right), \tag{4.13}$$

where the squared error between the true variable $\boldsymbol{\alpha}_k^t$ and the estimate $\hat{\boldsymbol{\alpha}}_k^t$ at time k is averaged over Monte Carlo simulations $t = 1$ to N_t. This general expression is

used to evaluate the MSE of the state space variables by replacing $\boldsymbol{\alpha}_k$ with each of the vectors $\mathbf{s}_k$, $\mathbf{h}_k$, $\mathbf{a}_k$, $\text{diag}(\boldsymbol{\Sigma}_w)$, $\text{diag}(\boldsymbol{\Sigma}_v)$. The plot of the MSE over time is referred to as the learning curve, and illustrates the convergence speed of the algorithm. The average performance over time is computed by averaging over the time indices $k = k_{ss}$ to K:

$$\text{MSE}(\hat{\boldsymbol{\alpha}}_{k_{ss}:K}) = 10\log_{10}\left(\frac{1}{K-k_{ss}+1}\sum_{k=k_{ss}}^{K}\left(\frac{1}{N_t}\sum_{t=1}^{N_t}\|\boldsymbol{\alpha}_k^t - \hat{\boldsymbol{\alpha}}_k^t\|_2^2\right)\right). \tag{4.14}$$

The effect of transient initialization error can be removed by selecting the steady-state time index $k_{ss} > 1$. The value $k_{ss} = 101$ is used for the simulations.

In order to evaluate the above MSE, it is necessary to resolve the inherent scale ambiguity in the blind estimation problem. Based on a least squares formulation, the unknown scale factor β_n for the n^{th} source is estimated as

$$\beta_n = \frac{\hat{\mathbf{s}}_{k_{ss}:K}[n]^{\mathrm{T}}\mathbf{s}_{k_{ss}:K}[n]}{\hat{\mathbf{s}}_{k_{ss}:K}[n]^{\mathrm{T}}\hat{\mathbf{s}}_{k_{ss}:K}[n]}. \tag{4.15}$$

The estimated source $\hat{\mathbf{s}}_n$ is first multiplied by β_n, and the corresponding estimated source variance is multiplied by β_n^2, before computing the MSE. A least squares problem is also solved for the channel scaling factor by concatenating the channel coefficients into a single vector.

4.3 Simulation Results

Simulation results are now presented to support the proposed algorithm. A SIMO system with $N = 1$ source and $J = 2$ sensors was run for $N_t = 50$ Monte Carlo trials. The dynamical state space model was simulated for time steps $k = 1$ to $K = 500$. The particle filter algorithm begins after the AR models for the state variables $\mathbf{s}_k$, $\mathbf{h}_k$, and $\mathbf{a}_k$ are initiated for $K_{init} = 2000$ time steps. This is to eliminate the starting transients and produce steady state AR processes for these variables.

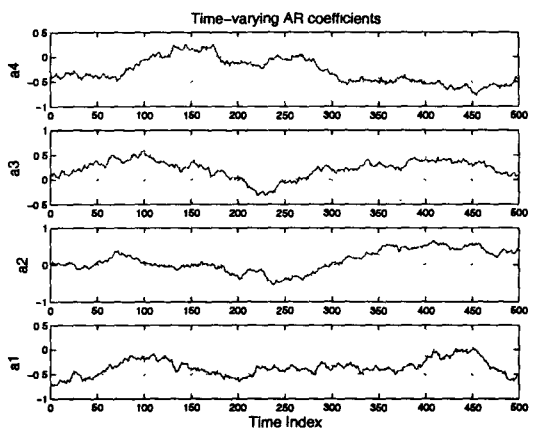


Figure 4.2: Time-varying AR coefficients

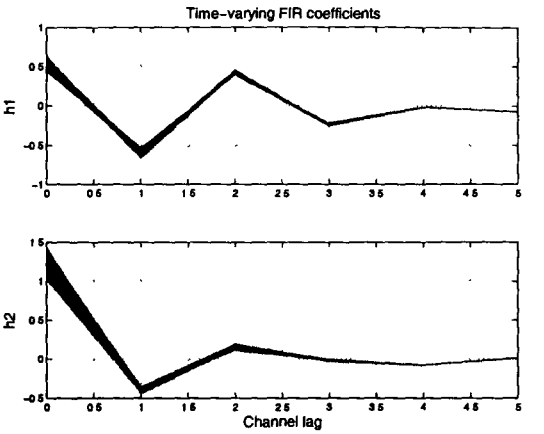


Figure 4.3: Time-varying FIR coefficients

The initial $P = 4$ order AR coefficient vector $\mathbf{a}_0$ is generated from a P-order low-pass Butterworth filter with normalized cutoff frequency $w_n = 0.25$. The time-varying state $\mathbf{a}_k$ is then generated from the first-order AR model in (3.14) using $a_a = 0.9999$ and $\Sigma_a = 0.001\mathbf{I}_P$. Figure 4.2 contains an example of the time-varying AR coefficients. Stability of the source AR model is enforced by scaling the maximum absolute value of the AR coefficient poles to just less than one whenever a pole passes outside the unit circle.

The initial $L = 6$ order FIR channel vectors $\mathbf{h}_{0,j}$ from the source to the j^{th} sensor are produced from independent draws from a zero-mean Gaussian distribution with exponentially decaying covariance matrix given by

$$\Sigma_{h,0} = \text{diag}\left([e^{-\frac{L-1}{WL}}, e^{-\frac{L-2}{WL}}, \ldots, e^{-\frac{0}{WL}}]\right). \tag{4.16}$$

The decay factor $W = 0.15$ was used in the simulations. The time-varying channel vector $\mathbf{h}_k$ is generated from the first-order AR model in (3.27) using $a_h = 0.9999$ and $\Sigma_h = (1 - a_h^2)\Sigma_{h,0}$. Figure 4.3 demonstrates typical FIR vectors from all time steps superimposed on each other. The simulations assume that the FIR channel is varying slowly compared to the AR coefficients to model a speech source vocal tract that changes more rapidly than the audio channel [6].

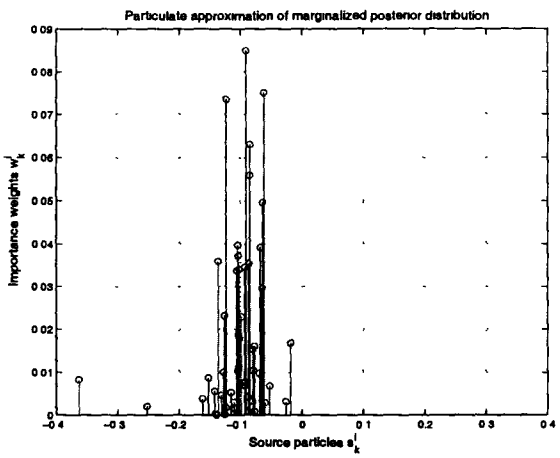


Figure 4.4: Particulate approximation to the marginalized posterior distribution

The source noise variance was set to $\sigma_v^2 = 0.01$. The measurement noise variance σ_w^2 was set equal for both sensors, and scaled relative to the source noise variance and AR/FIR coefficients to achieve an average SNR_y over the Monte Carlo runs of approximately 15 dB. The numerical results of simulation parameters averaged over the Monte Carlo runs are summarized in Table 4.1.

Parameter	σ_w^2	NVR	SNR_s	SNR_y
Value	0.00056	14.30	1.03	17.31

Table 4.1: Average simulation parameter settings

The particle filter was run with $N_p = 50$ particles, and the threshold N_{thresh} for the approximate effective sample size was set to $N_p/3$. Figure 4.4 plots an example of the source particles and importance weights that form the particulate approximation to the marginalized posterior distribution at an arbitrary time step. A single Monte Carlo run of the algorithm implemented in *Matlab* took approximately 3 minutes (0.18 seconds per time step) on a standard 2 GHz PC. The computational complexity of a general marginalized particle filtering algorithm is studied in [51], including an analysis of equivalent flop count, guidelines for optimal partitioning of the estimation problem, and suggestions for algorithm modifications to reduce complexity.

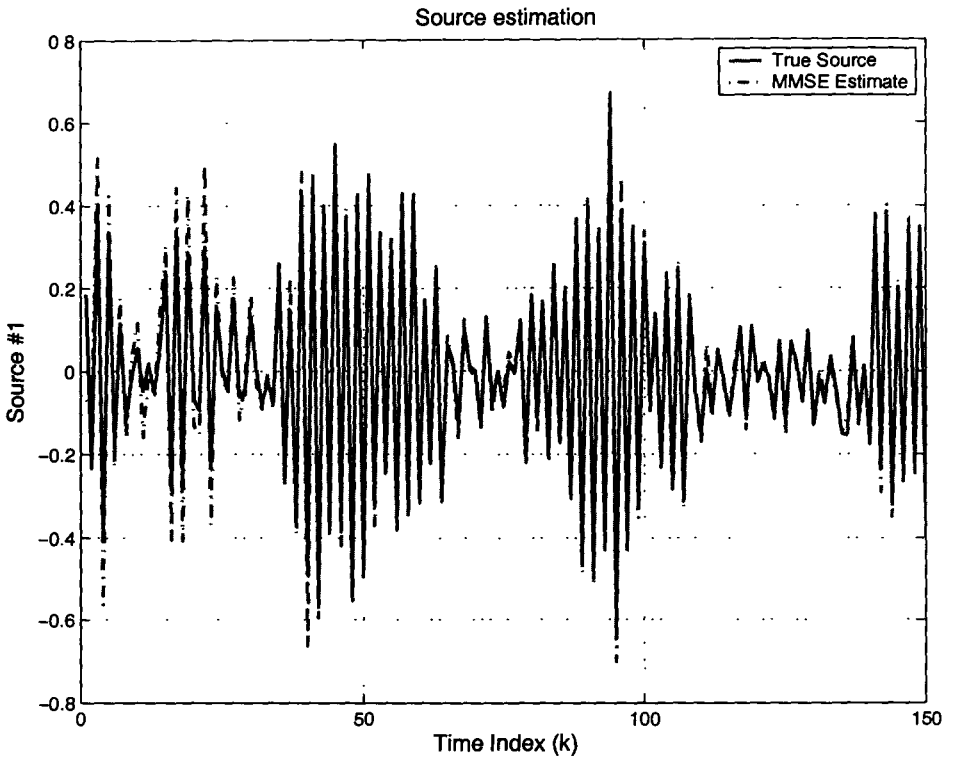
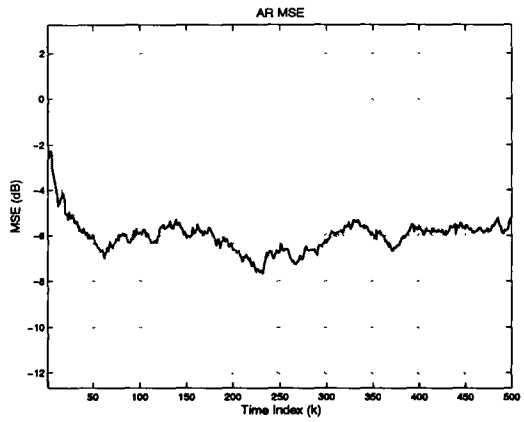


Figure 4.5: SIMO source estimation

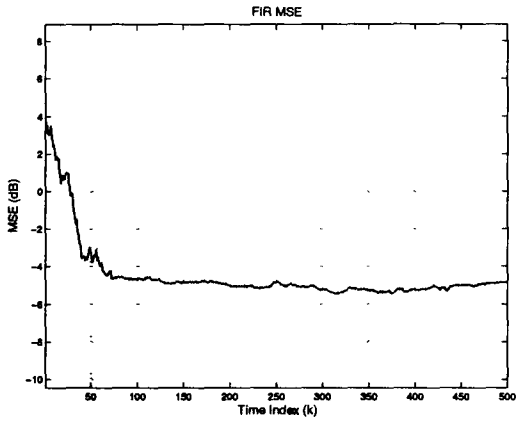
Figure 4.5 compares the MMSE estimate of the source with the true source up to time index $k = 150$ from a typical Monte Carlo run. The learning curves of MSE over time from (4.13) are presented in Figure 4.6 for the AR/FIR coefficients, source and measurement noise variances, and source. The time-averaged MSE values computed over the Monte Carlo trials using (4.14) are shown in Table 4.2 for all the unknown state space variables. Normalizing the MSE of the multidimensional variables $\mathbf{a}_k \in \mathbb{R}^{NP\times 1}$ and $\mathbf{h}_k \in \mathbb{R}^{JNL\times 1}$ by the state dimension results in values of -12.09 and -15.83 dB, respectively.

Variable	$\mathbf{s}_k$	$\mathbf{h}_k$	$\mathbf{a}_k$	σ_v^2	σ_w^2
MSE	-23.08	-5.03	-6.07	-46.31	-72.40

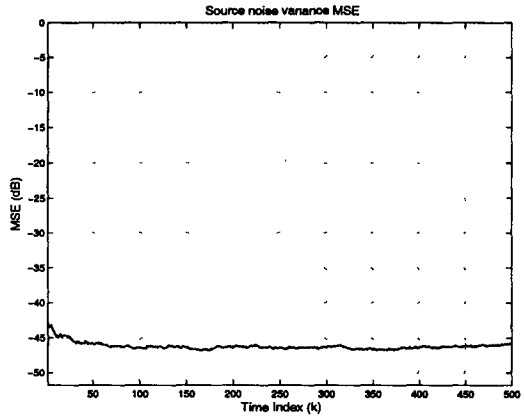
Table 4.2: MSE simulation results



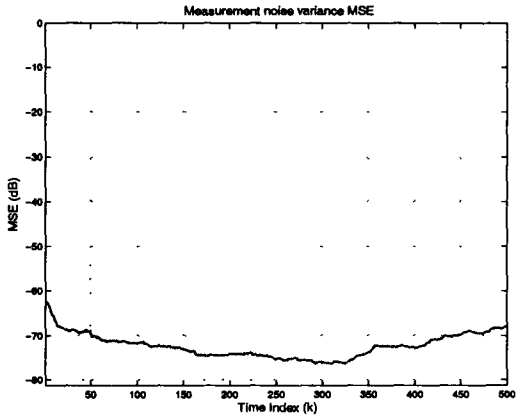
(a) AR coefficients



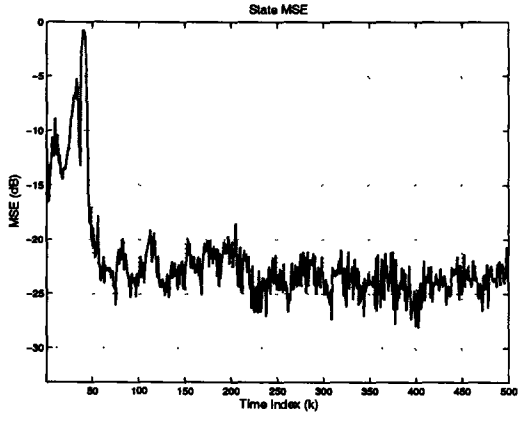
(b) FIR coefficients



(c) Source noise variance



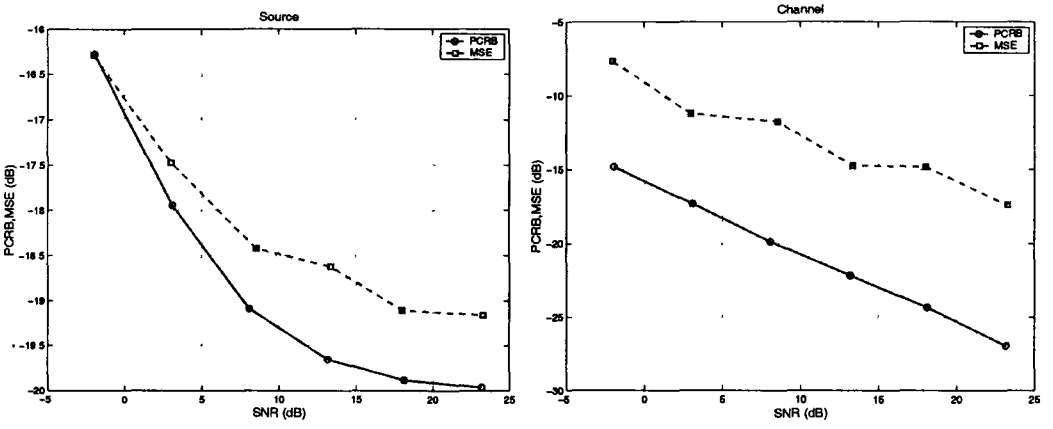
(d) Measurement noise variance



(e) Source

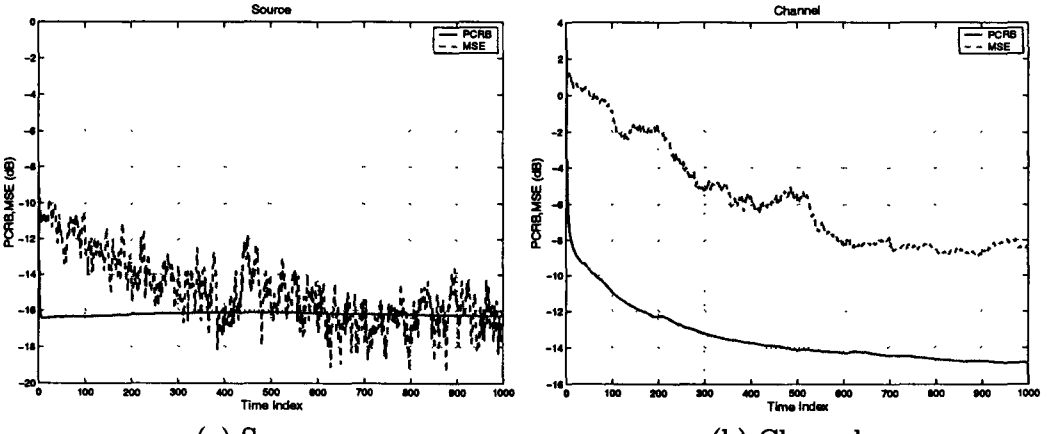
Figure 4.6: MSE learning curves

For comparison with the PCRB, the noise variances and AR coefficients are set to their true values as described in Section 4.1. It is noted that the algorithm had negligible performance loss when these parameters were estimated instead of assumed known. With the noise variances known the inherent scaling ambiguity reduces to a sign ambiguity, which is resolved before computing the MSE. The state space model was simulated for $K = 1000$ time steps using a time-invariant $P = 1$ order source AR model with coefficient value set to $a = 0.9535$, corresponding to $\text{SNR}_s = 10$ dB. The model order for the time-varying FIR channel was set to $L = 3$, while all other state space parameters are as described above. The required expectations for the PCRB matrices were evaluated by averaging 5000 *i.i.d.* realizations of the state space model. The MSE of the MMSE one-step-ahead predicted estimates of the source and channel were averaged over $N_t = 50$ Monte Carlo runs. One Monte Carlo run was discarded since the particle filter momentarily tracked a scaled version of the source, rather than to within a sign ambiguity. Figure 4.7 plots a range of numerical SNR_y values versus the PCRB and MSE both averaged over the last 100 time steps. The source PCRB saturates at high SNR_y to a value of -20 dB since the performance of the one-step-ahead predictor for the source is limited by the source noise variance $\sigma_v^2 = 0.01$. Figure 4.8 compares the PCRB and MSE for the source and channel over time $k = 1$ to 1000 with measurement SNR_y approximately 0 dB. The figures illustrate that the source MSE compares well with the PCRB, falling within 1 dB of the lower bound over the range of SNR_y. The performance of the 6-dimensional channel vector falls within $8 - 10$ dB of the PCRB. For the application of speech dereverberation, we are most interested in directly recovering the source, rather than in obtaining a channel estimate. As such, the simulation results confirm that the Monte Carlo-based particle filtering approach to the blind system identification problem achieves an acceptable level of performance. It is also noted that this performance is achieved using a moderate number of particles.



(a) Source (b) Channel

Figure 4.7: Comparison of MSE with PCRB over SNR_y



(a) Source (b) Channel

Figure 4.8: Comparison of MSE with PCRB over time

Chapter 5

Conclusions and Future Research

5.1 Conclusions

A marginalized particle filtering algorithm has been proposed for the blind system identification problem of estimating AR sources mixed by FIR channels in the presence of additive white noise. The sequential estimation of the dynamic state space model was implemented using sequential Monte Carlo methods which handle the nonlinearity in the model. The conditionally linear-Gaussian structure in the model was exploited using the Rao-Blackwellisation technique which resulted in a more efficient particle filter. By marginalizing out the AR and FIR coefficients, the source is directly recovered which makes the proposed algorithm suitable for the speech dereverberation problem. The performance of the particle filter was also improved through the use of a Gaussian approximation to the optimal importance function. Simulation results have demonstrated the effectiveness of the method in estimating each of the state space variables. Comparison of the results with the derived Posterior Cramér-Rao bound has also provided theoretical validation on the performance.

5.2 Contributions to the Scientific Literature

The research summarized in this thesis has also been published or submitted to a number of conferences. A journal paper is under preparation. Conference paper (1) is based on previous research not summarized here. Paper (5) is a result of collaborative research and is also not included in the thesis.

- Journal Papers

    1. Marginalized particle filtering for blind system identification. *Under preparation.*

- Conference Papers

    1. M.J. Daly and J.P. Reilly, "Blind deconvolution using Bayesian methods with application to the dereverberation of speech", *IEEE Proc. International Conference on Acoustics, Speech and Signal Processing*, Montreal, Canada. May 17-21 2004.
    2. M.J. Daly, J.P. Reilly and J.H. Manton, Invited Paper, "A Bayesian approach to blind source recovery", to appear *Thirty-Eighth Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA. Nov. 7-10, 2004.
    3. M.J. Daly, J.P. Reilly and J.H. Manton, "A Bayesian approach to blind source recovery", to appear *IMA Conference on Mathematics in Signal Processing VI*, Cirencester, England. Dec. 14-16 2004
    4. M.J. Daly, J.P. Reilly and M.R. Morelande, "Rao-Blackwellised particle filtering for blind system identification", submitted to *IEEE ICASSP 2005.*
    5. D. Yee, M.J. Daly, J.P. Reilly and T. Kirubarajan, "Rao-Blackwellised particle filters for mixture Gaussian processes". *Under preparation.*

5.3 Future Research

Two directions for future research are provided. First, a novel extension to sequential importance sampling is developed in detail. The application of the proposed marginalized particle filtering algorithm to the speech dereverberation problem using a filter bank structure is then outlined.

5.3.1 Block Sequential Importance Sampling

The block sequential importance sampling (BSIS) formulation of the particle filter is now introduced. We first revisit the proposed marginalized particle filter algorithm based on the classical SIS approach. The algorithm is analyzed to determine how the structure of the convolution state space model can be exploited to further improve the efficiency of the particle filtering algorithm.

In Section 3.3, the particle filtering problem for nonlinear estimation of the sources is formulated in terms of the state variable $\mathbf{s}_{M,k}$ given by

$$\mathbf{s}_{M,k} = [\mathbf{s}_{k-M+1}^{\mathrm{T}}, \mathbf{s}_{k-M+2}^{\mathrm{T}}, \ldots, \mathbf{s}_{k}^{\mathrm{T}}]^{\mathrm{T}}. \tag{5.1}$$

The importance function for $\mathbf{s}_{M,k}$ is shown to reuse the particle values for $\mathbf{s}_{k-M+1:k-1}$ from the previous time steps. This was a result of the state update for $\mathbf{s}_{M,k}$ illustrated in (3.52). The particles for the remaining quantity $\mathbf{s}_k$ in $\mathbf{s}_{M,k}$ are generated from the Gaussian importance function in (3.82). This form of the importance function in the SIS implementation leads to the following two observations:

1. The current measurement $\mathbf{y}_k$ is only incorporated into the particle generation process for $\mathbf{s}_k$ in the form of a Kalman filter update. From the form of the convolution model

$$\mathbf{y}_k = \sum_{\ell=0}^{L-1} \mathbf{H}_{k,\ell}\mathbf{s}_{k-\ell} + \mathbf{w}_k, \tag{5.2}$$

the current measurement is directly related to the past source vectors $\mathbf{s}_{k-\ell}$, $\ell = 1, 2, \ldots, L-1$ as well. Our first goal is to use the information in $\mathbf{y}_k$ to generate

new particle values for the quantities $\mathbf{s}_{k-M+1:k-1}$ appearing in $\mathbf{s}_{M,k}$.

2. Diversity is only introduced on $\mathbf{s}_k$, which can lead to sample impoverishment in the state vector $\mathbf{s}_{M,k}$. Since the particles for $\mathbf{s}_{k-M+1:k-1}$ are assumed known at time k in the recursive implementation, only Metropolis-Hastings candidates that are identical to these values are accepted which produces no diversity. The second goal is then to reintroduce statistical diversity on the quantities $\mathbf{s}_{k-M+1:k-1}$.

The BSIS formulation of the particle filter is developed next to achieve the two specified goals.

The development of the SIS particle in Section 2.4.3 makes use of following factorization property of the importance function:

$$q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) = q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k})q(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1}). \tag{5.3}$$

This assumption is used to prevent having to modify the particles for the past history $\mathbf{x}_{1:k-1}$ based on the new measurement $\mathbf{y}_k$. The benefit of this approximation is to reduce computational complexity since the particles for the state history $\mathbf{x}_{1:k-1}$ do not have to be redrawn.

The BSIS approach proposes to generate particles at time k for the block of B most recent states $\mathbf{x}_{k-B+1:k}$, denoted as $\mathbf{x}_{B,k}$. This offers a compromise to having to redraw particles for the entire state history $\mathbf{x}_{1:k-1}$. This is done at the expense of increased computational complexity in the algorithm. This leads us to consider an importance function with the following factorization property:

$$q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) = q(\mathbf{x}_{B,k}|\mathbf{x}_{1:k-B}, \mathbf{y}_{1:k})q(\mathbf{x}_{1:k-B}|\mathbf{y}_{1:k-B}). \tag{5.4}$$

This specifies that the block of states $\mathbf{x}_{B,k}$ are being jointly drawn from the importance distribution $q(\mathbf{x}_{B,k}|\mathbf{x}_{1:k-B}, \mathbf{y}_{1:k})$, and that the current measurement $\mathbf{y}_k$ is incorporated into the particle generation for past states $\mathbf{x}_{k-B+1:k-1}$. In addition, diversity can be introduced on these past states as well since $\mathbf{x}_{k-B+1:k-1}$ are not assumed known at

time k. By considering the state variable $\mathbf{s}_{M,k}$ with the use of a block size B of M states, it is seen that this formulation achieves the two desired goals specified in the previous section. Figure 5.1 compares the basic principles of the SIS and BSIS particle filters.

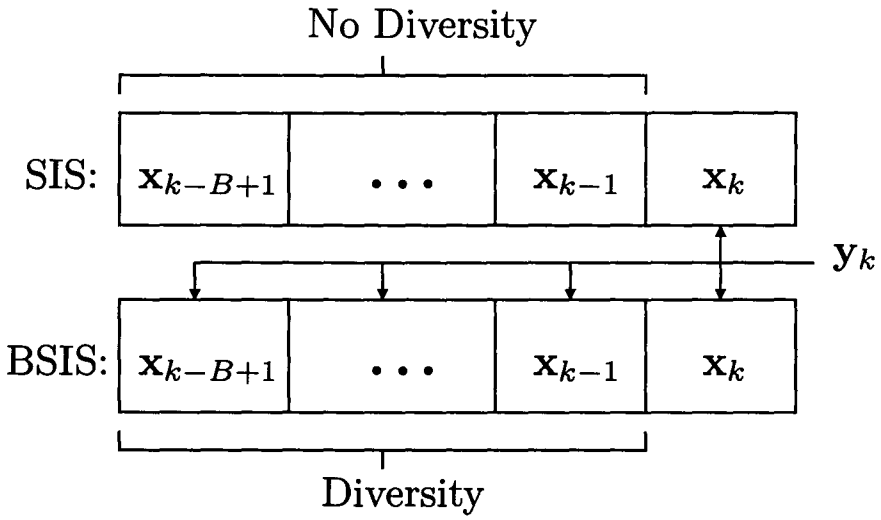


Figure 5.1: Comparison of the SIS and BSIS particle filters

The corresponding recursion for the posterior distribution is developed by applying Bayes' theorem to produce a likelihood for the block of measurements $\mathbf{y}_{B,k}$ and a prior for the block of states $\mathbf{x}_{B,k}$:

$$\begin{aligned} p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k}) &= \frac{p(\mathbf{y}_{B,k}|\mathbf{x}_{1:k},\mathbf{y}_{1:k-B})p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k-B})}{p(\mathbf{y}_{B,k}|\mathbf{y}_{1:k-B})} \\ &= \frac{p(\mathbf{y}_{B,k}|\mathbf{x}_{1:k},\mathbf{y}_{1:k-B})p(\mathbf{x}_{B,k}|\mathbf{x}_{1:k-B},\mathbf{y}_{1:k-B})}{p(\mathbf{y}_{k-B+1:k}|\mathbf{y}_{1:k-B})}p(\mathbf{x}_{1:k-B}|\mathbf{y}_{1:k-B}) \\ &= \frac{p(\mathbf{y}_{B,k}|\mathbf{x}_{1:k})p(\mathbf{x}_{B,k}|\mathbf{x}_{k-B})}{p(\mathbf{y}_{B,k}|\mathbf{y}_{1:k-B})}p(\mathbf{x}_{1:k-B}|\mathbf{y}_{1:k-B}). \end{aligned} \tag{5.5}$$

The recursive update for the importance weights using (5.4) and (5.5) then follows as

$$\begin{aligned} w_k^i &\propto \frac{p(\mathbf{y}_{B,k}|\mathbf{x}_{1:k}^i)p(\mathbf{x}_{B,k}^i|\mathbf{x}_{k-B}^i)}{q(\mathbf{x}_{B,k}^i|\mathbf{x}_{1:k-B}^i,\mathbf{y}_{1:k})}\frac{p(\mathbf{x}_{1:k-B}^i|\mathbf{y}_{1:k-B})}{q(\mathbf{x}_{1:k-B}^i|\mathbf{y}_{1:k-B})} \\ &\propto w_{k-B}^i\frac{p(\mathbf{y}_{B,k}|\mathbf{x}_{1:k}^i)p(\mathbf{x}_{B,k}^i|\mathbf{x}_{k-B}^i)}{q(\mathbf{x}_{B,k}^i|\mathbf{x}_{1:k-B}^i,\mathbf{y}_{1:k})}. \end{aligned} \tag{5.6}$$

The weight recursion implies that the previous weights $w_{k-B:k-1}$ need to be stored. Also, the importance function draws samples for $\mathbf{x}_{B,k}$ rather than just $\mathbf{x}_k$. These observations indicate that a reasonable rule of thumb for the increase in computational complexity for the BSIS approach is a factor of $O(B)$. The BSIS implementation is designed to reduce the number of particles required to achieve a desired level of performance by introducing additional measurement information and diversity into past states. In a convolution model, these past states directly influence the performance in estimating the current state. Therefore, the choice of a BSIS implementation is only beneficial if it can outperform a SIS implementation which is allowed to use B times as many particles. The choice of using SIS or BSIS would then be dependent on the specific application and the tradeoffs between computational cost and algorithm performance.

Future research to incorporate the BSIS formulation into the proposed marginalized particle filtering algorithm is required. It is also noted that the BSIS method may have application beyond the specific blind system identification problem presented in this thesis.

5.3.2 Dereverberation of Speech

For the problem of the dereverberation of speech, typical AIR can be in the order of 250ms, or 2000 samples at a sampling rate of 8kHz. The large computational cost of dealing with such a long channel can be addressed by decomposing the problem into smaller independent blind estimation problems using the complex subband decomposition [76]. In previous work [21], which was an extension of [92], the complex subband decomposition was used together with an MCMC algorithm that estimated the source. The use of the proposed marginalized particle filtering algorithm allows many of the previous assumptions to be relaxed, and to consider more realistic models for the speech sources and acoustic channel. Future work is required to implement the proposed blind estimation algorithm using the complex subband decomposition.

The use of the filter bank structure requires a prototype filter design that can accurately implement the complex subbanding decomposition. An efficient method for the design of oversampled near perfect reconstruction (NPR) generalized discrete Fourier transform (GDFT) filter banks is presented in [91]. This approach can be considered for the design of the prototype filter, in addition to previous methods described in [76].

It is then of interest to test the proposed blind estimation algorithm in addressing the practical problem of the dereverberation of speech. Real audio data has been collected in a reverberant room on campus in two experiments. In the first measurement setup, the acoustic impulse responses of the room were measured using a four-element array of omnidirectional microphones. The relative location of the speaker producing the test signal was varied, along with the spacing and heights of the array elements. In the second measurement setup shown in Figures 5.2,5.3, the testing procedure described in [90] was followed. The collected measurements were added to the R-HINT-E (Realistic Hearing In Noise Testing Environment) database. The audio signals were measured using KEMAR, a head-torso model which includes a three-element microphone array inserted in each ear. The data from these two experiments can be considered for testing the complex subbanding implementation of the proposed algorithm.

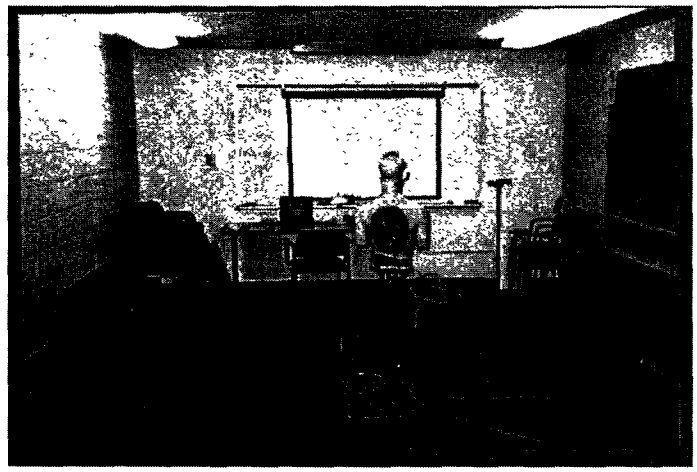


Figure 5.2: Experimental setup for measuring acoustic impulse responses

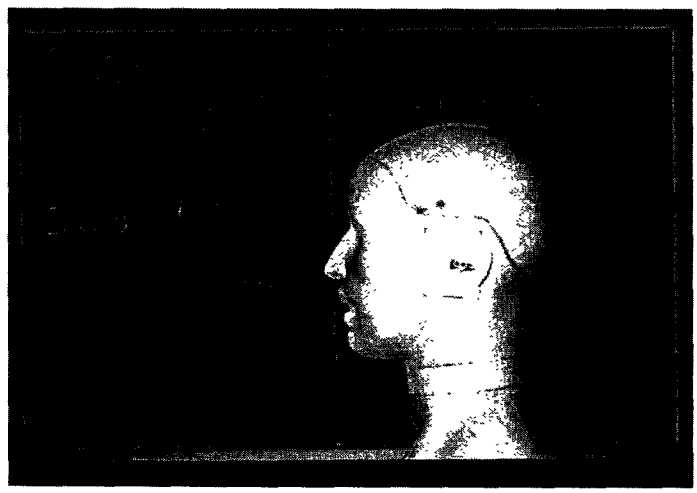


Figure 5.3: Three-element microphone array in KEMAR

Appendix A

Derivation of the Marginalized Prior and Likelihood

The marginalized likelihood $p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1})$ and prior $p(\mathbf{s}_k|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1})$ are derived for the SIS marginalized particle filtering algorithm in Section 3.3. The derivation follows the proof presented in [79], and only the key steps are shown here. In addition to deriving the marginalized distributions, an identity is derived for use in Appendix B to derive the Gaussian approximation to the optimal importance function.

The marginalized likelihood is derived from the measurement model by marginalizing over the channel $\mathbf{h}_k$:

$$p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1}) = \int p(\mathbf{y}_k|\mathbf{h}_k, \mathbf{s}_{L,k}) p(\mathbf{h}_k|\mathbf{s}_{1:k}, \mathbf{y}_{1:k-1}) d\mathbf{h}_k, \tag{A.1}$$

where from Section 3.2

$$\begin{aligned} p(\mathbf{y}_k|\mathbf{h}_k, \mathbf{s}_{L,k}) &= \mathcal{N}(\mathbf{T}_k\mathbf{h}_k, \mathbf{\Sigma}_w) \\ &= \frac{1}{2\pi^{\frac{J}{2}}|\mathbf{\Sigma}_w|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)\right], \end{aligned} \tag{A.2}$$

and

$$p(\mathbf{h}_k|\mathbf{s}_{1:k},\mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{h}}_{k|k-1},\mathbf{\Phi}_{h,k|k-1}) = \frac{1}{2\pi^{\frac{NJL}{2}}|\mathbf{\Phi}_{h,k|k-1}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{h}_k-\hat{\mathbf{h}}_{k|k-1})^{\mathrm{T}}\mathbf{\Phi}^{-1}_{h,k|k-1}(\mathbf{h}_k-\hat{\mathbf{h}}_{k|k-1})\right]. \tag{A.3}$$

The approach of the proof is to factorize the integrand in (A.1) into the product of a Gaussian distribution for $\mathbf{h}_k$ and a Gaussian distribution for $\mathbf{y}_k$ independent of $\mathbf{h}_k$. Marginalizing over the channel then integrates the Gaussian distribution for $\mathbf{h}_k$ to one, and the remaining Gaussian distribution is identified as $p(\mathbf{y}_k|\mathbf{s}_{L,k})$. Defining

$$\tilde{\mathbf{h}}_k = \mathbf{h}_k - \hat{\mathbf{h}}_{k|k-1}, \tag{A.4}$$

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k - \mathbf{T}_k\hat{\mathbf{h}}_{k|k-1}, \tag{A.5}$$

and grouping expressions results in

$$p(\mathbf{y}_k|\mathbf{s}_{M,1:k},\mathbf{y}_{1:k-1}) = \int \frac{1}{2\pi^{\frac{J+NJL}{2}}(|\mathbf{\Sigma}_w||\mathbf{\Phi}_{h,k|k-1}|)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left((\tilde{\mathbf{y}}_k-\mathbf{T}_k\tilde{\mathbf{h}}_k)^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}(\tilde{\mathbf{y}}_k-\mathbf{T}_k\tilde{\mathbf{h}}_k)+\tilde{\mathbf{h}}_k^{\mathrm{T}}\mathbf{\Phi}^{-1}_{h,k|k-1}\tilde{\mathbf{h}}_k\right)\right]d\mathbf{h}_k. \tag{A.6}$$

The exponent can be written as the quadratic form

$$\begin{bmatrix}\tilde{\mathbf{h}}_k\\ \tilde{\mathbf{y}}_k\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\mathbf{\Phi}^{-1}_{h,k|k-1}+\mathbf{T}_k^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}\mathbf{T}_k & -\mathbf{T}_k^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}\\ -\mathbf{\Sigma}_w^{-1}\mathbf{T}_k^{\mathrm{T}} & \mathbf{\Sigma}_w^{-1}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{h}}_k\\ \tilde{\mathbf{y}}_k\end{bmatrix}. \tag{A.7}$$

The matrix can be factored as

$$\begin{bmatrix}\mathbf{I} & -\mathbf{W}_k\\ \mathbf{0} & \mathbf{I}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\mathbf{\Phi}^{-1}_{h,k|k} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_k^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I} & -\mathbf{W}_k\\ \mathbf{0} & \mathbf{I}\end{bmatrix}, \tag{A.8}$$

where the following terms are defined

$$\mathbf{W}_k = (\mathbf{\Phi}^{-1}_{h,k|k-1}+\mathbf{T}_k^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}\mathbf{T}_k)\mathbf{T}_k^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}, \tag{A.9}$$

$$\mathbf{\Phi}^{-1}_{h,k|k} = \mathbf{\Phi}^{-1}_{h,k|k-1}+\mathbf{T}_k^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}\mathbf{T}_k, \tag{A.10}$$

$$\mathbf{Q}_k^{-1} = \mathbf{\Sigma}_w^{-1}-\mathbf{\Sigma}_w^{-1}\mathbf{T}_k(\mathbf{T}_k^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}\mathbf{T}_k+\mathbf{\Phi}^{-1}_{h,k|k-1})^{-1}\mathbf{T}_k^{\mathrm{T}}\mathbf{\Sigma}_w^{-1}. \tag{A.11}$$

The repeated use of the matrix inversion lemma given by

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}, \tag{A.12}$$

results in the expressions

$$\mathbf{W}_k = \mathbf{\Phi}_{h,k|k-1}\mathbf{T}_k^{\mathrm{T}}\mathbf{Q}_k^{-1}, \tag{A.13}$$

$$\mathbf{\Phi}_{h,k|k} = \mathbf{\Phi}_{h,k|k-1} - \mathbf{W}_k\mathbf{T}_k\mathbf{\Phi}_{h,k|k-1}, \tag{A.14}$$

$$\mathbf{Q}_k = \mathbf{T}_k\mathbf{\Phi}_{h,k|k-1}\mathbf{T}_k^{\mathrm{T}} + \mathbf{\Sigma}_w. \tag{A.15}$$

From the description of the Kalman filter in Section 2.3, it can be seen that these are expressions for the Kalman gain $\mathbf{W}_k$, updated Kalman covariance $\mathbf{\Phi}_{h,k|k}$, and innovation covariance $\mathbf{Q}_k$. Using properties of the determinant operator, it can also be shown that

$$|\mathbf{\Sigma}_w||\mathbf{\Phi}_{h,k|k-1}| = |\mathbf{\Phi}_{h,k|k}||\mathbf{Q}_k|. \tag{A.16}$$

Using this property and expanding the quadratic form for the exponent using the factorized matrix allows the integrand in (A.6) to be factorized as

$$\begin{aligned}
&p(\mathbf{y}_k|\mathbf{s}_{M,1:k},\mathbf{y}_{1:k-1}) \\
&= \int \frac{1}{2\pi^{\frac{NJL}{2}}|\mathbf{\Phi}_{h,k|k}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\tilde{\mathbf{h}}_k - \mathbf{W}_k\tilde{\mathbf{y}}_k)^{\mathrm{T}}\mathbf{\Phi}_{h,k|k}^{-1}(\tilde{\mathbf{h}}_k - \mathbf{W}_k\tilde{\mathbf{y}}_k)\right] \\
&\quad \frac{1}{(2\pi)^{\frac{J}{2}}|\mathbf{Q}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\tilde{\mathbf{y}}_k\mathbf{Q}_k^{-1}\tilde{\mathbf{y}}_k\right] d\mathbf{h}_k \\
&= \int \frac{1}{2\pi^{\frac{NJL}{2}}|\mathbf{\Phi}_{h,k|k}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{h}_k - \hat{\mathbf{h}}_{k|k-1} - \mathbf{W}_k\tilde{\mathbf{y}}_k)^{\mathrm{T}}\mathbf{\Phi}_{h,k|k}^{-1}(\mathbf{h}_k - \hat{\mathbf{h}}_{k|k-1} - \mathbf{W}_k\tilde{\mathbf{y}}_k)\right] \\
&\quad \frac{1}{(2\pi)^{\frac{J}{2}}|\mathbf{Q}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}_k - \mathbf{T}_k\hat{\mathbf{h}}_{k|k-1})^{\mathrm{T}}\mathbf{Q}_k^{-1}(\mathbf{y}_k - \mathbf{T}_k\hat{\mathbf{h}}_{k|k-1})\right] d\mathbf{h}_k.
\end{aligned} \tag{A.17}$$

By comparing the original integrand in (A.1) with the integrand in (A.17), it is noted that the following identity has been developed:

$$\begin{aligned}
&\mathcal{N}(\mathbf{y}_k;\mathbf{T}_k\mathbf{h}_k,\mathbf{\Sigma}_w)\mathcal{N}(\mathbf{h}_k;\hat{\mathbf{h}}_{k|k-1},\mathbf{\Phi}_{h,k|k-1}) \\
=&\mathcal{N}(\mathbf{y}_k;\mathbf{T}_k\hat{\mathbf{h}}_{k|k-1},\mathbf{T}_k\mathbf{\Phi}_{h,k|k-1}\mathbf{T}_k^{\mathrm{T}}+\mathbf{\Sigma}_w) \\
&\mathcal{N}(\mathbf{h}_k;\hat{\mathbf{h}}_{k|k-1}+\mathbf{W}_k(\mathbf{y}_k-\mathbf{T}_k\hat{\mathbf{h}}_{k|k-1}),\mathbf{\Phi}_{h,k|k-1}+\mathbf{W}_k\mathbf{T}_k\mathbf{\Phi}_{h,k|k-1}).
\end{aligned} \tag{A.18}$$

The random variable in each Gaussian distribution has been noted for clarity, along with the mean and covariance. This identity is also used in the development of the Gaussian approximation to the optimal importance function in Appendix B. By integrating over $\mathbf{h}_k$, the Gaussian distribution for $\mathbf{h}_k$ integrates to one, and the Gaussian distribution for $\mathbf{y}_k$ is constant with respect to $\mathbf{h}_k$ and therefore

$$p(\mathbf{y}_k|\mathbf{s}_{M,1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{T}_k\hat{\mathbf{h}}_{k|k-1}, \mathbf{T}_k\mathbf{\Phi}_{h,k|k-1}\mathbf{T}_k^{\mathrm{T}} + \mathbf{\Sigma}_w). \tag{A.19}$$

The same method can also be applied to determine the marginalized prior distribution, and the analogous result is

$$p(\mathbf{s}_k|\mathbf{s}_{M,1:k-1}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{S}_{k-1}\hat{\mathbf{a}}_{k|k-1}, \mathbf{S}_{k-1}\mathbf{\Phi}_{a,k|k-1}\mathbf{S}_{k-1}^{\mathrm{T}} + \mathbf{\Sigma}_v). \tag{A.20}$$

□

Appendix B

Gaussian Approximation to the Optimal Importance Function

The expression for the approximation to the optimal importance function in (3.78) specifies that it is proportional to the product of two Gaussian distributions:

$$q(\mathbf{s}_k|\mathbf{s}_{M,k-1}, \mathbf{y}_{1:k}) \propto \hat{p}(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}, \mathbf{s}_k)p(\mathbf{s}_k|\mathbf{s}_{P,k-1}), \tag{B.21}$$

where

$$\hat{p}(\mathbf{y}_k|\mathbf{s}_{L-1,k-1}, \mathbf{s}_k) = \mathcal{N}(\hat{\mathbf{H}}_{k|k-1,0}\mathbf{s}_k + \hat{\mathbf{y}}_{k|k-1}, \hat{\mathbf{Q}}_k), \tag{B.22}$$

$$p(\mathbf{s}_k|\mathbf{s}_{P,k-1}) = \mathcal{N}(\mathbf{S}_{k-1}\hat{\mathbf{a}}_{k|k-1}, \mathbf{R}_k). \tag{B.23}$$

A modification of the identity (A.18) developed in Appendix A is now used to rewrite the product of Gaussian distributions as

$$\begin{aligned}
&\mathcal{N}(\mathbf{y}_k; \hat{\mathbf{H}}_{k|k-1,0}\mathbf{s}_k + \hat{\mathbf{y}}_{k|k-1}, \hat{\mathbf{Q}}_k)\mathcal{N}(\mathbf{s}_k; \hat{\mathbf{s}}_{k|k-1}, \mathbf{R}_k) \\
=&\mathcal{N}(\mathbf{y}_k; \hat{\mathbf{H}}_{k|k-1,0}\hat{\mathbf{s}}_{k|k-1} + \hat{\mathbf{y}}_{k|k-1}, \hat{\mathbf{H}}_{k|k-1,0}\mathbf{R}_k\hat{\mathbf{H}}^{\mathrm{T}}_{k|k-1,0} + \hat{\mathbf{Q}}_k) \\
&\mathcal{N}(\mathbf{s}_k; \hat{\mathbf{s}}_{k|k-1} + \mathbf{W}_k(\mathbf{y}_k - \hat{\mathbf{H}}_{k|k-1,0}\hat{\mathbf{s}}_{k|k-1} - \hat{\mathbf{y}}_{k|k-1}), \mathbf{R}_k - \mathbf{W}_k\hat{\mathbf{H}}_{k|k-1,0}\mathbf{R}_k),
\end{aligned} \tag{B.24}$$

where now the Kalman gain is

$$\mathbf{W}_k = \mathbf{R}_k\hat{\mathbf{H}}^{\mathrm{T}}_{k|k-1,0}[\hat{\mathbf{H}}_{k|k-1,0}\mathbf{R}_k\hat{\mathbf{H}}^{\mathrm{T}}_{k|k-1,0} + \hat{\mathbf{Q}}_k]^{-1}. \tag{B.25}$$

The presence of the term $\hat{\mathbf{y}}_{k|k-1}$ is addressed in the proof of the identity in Appendix A by defining

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1} - \hat{\mathbf{H}}_{k|k-1,0}\mathbf{s}_k. \tag{B.26}$$

This does not change other aspects of the proof and results in the modification of the identity used in this section.

The benefit of using this identity is that now the approximation to the optimal importance function is the product of a Gaussian distribution for the variable of interest $\mathbf{s}_k$, and a Gaussian distribution for $\mathbf{y}_k$ which is independent of $\mathbf{s}_k$, and is therefore only a scaling term. The importance function can now be identified as drawing samples from

$$\begin{aligned} &q(\mathbf{s}_k|\mathbf{s}_{M,k-1}, \mathbf{y}_{1:k}) = \\ &\mathcal{N}(\hat{\mathbf{s}}_{k|k-1} + \mathbf{W}_k(\mathbf{y}_k - \hat{\mathbf{H}}_{k|k-1,0}\hat{\mathbf{s}}_{k|k-1} - \hat{\mathbf{y}}_{k|k-1}), \mathbf{R}_k - \mathbf{W}_k\hat{\mathbf{H}}_{k|k-1,0}\mathbf{R}_k). \end{aligned} \tag{B.27}$$

□

Appendix C

Derivation of the PCRB

The recursion for the PCRB matrix $\mathbf{P}_{k+1}$ shown in (2.54) is reproduced here:

$$\mathbf{P}_{k+1} = \mathbf{F}_k(\mathbf{P}_k^{-1} + \mathbf{R}_k^{-1})^{-1}\mathbf{F}_k^{\mathrm{T}} + \mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^{\mathrm{T}}. \tag{C.28}$$

The matrix expressions $\mathbf{F}_k$, $\mathbf{R}_k^{-1}$, $\mathbf{G}_k$, $\mathbf{Q}_k$ and $\mathbf{P}_0$ defined in Section 2.5 are now developed using the state space formulation for the state variable $\boldsymbol{\phi}_k$ and noise variable $\mathbf{u}_k$ in Section 4.1.· All expectations are evaluated numerically by averaging Monte Carlo realizations of the state space model.

Derivation of $\mathbf{G}_k$:

$$\mathbf{G}_k^{\mathrm{T}} = \mathbb{E}\left\{\nabla_{\mathbf{u}_k}\mathbf{f}_k^{\mathrm{T}}(\boldsymbol{\phi}_k, \mathbf{u}_k)\right\} \tag{C.29}$$

Using the definition of $\mathbf{f}_k(\boldsymbol{\phi}_k, \mathbf{u}_k)$ in (4.4) results in

$$\begin{aligned}\mathbf{G}_k^{\mathrm{T}} &= \mathbb{E}\left\{\nabla_{\mathbf{u}_k}(\tilde{\mathbf{F}}_k\boldsymbol{\phi}_k + \mathbf{u}_k)^{\mathrm{T}}\right\} \\ &= \mathbf{I}_r.\end{aligned} \tag{C.30}$$

Derivation of $\mathbf{F}_k$:

$$\begin{aligned}\mathbf{F}_k^{\mathrm{T}} &= \mathbb{E}\left\{\nabla_{\boldsymbol{\phi}_k}\mathbf{f}_k^{\mathrm{T}}(\boldsymbol{\phi}_k, \mathbf{u}_k)\right\} \\ &= \mathbb{E}\left\{\begin{bmatrix} \mathbf{a}_h\mathbf{I}_{NJL} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{A}}_{k+1}^{\mathrm{T}} \end{bmatrix}\right\}\end{aligned} \tag{C.31}$$

Derivation of $\mathbf{Q}_k$:

$$\mathbf{Q}_k = \left(\mathbb{E}\left\{-\Delta_{\mathbf{u}_k}^{\mathbf{u}_k} \log p(\mathbf{u}_k)\right\}\right)^{-1} \tag{C.32}$$

The noise vector $\mathbf{u}_k$ in (4.4) is defined in terms of the noise vector $\mathbf{v}_{h,k}$ from Section 3.1.1, and $\tilde{\mathbf{v}}_k$ from Section 3.3. The distributions for these noise vectors are summarized here:

$$\mathbf{v}_{h,k} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_h), \tag{C.33}$$

$$\tilde{\mathbf{v}}_k \sim \mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}_v), \tag{C.34}$$

where from the definition of $\tilde{\mathbf{v}}_k$ in (3.54):

$$\tilde{\boldsymbol{\Sigma}}_v = \begin{bmatrix} \mathbf{0}_{(M-1)N,(M-1)N} & \mathbf{0}_{(M-1)N,N} \\ \mathbf{0}_{N,(M-1)N} & \boldsymbol{\Sigma}_v \end{bmatrix}. \tag{C.35}$$

Assuming the noise vectors $\mathbf{v}_{h,k}$ and $\tilde{\mathbf{v}}_k$ are mutually independent, $\mathbf{u}_k$ is then Gaussian distributed with mean zero and covariance given by

$$\mathbf{Q}_{u,k} = \begin{bmatrix} \boldsymbol{\Sigma}_h & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Sigma}}_v \end{bmatrix}. \tag{C.36}$$

This results in

$$\begin{aligned} \mathbf{Q}_k &= \left(\mathbb{E}\left\{\Delta_{\mathbf{u}_k}^{\mathbf{u}_k} \frac{1}{2} \mathbf{u}_k^{\mathrm{T}} \mathbf{Q}_{u,k}^{-1} \mathbf{u}_k\right\}\right)^{-1} \\ &= \mathbf{Q}_{u,k}. \end{aligned} \tag{C.37}$$

From (C.35), the diagonal matrix $\tilde{\boldsymbol{\Sigma}}_v$ contains zero elements on the main diagonal. Therefore, the diagonal matrix $\mathbf{Q}_{u,k}$ is not invertible. However, since only the term $\mathbf{Q}_k$ appears in (C.28), the PCRB recursion handles the case of a singular transition prior.

Derivation of $\mathbf{R}_k^{-1}$:

$$\mathbf{R}_k^{-1} = \mathbb{E}\left\{-\Delta_{\boldsymbol{\phi}_k}^{\boldsymbol{\phi}_k} \log p(\mathbf{y}_k|\boldsymbol{\phi}_k)\right\} \tag{C.38}$$

Based on the structure of the state variable $\boldsymbol{\phi}_k = [\mathbf{h}_k^{\mathrm{T}}, \mathbf{s}_{M,k}^{\mathrm{T}}]^{\mathrm{T}}$, the matrix $\mathbf{R}_k^{-1}$ is partitioned into submatrices as follows using the notation $p_k = p(\mathbf{y}_k|\boldsymbol{\phi}_k)$:

$$\begin{aligned}\mathbf{R}_k^{-1} &= \begin{bmatrix} \mathbb{E}\left\{-\Delta_{\mathbf{h}_k}^{\mathbf{h}_k} \log p_k\right\} & \mathbb{E}\left\{-\Delta_{\mathbf{h}_k}^{\mathbf{s}_{M,k}} \log p_k\right\} \\ \mathbb{E}\left\{-\Delta_{\mathbf{s}_{M,k}}^{\mathbf{h}_k} \log p_k\right\} & \mathbb{E}\left\{-\Delta_{\mathbf{s}_{M,k}}^{\mathbf{s}_{M,k}} \log p_k\right\} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \mathbf{R}_{k,hh}^{-1} & \mathbf{R}_{k,hs}^{-1} \\ \mathbf{R}_{k,sh}^{-1} & \mathbf{R}_{k,ss}^{-1} \end{bmatrix}. \end{aligned} \tag{C.39}$$

The submatrices in the partitioned $\mathbf{R}_k^{-1}$ are now derived individually. We make use of the two equivalent representations of the measurement equation in Section 4.1. The first form in (4.5) is convenient for differentiating with respect to $\mathbf{s}_{M,k}$, and the second in (4.6) for differentiating with respect to $\mathbf{h}_k$. The submatrices are now derived:

$$\begin{aligned}\mathbf{R}_{k,hh}^{-1} &= \mathbb{E}\left\{\Delta_{\mathbf{h}_k}^{\mathbf{h}_k} \frac{1}{2}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)\right\} \\ &= \mathbb{E}\left\{-\nabla_{\mathbf{h}_k}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k\right\} \\ &= \mathbb{E}\left\{\mathbf{T}_k^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k\right\}, \end{aligned} \tag{C.40}$$

$$\begin{aligned}\mathbf{R}_{k,ss}^{-1} &= \mathbb{E}\left\{\Delta_{\mathbf{s}_{M,k}}^{\mathbf{s}_{M,k}} \frac{1}{2}(\mathbf{y}_k - \tilde{\mathbf{H}}_k\mathbf{s}_{M,k})^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_k - \tilde{\mathbf{H}}_k\mathbf{s}_{M,k})\right\} \\ &= \mathbb{E}\left\{-\nabla_{\mathbf{s}_{M,k}}(\mathbf{y}_k - \tilde{\mathbf{H}}_k\mathbf{s}_{M,k})^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\tilde{\mathbf{H}}_k\right\} \\ &= \mathbb{E}\left\{\tilde{\mathbf{H}}_k^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\tilde{\mathbf{H}}_k\right\}. \end{aligned} \tag{C.41}$$

Before continuing with the derivation of $\mathbf{R}_{k,sh}^{-1}$, we simplify the presentation by observing that since the likelihood is only dependent on $\mathbf{s}_{L,k}$, the derivative with respect to the previous quantities in $\mathbf{s}_{M,k}$ is zero. This is reflected in:

$$\begin{aligned}\mathbf{R}_{k,sh}^{-1} &= \mathbb{E}\left\{\Delta_{\mathbf{s}_{M,k}}^{\mathbf{h}_k} \frac{1}{2}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)\right\} \\ &= \begin{bmatrix} \mathbb{E}\left\{\Delta_{\mathbf{s}_{M,k}[1:(M-L)N]}^{\mathbf{h}_k} \frac{1}{2}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)\right\} \\ \mathbb{E}\left\{\Delta_{\mathbf{s}_{L,k}}^{\mathbf{h}_k} \frac{1}{2}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_k - \mathbf{T}_k\mathbf{h}_k)\right\} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0}_{(M-L)N,NL} \\ \tilde{\mathbf{R}}_{k,sh}^{-1} \end{bmatrix}. \end{aligned}$$

The matrix $\tilde{\mathbf{R}}_{k,sh}^{-1}$ is now derived:

$$\begin{aligned}\mathbf{R}_{k,sh}^{-1} &= \mathbb{E}\left\{\Delta_{\mathbf{s}_{L,k}}^{\mathbf{h}_k}\frac{1}{2}(\mathbf{y}_k-\mathbf{T}_k\mathbf{h}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_k-\mathbf{T}_k\mathbf{h}_k)\right\}\\ &= \mathbb{E}\left\{-\nabla_{\mathbf{s}_{L,k}}(\mathbf{y}_k-\tilde{\mathbf{H}}_k\mathbf{s}_{M,k})^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k\right\}\\ &= \mathbb{E}\left\{-\nabla_{\mathbf{s}_{L,k}}(\mathbf{y}_k^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k-(\tilde{\mathbf{H}}_k\mathbf{s}_{M,k})^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k)\right\}. \qquad \text{(C.42)}\end{aligned}$$

To complete the derivation of $\mathbf{R}_{k,sh}^{-1}$, we first examine the term $\mathbf{y}_k^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k \in \mathbb{R}^{1\times JNL}$ using the definition of $\mathbf{T}_k$ in (3.25):

$$\mathbf{y}_k^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k = \left[\frac{\mathbf{y}_k[1]}{\sigma_{w,1}^2}\mathbf{s}_{L,k}^{\mathrm{T}}, \frac{\mathbf{y}_k[2]}{\sigma_{w,2}^2}\mathbf{s}_{L,k}^{\mathrm{T}}, \ldots, \frac{\mathbf{y}_k[J]}{\sigma_{w,J}^2}\mathbf{s}_{L,k}^{\mathrm{T}}\right]. \qquad \text{(C.43)}$$

Differentiating with respect to $\mathbf{s}_{L,k}$ results in:

$$\begin{aligned}\nabla_{\mathbf{s}_{L,k}}(\mathbf{y}_k^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k) &= \left[\frac{\mathbf{y}_k[1]}{\sigma_{w,1}^2}\mathbf{I}_{NL}, \frac{\mathbf{y}_k[2]}{\sigma_{w,2}^2}\mathbf{I}_{NL}, \ldots, \frac{\mathbf{y}_k[J]}{\sigma_{w,J}^2}\mathbf{I}_{NL}\right] \qquad \text{(C.44)}\\ &= (\mathbf{y}_k^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1})\otimes\mathbf{I}_{NL}.\end{aligned}$$

We now examine the remaining term $(\tilde{\mathbf{H}}_k\mathbf{s}_{M,k})^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k \in \mathbb{R}^{1\times JNL}$ in (C.42) using the definitions of $\tilde{\mathbf{H}}_k$ in (3.56) and $\mathbf{H}_k$ in (3.23),(3.24):

$$\begin{aligned}(\tilde{\mathbf{H}}_k\mathbf{s}_{M,k})^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k &= \left[\frac{\mathbf{h}_1^{\mathrm{T}}\mathbf{s}_{L,k}}{\sigma_{w,1}^2}\mathbf{s}_{L,k}^{\mathrm{T}}, \frac{\mathbf{h}_2^{\mathrm{T}}\mathbf{s}_{L,k}}{\sigma_{w,2}^2}\mathbf{s}_{L,k}^{\mathrm{T}}, \ldots, \frac{\mathbf{h}_J^{\mathrm{T}}\mathbf{s}_{L,k}}{\sigma_{w,J}^2}\mathbf{s}_{L,k}^{\mathrm{T}}\right] \qquad \text{(C.45)}\\ &= [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_J],\end{aligned}$$

where the following definition has been introduced

$$\boldsymbol{\alpha}_j = \frac{\mathbf{h}_j^{\mathrm{T}}\mathbf{s}_{L,k}}{\sigma_{w,j}^2}\mathbf{s}_{L,k}^{\mathrm{T}} \in \mathbb{R}^{1\times NL}. \qquad \text{(C.46)}$$

For the vectors $\mathbf{h}_j$ and $\mathbf{s}_{L,k} \in \mathbb{R}^{NL\times 1}$:

$$\mathbf{h}_j^{\mathrm{T}}\mathbf{s}_{L,k} = \sum_{i=1}^{NL}\mathbf{h}_j[i]\mathbf{s}_{L,k}[i], \qquad \text{(C.47)}$$

which allows the individual elements of $\boldsymbol{\alpha}_j$ to written as

$$\boldsymbol{\alpha}_j[\ell] = \frac{\sum_{i=1}^{NL}\mathbf{h}_j[i]\mathbf{s}_{L,k}[i]}{\sigma_{w,j}^2}\mathbf{s}_{L,k}[\ell], \quad \ell = 1,2,\ldots,NL. \qquad \text{(C.48)}$$

Differentiating $\boldsymbol{\alpha}_j[\ell]$ with respect to $\mathbf{s}_{L,k}[m]$, $m = 1, 2, \ldots, NL$ results in

$$\frac{\partial \boldsymbol{\alpha}_j[\ell]}{\partial \mathbf{s}_{L,k}[m]} = \begin{cases} \frac{\mathbf{h}_j[\ell]\mathbf{s}_{L,k}[\ell]}{\sigma^2_{w,j}} + \frac{\mathbf{h}_j^{\mathrm{T}}\mathbf{s}_{L,k}}{\sigma^2_{w,j}}, & \ell = m \\ \frac{\mathbf{h}_j[m]\mathbf{s}_{L,k}[\ell]}{\sigma^2_{w,j}}, & \ell \neq m \end{cases} \tag{C.49}$$

This can be written compactly in matrix format as

$$\nabla_{\mathbf{s}_{L,k}}(\boldsymbol{\alpha}_j) = \frac{\mathbf{h}_j^{\mathrm{T}}\mathbf{s}_{L,k}}{\sigma^2_{w,j}}\mathbf{I}_{NL} + \frac{\mathbf{s}_{L,k}\mathbf{h}_j^{\mathrm{T}}}{\sigma^2_{w,j}}. \tag{C.50}$$

It is convenient to define the noise-free measurement vector $\tilde{\mathbf{y}}_k \in \mathbb{R}^{J\times 1}$ as

$$\tilde{\mathbf{y}}_k[j] = \mathbf{h}_j^{\mathrm{T}}\mathbf{s}_{L,k}, \; j = 1, 2, \ldots, J. \tag{C.51}$$

Now differentiating the expression in (C.45) results in

$$\begin{aligned} \nabla_{\mathbf{s}_{L,k}}(\tilde{\mathbf{H}}_k\mathbf{s}_{M,k})^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}_k &= \left[\nabla_{\mathbf{s}_{L,k}}(\boldsymbol{\alpha}_1), \ldots, \nabla_{\mathbf{s}_{L,k}}(\boldsymbol{\alpha}_J)\right] \\ &= \left[\frac{\tilde{\mathbf{y}}_1}{\sigma^2_{w,1}}\mathbf{I}_{NL} + \frac{\mathbf{s}_{L,k}\mathbf{h}_1^{\mathrm{T}}}{\sigma^2_{w,1}}, \ldots, \frac{\tilde{\mathbf{y}}_J}{\sigma^2_{w,J}}\mathbf{I}_{NL} + \frac{\mathbf{s}_{L,k}\mathbf{h}_J^{\mathrm{T}}}{\sigma^2_{w,J}}\right] \\ &= (\tilde{\mathbf{y}}^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1}) \otimes \mathbf{I}_{NL} + \mathbf{s}_{L,k}\mathbf{h}_k^{\mathrm{T}}(\boldsymbol{\Sigma}_w^{-1} \otimes \mathbf{I}_{NL}). \end{aligned} \tag{C.52}$$

Using both (C.44) and (C.52) in the expression for $\tilde{\mathbf{R}}^{-1}_{k,sh}$ in (C.42) produces

$$\mathbf{R}^{-1}_{k,sh} = \mathbb{E}\left\{(\tilde{\mathbf{y}}_k - \mathbf{y}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1} \otimes \mathbf{I}_{NL} + \mathbf{s}_{L,k}\mathbf{h}_k^{\mathrm{T}}(\boldsymbol{\Sigma}_w^{-1} \otimes \mathbf{I}_{NL})\right\}. \tag{C.53}$$

The difference between the noise-free measurement $\tilde{\mathbf{y}}_k$ and the true measurement $\mathbf{y}_k$ is equal to the measurement noise vector $\mathbf{w}_k$. Since the measurement noise is assumed zero-mean, the following holds:

$$\begin{aligned} \mathbb{E}\left\{(\tilde{\mathbf{y}}_k - \mathbf{y}_k)^{\mathrm{T}}\right\} &= \mathbb{E}\left\{\mathbf{w}_k^{\mathrm{T}}\right\} \\ &= \mathbf{0}_{J,1}, \end{aligned} \tag{C.54}$$

and therefore the final expression for $\tilde{\mathbf{R}}^{-1}_{k,sh}$ is

$$\tilde{\mathbf{R}}^{-1}_{k,sh} = \mathbb{E}\left\{\mathbf{s}_{L,k}\mathbf{h}_k^{\mathrm{T}}(\boldsymbol{\Sigma}_w^{-1} \otimes \mathbf{I}_{NL})\right\}. \tag{C.55}$$

The derivation of the submatrices forming $\mathbf{R}_k^{-1}$ is completed by noting that $\mathbf{R}^{-1}_{k,hs} = (\mathbf{R}^{-1}_{k,sh})^{\mathrm{T}}$.

Derivation of $\mathbf{P}_0$:

$$\mathbf{P}_0 = \left(\mathbb{E}\{-\Delta_{\boldsymbol{\phi}_0}^{\boldsymbol{\phi}_0} \log p(\boldsymbol{\phi}_0)\}\right)^{-1} \tag{C.56}$$

The initial state $\boldsymbol{\phi}_0$ is assumed to be Gaussian distributed with mean zero and covariance given by

$$\mathbf{Q}_{0,k} = \begin{bmatrix} \boldsymbol{\Sigma}_{h,0} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Sigma}}_{v,0} \end{bmatrix}. \tag{C.57}$$

This results in

$$\begin{aligned} \mathbf{P}_0 &= \left(\mathbb{E}\left\{\Delta_{\boldsymbol{\phi}_0}^{\boldsymbol{\phi}_0} \frac{1}{2} \boldsymbol{\phi}_0^{\mathrm{T}} \mathbf{Q}_{0,k}^{-1} \boldsymbol{\phi}_0\right\}\right)^{-1} \\ &= \mathbf{Q}_{0,k}. \end{aligned} \tag{C.58}$$

□

Bibliography

[1] K. Abed-Meraim, J. Cardoso, A. Gorokhov, P. Loubaton, and E. Moulines, "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 42–55, Jan. 1997.

[2] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.

[3] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, pp. 1310–1322, Aug. 1997.

[4] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, New Jersey: Prentice Hall, 1979.

[5] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Processing*, vol. 47, pp. 2667–2676, Oct. 1999.

[6] C. Andrieu and S. J. Godsill, "A particle filter for model based audio source separation," in *Proc. International Workshop on Independent Component Analysis and Blind Signal Separation*, Helsinki, Finland, June 2000, pp. 381–386.

[7] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, pp. 174–188, Feb. 2002.

[8] A. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, Aug. 1995.

[9] N. Bergman, "Recursive Bayesian estimation: Navigation and tracking applications," Ph.D. thesis, Linkoping University, Linkoping, Sweden, 1999.

[10] J. Bernardo and A. Smith, *Bayesian Theory.* New York: John Wiley & Sons, 1994.

[11] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control.* San Francisco, California: Holden-Day, 1976.

[12] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, vol. 86, pp. 117–128, Jan. 2000.

[13] M. Bruno and A. Pavlov, "Improved particle filters for ballistic target tracking," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 2, Montreal, Canada, May 2004, pp. 705–708.

[14] J. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 8, pp. 2009–2025, Oct. 1998.

[15] G. Casella and C. P. Robert, "Rao-Blackwellisation of sampling schemes," *Biometrika*, vol. 83, pp. 81–94, Mar. 1996.

[16] R. Chen and T. H. Li, "Blind restoration of linearly degraded discrete signals by Gibbs sampling," *IEEE Trans. Signal Processing*, vol. 43, pp. 2410–2413, Oct. 1995.

[17] R. Chen and J. S. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society B.*, vol. 62, pp. 493–508, Sept. 2000.

[18] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," Adaptive Syst. Lab., McMaster University, Hamilton, ON, Canada, Tech. Rep., 2003.

[19] S.-S. Chin and S. Hong, "VLSI design of high-throughput processing element for real-time particle filtering," in *International Symposium on Signals, Circuits, and Systems*, Iasi, Romania, July 2003, pp. 617 – 620.

[20] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, Apr. 1994.

[21] M. J. Daly and J. P. Reilly, "Blind deconvolution using Bayesian methods with application to the dereverberation of speech," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 2, Montreal, Canada, May 2004, pp. 1009–1012.

[22] N. de Freitas, "Rao-Blackwellised particle filtering for fault diagnosis," *IEEE Trans. Aerosp.*, vol. 4, pp. 1767–1772, Mar. 2002.

[23] P. M. Djurić, "Sequential estimation of random parameters under model uncertainty," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 1, Instanbul, Turkey, June 2000, pp. 297–300.

[24] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering," *IEEE Signal Processing Mag.*, vol. 20, pp. 19–38, Sept. 2003.

[25] A. Doucet, "On sequential simultion-based methods for Bayesian filtering," University of Cambridge, Department of Engineering, Signal Processing Group, England, Tech. Rep. TR.310, 1998.

[26] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo Methods in Practice.* New York: Springer-Verlag, 2000.

[27] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proc. Sixteenth Conference on Uncertainty in Artificial Intelligence*, Stanford, California, July 2000, pp. 176–183.

[28] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, Dec. 2000.

[29] A. Doucet, N. J. Gordon, and V. Krishnamurty, "Particle filters for state estimation of jump Markov linear systems," *IEEE Trans. Signal Processing*, vol. 49, pp. 613–624, Mar. 2001.

[30] W. Fong, S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing with application to speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, pp. 438–449, Feb. 2002.

[31] T. Ghirmai, M. F. Bugallo, and P. M. Djurić, "Joint data detection and symbol timing and symbol timing estimation using particle filtering," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 4, Hong Kong, June 2003, pp. 596–599.

[32] G. Giannakis and J. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 360–377, Mar. 1989.

[33] W. Gilks, S. Richardson, and D. Spiegelhalterm, *Markov Chain Monte Carlo in Practice.* New York: Chapman and Hall, 1998.

[34] S. J. Godsill and C. Andrieu, "Bayesian separation and recovery of convolutively mixed autoregressive sources," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, Phoenix, Arizona, Mar. 1999.

[35] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing for non-linear time series," *Journal of American Statistical Association*, vol. 50, pp. 438–449, Mar. 2004.

[36] N. Gordon, D. Salmond, and A. Smith, "Novel approach to non-linear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, pp. 107–113, Apr. 1993.

[37] A. Gorokhov and P. Loubaton, "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. Circuits Syst. I*, vol. 44, pp. 813–820, Sept. 1997.

[38] P. Green, "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, Jan. 1995.

[39] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, New Jersey: Prentice-Hall, 2002.

[40] S. Haykin, K. Huber, and Z. Chen, "Bayesian sequential state estimation for MIMO wireless communications," *Proceedings of the IEEE*, vol. 92, pp. 439–454, Mar. 2004.

[41] S. M. Herman, "A particle filtering approach to joint passive radar tracking and target classification," Ph.D. thesis, University of Illinois, Urbana, Illinois, 2002.

[42] J. R. Hopgood, "Nonstationary signal processing with application to reverberation cancellation in acoustic environments," Ph.D. thesis, University of Cambridge, Cambridge, UK, 2000.

[43] ——, "Bayesian blind MIMO deconvolution of nonstationary autoregressive sources mixed through all-pole channels," in *IEEE Workshop on Statistical Signal Processing*, St. Louis, Missouri, Oct. 2003, pp. 422–425.

[44] J. R. Hopgood and P. J. W. Rayner, "Blind single channel deconvolution using nonstationary signal processing," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 2926–2937, May 2003.

[45] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.

[46] Y. Hua, S. An, and Y. Xiang, "Blind identification and equalization of FIR MIMO channels by BIDS," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 4, Pittsburg, Pennsylvania, May 2001, pp. 3124–3127.

[47] ——, "Blind identification of FIR MIMO channels by decorrelating subchannels," *IEEE Trans. Signal Processing*, vol. 51, pp. 1143–1155, May 2003.

[48] Y. Hua and J. Tugnait, "Blind identifiability of FIR-MIMO systems with colored input using second order statistics," *IEEE Signal Processing Lett.*, vol. 7, pp. 348–350, Dec. 2000.

[49] R. Iltis, "A sequential Monte Carlo filter for joint linear/nonlinear state estimation with application to DS-CDMA," *IEEE Trans. Signal Processing*, vol. 51, pp. 417–426, Feb. 2003.

[50] T. Kailath, *Linear Systems.* Englewood Cliffs, New Jersey: Prentice-Hall, 1980.

[51] R. Karlsson, T. Schon, and F. Gustafsson, "Complexity analysis of the marginalized particle filter," Department of Electrical Engineering, Linkoping University, SE-581 83 Linkoping, Sweden, Tech. Rep. LiTH-ISY-R-2611, June 2004.

[52] S. M. Kay, *Modern Spectral Estimation.* Englewood Cliffs, New Jersey: Prentice-Hall, 1988.

[53] D. Kirkland, "Modelling of acoustic impulse responses for deconvolution," M.A.Sc. thesis, Dept. of Elec. and Comp. Eng., McMaster University, Hamilton, ON, Canada, 1996.

[54] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, pp. 1–25, Mar. 1996.

[55] C. Kwok, D. Fox, and M. Meilă, "Real-time particle filters," *Proceedings of the IEEE*, vol. 92, pp. 469–484, Mar. 2004.

[56] J.-R. Larocque, J. P. Reilly, and W. Ng, "Particle filters for tracking an unknown number of sources," *IEEE Trans. Signal Processing*, vol. 50, pp. 2926–2937, Dec. 2001.

[57] T. H. Li, "Blind identification and deconvolution of linear systems driven by binary random sequences," *IEEE Trans. Inform. Theory*, vol. 38, pp. 26–38, Jan. 1992.

[58] X. R. Li, Y. Bar-Shalom, and T. Kirubarajan, *Estimation, Tracking and Navigation: Theory, Algorithms and Software.* New York: John Wiley & Sons, 2001.

[59] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 95, pp. 2714–2723, Nov. 1995.

[60] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputations," *Journal of the American Statistical Association*, vol. 90, pp. 567–576, Sept. 1995.

[61] ——, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, pp. 1032–1044, Sept. 1998.

[62] S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline, "Semi-blind deconvolution of neural impulse response in fMRI using a Gibbs sampling method," in *Proc.*

IEEE International Conference on Acoustic, Speech and Signal Processing, vol. 5, Montreal, Canada, May 2004, pp. 601–604.

[63] J. H. Manton, private communication, The University of Melbourne, Melbourne, Australia, 2004.

[64] J. Míguez and P. M. Djurić, "Blind equalization by sequential importance sampling," in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 1, Phoenix, Arizona, June 2000, pp. 845–848.

[65] M. Morelande, S. Challa, and N. Gordon, "A study of the application of particle filters to single target tracking problems," in *Proceedings of the SPIE*, vol. 5204, San Diego, California, Dec. 2003.

[66] W. Ng, J. P. Reilly, and T. Kiruburajan, "A Bayesian approach to tracking wideband targets using sensor arrays and particle filters," in *IEEE Workshop on Statistical Signal Processing*, St. Louis, Missouri, Oct. 2003, pp. 510–513.

[67] A. Papoulis, *Probability, Random Variables and Stochastic Processes.* New York: McGraw-Hill, 1984.

[68] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 320–327, May 2000.

[69] D. T. Pham and J. Cardoso, "Blind source separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 49, pp. 1837–1848, Sept. 2001.

[70] D. T. Pham, C. Servière, and H. Boumaraf, "Blind separation of convolutive audio mixtures using nonstationarity," in *Proc. International Workshop on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, Apr. 2003, pp. 975–980.

[71] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, pp. 590–599, June 1999.

[72] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Particle filtering for demodulation in fading channels with non-Gaussian additive noise," *IEEE Trans. Commun.*, vol. 49, pp. 579–582, Apr. 2001.

[73] K. Rahbar, "Multichannel blind estimation techniques: Blind system identification and blind source separation," Ph.D. thesis, Dept. of Elec. and Comp. Eng., McMaster University, Hamilton, ON, Canada, 2002.

[74] K. Rahbar and J. P. Reilly, "A new frequency domain method for blind source separation of convolutive audio mixtures," *to appear IEEE Transactions on Speech and Audio Processing.*

[75] K. Rahbar, J. P. Reilly, and J. H. Manton, "Blind identification of MIMO FIR systems driven by quasistationary sources using second-order statistics: A frequency domain approach," *IEEE Trans. Signal Processing*, vol. 52, pp. 406–417, Feb. 2004.

[76] J. P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," *IEEE Trans. Signal Processing*, vol. 50, pp. 2730–2743, Nov. 2002.

[77] C. Robert and G. Casella, *Monte Carlo Statistical Methods.* New York: Springer-Verlag, 1999.

[78] L. L. Scharf, *Detection, Estimation and Time Series Analysis.* Reading, Massachusetts: Addison-Wesley, 1991.

[79] T. Schon, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for nonlinear state-space models," Department of Electrical Engineering, Linkoping

[88] ——, "Particle methods for Bayesian modelling and enhancement of speech signals," *IEEE Trans. Speech Audio Processing*, vol. 10, pp. 173–185, Mar. 2002.

[89] M. Šimandl, J. Královec, and P. Tichavský, "Filtering, predictive, and smoothing Cramér-Rao bounds for discrete-time nonlinear dynamic systems," *Automatica*, vol. 37, pp. 1703–1716, Nov. 2001.

[90] K. Wiklund, R. Sonnadara, L. J. Trainor, and S. Haykin, "R-HINT-E: A realistic hearing in noise testing environment," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 4, Montreal, Canada, May 2004, pp. 5–8.

[91] M. Wilbur, T. N. Davidson, and J. P. Reilly, "Efficient design of oversampled NPR GDFT filter banks," *IEEE Trans. Signal Processing*, vol. 52, pp. 1947–1963, July 2004.

[92] H. Wu, "Blind deconvolution using Bayesian methods and its application to the dereverberation of speech," M.A.Sc. thesis, Dept. of Elec. and Comp. Eng., McMaster University, Hamilton, ON, Canada, 2003.