INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

**SOME IMPROVEMENTS IN STATE/PARAMETER ESTIMATION
USING THE CELL-TO-CELL MAPPING TECHNIQUE**

DISSERTATION

**Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University**

By

Peng Wang , B.S., M.S.

**The Ohio State University
2001**

Dissertation Committee:

**Dr. Tunc Aldemir, Adviser
Dr. V.I.Utkin
Dr Don W. Miller
Dr. Richard S. Denning**

Approved By

**Adviser
Nuclear Engineering Graduate Program**

UMI Number: 3039539

UMI[®]

UMI Microform 3039539

Copyright 2002 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

The DSD (Dynamic System Doctor) is a newly developed parameter/state estimation methodology which is based on the cell-to-cell mapping technique (CCMT) and can be used for parameter/state estimation in an arbitrary non-linear dynamic system. Compared with conventional estimators, the most important feature of the technique is that it is capable of assigning probabilities to the parameter/state variables occupying user specified magnitude intervals. Such an assignment allows rank ordering the likelihood of the possible states the system can be in at a given time, as well as identifying the uncertainty in the system state. In this respect, the procedure is particularly suited for probabilistic risk analysis and uncertainty analysis in the case of decision-making and safety concerns.

One limitation of this technique is that it assumes uniformly distributed measurement noise and model uncertainty. Also, only individual state variables are assumed to be observable and arbitrary observers are not allowed. This dissertation presents a new theoretical basis for CCMT based estimation which waives these assumptions and increases the applicability of DSD to arbitrary observers and arbitrary noise distributions in the monitored data as well as to arbitrary uncertainties in the system model. The new theoretical basis also leads to improvements in estimation speed

and accuracy and makes DSD capable of validating noisy measurement. The relationship between DSD and several conventional estimation techniques is also studied theoretically and numerically.

Dedicated to my wife Fei Pan

ACKNOWLEDGMENTS

I wish to thank my adviser, Dr Tunc Aldemir, for intellectual support, encouragement, and enthusiasm which made this dissertation possible, and for his patience in correcting both my stylistic and scientific errors.

Also, thanks go to Drs V.I.Utkin, Don W. Miller and Richard S. Denning for their encouragements and suggestions.

To my wife, I thank you for your support. To my parents, I thank you for understanding my long absence.

VITA

March 17, 1971	Born - HeBei, China
July 1994	B.E. Engineering Physics, Tsinghua University
September 1994-July 1996	Research Assistant INET, Tsinghua University
March 2000	M.S. N.E., The Ohio State University Columbus, Ohio
September 1996-present	Graduate Research Assistant The Ohio State University

PUBLICATIONS

Research Publication

1. P.Wang, XiaoMing Chen, Tunc Aldemir, "DSD: A Generic Software Package for Model-Based Fault Diagnosis in Dynamic Systems", Reliability Engineering & System Safety (Accepted).
2. P.Wang, T.Aldemir, Vadim I. Utkin, "Estimation of Xenon Concentration and Reactivity in Nuclear Reactor Using Sliding Mode Observer", 40th IEEE Conference on Decision and Control, Orlando, Florida (Dec. 2001).
3. T.Aldemir, P.Wang, D.Miller, "Parameter and State Estimation Using DSD", Proceeding of ANS 2001 Annual Meeting, Milwaukee, WI (June 2001).
4. Hailong Jin, Matteo Gadioli, Peng Wang, Tunc Aldemir, Don W.Miller, Joseph W. Talnagi Jr "Validation of A Point Reactor Model For Global Xenon Estimation", Proceeding of ICONE 8, Baltimore, MD (April 2000).

5. P.Wang, T.Aldemir, "Real-Time Xenon Estimation in Nuclear Power Plants", Transactions of American Nuclear Society, **81**,154-156 (1999).
6. P.Wang, T.Aldemir, E.Zio, "An On-line Tool for Risk-based Accident Management", Safety and Reliability, G.I.Schueller,P.Kafka (Eds.), 439-444, A.A.Balkema, Rotterdam (1999).
7. X.M.Chen, P.Wang, T.Aldemir, "DSD; A Parameter/State Estimation Tool For Model-based Fault Diagnosis in Non-linear Dynamic Systems", Dynamic Reliability: Future Directions, International Workshop Series on Advanced Topics in Reliability And Risk Analysis Center for Reliability Engineering, University of Maryland (1999).
8. T.Aldemir, P.Wang, "The Use of the Cell-to-Cell Mapping Technique as a Model-Based Diagnostic Tool", M&C'99-Madrid, J.M.Aragones, C.Ahnert, O.Cabellos (Eds.), 1330-1338, Senda Editorial, S.A., Madrid, Spain (1999) (invited paper).
9. X.M.Chen, P.Wang, T.Aldemir, "Development of a Software Package for Parameter/State Estimation in Non-Linear Systems", Transactions of American Nuclear Society, **79**,152-154 (1998).
10. P.Wang, T.Aldemir, D.W.Miller, " Reactivity Ramp Rate Estimation Using A Probabilistic Approach", Transactions of American Nuclear Society, **78**, 121-123 (1998).

FIELDS OF STUDY

Major Field: Nuclear Engineering

Studies in: Reliability Engineering

Statistical Dynamic System Analysis

Reactor Dynamics and Control

TABLE OF CONTENTS

ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGMENTS	v
VITA.....	vi
LIST OF TABLES.....	x
LIST OF FIGURES	xi
NOMENCLATURE	xv
CHAPTERS:	
1 INTRODUCTION	1
1.1 Overview	1
1.2 Literature Review.....	2
1.3 Objectives	8
1.4 Dissertation Organization	11
2. PREVIOUS WORK ON DSD.....	13
2.1 Theoretical Basis of the Current DSD Algorithm.....	13
2.2 The Current DSD Algorithm	20
3. THE DERIVATION OF DSD FOR GENERALIZED NOISE DISTRIBUTION	24
3.1 A Continuous Bayesian Estimator	24
3.2 Derivation of the Discrete Form of the Bayesian Estimator	31
3.3 Example Application	39

3.3.1 System Description.....	39
3.3.2 Estimation Results With Gaussian Noise (Case I)	41
3.3.3 Estimation Results With Uniform Noise (Case II).....	51
3.3.4 Estimation Results for Case III.....	62
4.RELATIONSHIP BETWEEN DSD, GENERALIZED MAXIMUM LIKELIHOOD ESTIMATOR AND THE LEAST SQUARES ESTIMATOR	68
4.1 Relationship Between DSD and GMLE	68
4.2 Equivalency of GMLE and LSE	74
4.3 Illustration	76
4.3.1 System Description.....	76
4.3.2 Estimation Results	81
5. THE DSD AND THE KALMAN FILTER	94
5.1 The Continuous Bayesian Estimator and The Kalman Filter	94
5.2 Numerical Examples.....	110
6. CONCLUSIONS AND RECOMMENDATIONS.....	118
6.1 Conclusions.....	118
6.2 Recommendations for Future Work.....	120
REFERENCES.....	121

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1 Parameters Used for Simulation of System Evolution for Case I	42
3.2 Partitioning Data Used for Case I & II	42
3.3 Parameters Used for Simulation of System Evolution for Case II	51
3.4 Parameters Used for Simulation of System Evolution for Case III	62
3.5 Partitioning Data Used for Case III	64
4.1 Estimation Results of the LS Estimator	82
4.2 Estimation Results of the GMLE Estimator	85
4.3 Partitioning Data Used for the CTPS Sensor Model	87
4.4 Estimated Most Probable Values at the Final Time Step for the CTPS Sensor Model.....	90
4.5 Estimated Credibility Intervals for the CTPS Sensor Model	92
4.6 Probability of Selected Intervals for the CTPS Sensor Model	92
5.1 Partitioning Data Used for Case $A = 1, C = 1, M = 1, R = 4$	111
5.2 Partitioning Data Used for Case $A = 0.9, C = 1, M = 9, R = 4$	115
5.3 Partitioning Data Used For Case $A = 1.05, C = 1, M = 4, R = 1$	115

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
2.1	Geometric Representation of Transition Probability Calculation	23
3.1	Simulated $x_1(t)$ and $y_1(t)$ for Case I.....	43
3.2	Simulated $x_2(t)$ and $y_2(t)$ for Case I.....	43
3.3	Estimated Interval of x_1 as a Function of Time for Case I	46
3.4	Estimated Interval of x_2 as a Function of Time for Case I.....	46
3.5	Estimated Interval of α as a Function of Time for Case I.....	47
3.6	Mean Value of the Estimated Probability Distribution Function of x_1 for Case I	47
3.7	Mean Value of the Estimated Probability Distribution Function of x_2 for Case I	48
3.8	Mean Value of the Estimated Probability Distribution Function of α for Case I	48
3.9	The Ratio of the Estimated Posterior Standard Deviation to the Standard Deviation of the Noise for x_1	49
3.10	The Ratio of the Estimated Posterior Standard Deviation to the Standard Deviation of the Noise for x_2	49
3.11	The Evolution of $p(\alpha \bar{y}_t)$ for the van der Pol Oscillator for Case I: a) $t = \tau$, b) $t = 20\tau$, c) $t = 40\tau$, d) $t = 80\tau$	50
3.12	Simulated $x_1(t)$ and $y_1(t)$ for Case II	53

3.13	Simulated $x_2(t)$ and $y_2(t)$ for Case II.....	53
3.14	Estimated Interval of α as a Function of Time Using the Extended DSD Algorithm for Case II	54
3.15	Estimated Interval of x_1 as a Function of Time Using the Extended DSD Algorithm for Case II	54
3.16	Estimated Interval of x_2 as a Function of Time Using the Extended DSD Algorithm for Case II	55
3.17	Mean Value of the Estimated Probability Distribution Function of α for Case II	55
3.18	Mean Value of the Estimated Probability Distribution Function of x_1 for Case II	56
3.19	Mean Value of the Estimated Probability Distribution Function of x_2 for Case II	56
3.20	Estimated Interval of α as a Function of Time Using the Rules in [1] for Case II	57
3.21	Estimated Interval of x_1 as a Function of Time for Case II Using Equation (3.40)	57
3.22	Estimated Interval of x_2 as a Function of Time for Case II Using Equation (3.40)	58
3.23	Estimated pdf of α for Case II at $t = 80\tau$: (a) DSD Algorithm Results, (b) Extended DSD Algorithm Results.....	60
3.24	Estimated pdf for State Variables for Case II at $t = 50\tau$ Using the Extended DSD Algorithm: (a) $p(x_1 \bar{y}_{50})$, (b) $p(x_2 \bar{y}_{50})$	61
3.25	Simulated $y(t)$ for Case III.....	63
3.26	Estimated Interval of x_1 as a Function of Time for Case III.....	65
3.27	Estimated Interval of x_2 as a Function of Time for Case III.....	65
3.28	Estimated Interval of α as a Function of Time for Case III	66

3.29	Mean Value of the Estimated Probability Distribution Function of x_1 for Case III.....	66
3.30	Mean Value of the Estimated Probability Distribution Function of x_2 for Case III.....	67
3.31	Mean Value of the Estimated Probability Distribution Function of α for Case III.....	67
4.1	A Schematic Diagram of CTPS sensor	78
4.2	Comparison of the Measured Data and Fitting Results By the Least Square Method.....	84
4.3	Plot of the Residuals.....	84
4.4	Estimated Interval of θ_1 as a Function of Time for CTPS Sensor Model.....	88
4.5	Estimated Interval of θ_2 as a Function of Time for CTPS Sensor Model.	88
4.6	Estimated Interval of θ_3 as a Function of Time for CTPS Sensor Model	89
4.7	Estimated Interval of θ_4 as a Function of Time for CTPS Sensor Model	89
4.8	Estimated pdf of the Unknown Parameters at the Final Time Step $k = 250$: (a) $p(\theta_1 \bar{T}_{1,k})$, (b) $p(\theta_2 \bar{T}_{1,k})$, (c) $p(\theta_3 \bar{T}_{1,k})$, (d) $p(\theta_4 \bar{T}_{1,k})$	93
5.1	Actual vs. Approximate Posterior Probability Distribution at Time Step $k = 10$	112
5.2	Relative Error In Posterior Mean and Variance Between the Actual and Approximate Probability Distribution Function At Each Time Step ($A = 1, C = 1, M = 1, R = 4$): (a) Relative Error in Posterior Mean, (b) Relative Error in Posterior Variance.....	113
5.3	Relative Error In Posterior Mean and Variance Between the Actual and Approximate Probability Distribution Function At Each Time Step ($A = 0.9, C = 1, M = 9, R = 4$): (a) Relative Error in Posterior Mean, (b) Relative Error in Posterior Variance	116

5.4	Relative Error In Posterior Mean and Variance Between the Actual and Approximate Probability Distribution Function At Each Time Step ($A = 1.05, C = 1, M = 4, R = 1$): (a) Relative Error in Posterior Mean, (b) Relative Error in Posterior Variance	117
-----	--	-----

NOMENCLATURE

Abbreviations

ARMA	: Autoregressive Moving Average
CCMT	: Cell to Cell Mapping Technique
CTPS	: Constant Temperature Power Sensor
DSD	: Dynamic System Doctor
GMLE	: Generalized Maximum Likelihood Estimator
LSE	: Least Square Estimator
MIMO	: Multiple Input Multiple Output
MSLO	: Modified Stochastic Luenberger Observer
NLP	: Non-Linear Programming
PLID	: Pseudo-Linear Identification
pdf	: Probability Distribution Function
SISO	: Single Input Single Output

Roman Symbols

f	: time derivative function
v	: system noise
w	: measurement noise
h	: observer function

x	: state variable vector
y	: measured data
g	: transition probability
j	: cell location in the discretized space
k	: time step index

Greek/Mathematical Symbols:

τ	: time step interval
δ	: Kronecker delta function
θ	: system parameter vector
σ	: standard deviation
Σ	: covariance matrix

CHAPTER 1

INTRODUCTION

1.1 Overview

Parameter estimation techniques play an important role in dynamic system analysis for fault detection, system identification, and adaptive control. Although parameter estimation in linear systems is by now a well-established field, nonlinear system parameter estimation is still a popular research area. Recently, a model-based statistical method using the cell-to-cell mapping technique (CCMT) has been proposed^{1,2}. The basic idea of this technique is to use the CCMT with the Chapman-Kolmogorov equation to locate the system in the parameter/state space. The research presented here is a new theoretical basis for this approach and also shows its relation to other conventional parameter estimation techniques.

This chapter provides the background for the research presented in this dissertation (Section 1.2) and gives the objectives of the research (Section 1.3). The dissertation organization is given in Section 1.4.

1.2 Literature Review

Among the large number of conventional parameter estimation techniques used for nonlinear dynamic system analysis, the largest category is based on optimizing the value of the estimates by minimizing a pre-defined objective function or loss function (i.e. least squares, weighted least squares, minimum mean square error). These techniques usually end up with solving a linear or nonlinear minimization problem with or without constraints. Closed form solution of a nonlinear minimization problem is usually not available and numerical methods (i.e. dynamic programming, discrete Euler-Lagrange equations, Gaussian-Newton method) have been proposed and widely used. Mook³ considers the optimal state and parameter estimation problems in conditions of model error for nonlinear dynamic systems. The unknown parameters are estimated using the least square method by minimizing the model error estimate. The minimization problem is solved by solving a two-point boundary-value problem. The authors argue that this method is accurate and robust with respect to large errors in system model and measured data but the model error estimate may contain jump discontinuities in case of noisy measurement. In Albuquerque and Biegler's paper⁴, the objective function is defined as the sum of a set of functions which depend on the states and inputs at one time, and the constraints are the differential equations which describe the dynamic process. The authors define a nonlinear dynamic programming (NLP) problem by discretizing the differential equations. Even though the authors propose a faster solver for the NLP problem, the major problem with this approach is that size of the NLP problems grows linearly with the number of data sets and heavy computational load

makes it impossible to solve large problems directly. Guay and McLean⁵ present a method for parameter estimation in nonlinear systems modeled by a set of ordinary differential equations by optimization of the Box and Draper (or the least squares) criterion⁶. The optimization of the Box and Draper criterion is performed using four different techniques including the Gauss-Newton method, full Newton method, and two hybrid methods. The decoupled direct method is applied to evaluate the gradient and Hessian matrix of the objective function with respect to the parameters. Authors state that using second order sensitivity coefficients to evaluate the Hessian matrix can lead to more accurate results. However, the computation of higher order sensitivity coefficients increases drastically the computational time and storage. The computational load and storage is also the major problem associated with Hjelmstad's⁷ method in order to obtain acceptable accuracy. Another least-squares method, developed by Liu⁸, includes a fast adaptive least-squares algorithm for parameter estimation which is based on Householder transformations. The author indicates that this algorithm requires computation and storage load in the order of $O(N)$ instead of $O(N^2)$ where N is the number of parameters to be estimated. One limitation of this algorithm is that it requires the system to be linear with respect to the unknown parameters. In addition, the estimated parameter does not converge to the true value in case of correlated noise. Dimogianopoulos and Lozano⁹ propose a least-squares-based non-recursive identification algorithm in their paper. The author chose the L_2 norm of the identification error with a forgetting factor as the objective function to be minimized. This technique is capable of dealing with slowly time-varying parameters without explicit knowledge of

the noise bound or the region where the true parameters lie. However, the bounds on the noise and the parameter variations should be small to obtain meaningful properties of the estimates.

Another widely used technique in nonlinear system parameter estimation problems is to use linearized system models about an operating point. Ghanadan and Blakenship¹⁰ propose an approximate model and parameter update computed in terms of Taylor series for adaptive approximate feedback linearization of non-linear systems under parameter uncertainty. Young and Suresh¹¹ consider parameter estimation in presence of noise for discrete time linear models with time delays, which may introduce nonlinearities in identifying system parameters. Hopkins and Van Landingham^{12,13} propose a method for stochastic linear time-invariant discrete-time systems with single input single output (SISO)¹² and with multiple input multiple outputs (MIMO)¹³ which are completely controllable and observable. This method is called pseudo-linear identification (PLID) and is a nonlinear method of simultaneous parameter and state estimation. Authors state that PLID is known to converge but this convergence may be a practical impossibility in the presence of a large amount of noise. Douce and Zhu¹⁴ describe a method for the modeling of nonlinear single-input single-output systems using a modified least square method. The application of this method requires that the nonlinear systems are perturbed only in a restricted operating range and the system can be linearized inside this operating range. Douce and Zhu use the autoregressive moving average (ARMA) model in order to approximate a wide range of nonlinear systems. A weighted least square algorithm for parameter estimation is proposed in order to handle

the nonlinear characteristics of the dynamic system. Chen¹⁵ presents a structural classification of the parameter estimation methods which can be applied using Voltera and Wiener kernels. The kernels are evaluated from the input-output measurements. The noise present in the measurement may cause kernel estimation errors. Strongly nonlinear systems require higher order Voltera and Wiener kernels to obtain acceptable accuracy, however, calculating the higher order Voltera and Wiener kernels leads to an appreciable increase in computational load. Recently, Lyashevskiy and Chen¹⁶ have developed an innovative identification procedure by applying the so-called harmonic linearization method. Using this technique, the nonlinearity of the system is replaced by a set of harmonic linearized elements which are obtained through the describing function method. The unknown parameters are found from a set of algebraic equations which are obtained from the amplitude and frequency of the self-oscillations. One limitation of this technique is that it can be applied only for dynamic systems with limit cycle.

A technique that can perform parameter estimation and also deal with unobservable states and variables in nonlinear dynamic systems is to generate observers. Ricardo and Tomei¹⁷ demonstrate adaptive observers that guarantee fast exponential convergence. However, this method is only applicable for a certain class of nonlinear systems under persistency of excitation conditions. The construction of the adaptive observers requires the system to be in adaptive observer form. If the system is not in an adaptive observer form, then filtered transformations can be used to bring the system in adaptive observer form. Hsieh and Chen¹⁸ propose a Modified Stochastic Luenberger observer (MSLO). In this paper, the MSLO is derived directly from the general two-

stage Kalman filter and the designed observer is also proved to be optimal in time-varying, linear discrete-time systems. Sliding mode control and estimation techniques have also been shown to be effective tool in parameter and states estimation problems. McCann, Islam and Husain¹⁹ have applied the sliding-mode observer method to the operation of a switched reluctance motor. A sliding-mode observer is designed to estimate the rotor position and velocity. Sliding mode observers have also been used in reactivity and xenon concentration estimation problems by Wang, Aldemir and Utkin²⁰. Kim, Rizzoni and Utkin²¹ apply the sliding-mode estimation idea into an automotive engine diagnosis and control problem. The unknown charge efficiency, mean mass flow rate of air and throttle angle are estimated through a well-designed sliding mode observer. The results show that the observer estimates the actual states within a reasonable accuracy. The sliding mode approach to parameter/state estimation in nonlinear dynamic systems consists of designing the model with unknown parameters/states replaced by discontinuous functions and enforcing sliding modes such that the model and plant outputs coincide. Then the average values of the discontinuous functions depend on the unknown states/parameters and can be used for their evaluation.

Kalman filter is a well-known estimation technique which has been applied widely in control area. D'Attelis and Cortina²² use Kalman filters to obtain simultaneous estimation of the neutron density and reactivity in an operating nuclear reactor. A probabilistic test is suggested to determine the filter of the bank with the best performance. The method can account for the possible corruption of neutron density measurement by Gaussian white noise. A problem which can arise with using this

method is that Kalman filters are at best locally convergent. Glielmo et al²³ propose an interlaced extended Kalman filter which deals with a class of nonlinear systems whose structure can be partitioned into subsystems. The interlaced Kalman filter consists of a set of Kalman filters, each of them is designed to estimate a subset of the state variables, considering the other state variables as known parameters. The error introduced is partially compensated by increasing the noise covariance matrix. The author states that this technique has proved working well, but heavy computational burden is a problem. A recent study by Boutayeb²⁴ provides a recursive identification technique for nonlinear system, in the presence of unknown but bounded disturbances. This technique is designed based on the modified extended Kalman filter technique and weak conditions are given to ensure local asymptotic convergence.

Statistical parameter estimation methods are also used in dynamic system analysis. The advantage of statistical method is that these methods will deal with uncertainty and noise directly. One commonly used statistical parameter estimation technique is Bayesian estimation. Bolviken et al²⁵ use Bayesian state estimation for nonlinear systems. The unknown state vectors are drawn many times from the posterior distribution and the average is used to be the approximate of the posterior mean. The authors apply the Monte Carlo technique to randomly pick up points in the range of interest and argue that the Monte Carlo uncertainty is small compared to the inherent uncertainty in the optimal Bayesian estimate. The problem of this technique is that it gives only sample points of the posterior probability distribution, the real probability distribution is still unavailable. Charalambous and Logothetis²⁶ apply the maximum

likelihood idea into the nonlinear stochastic system parameter estimation problems. The expectation maximization algorithm, an iterative numerical method, is used to generate the parameter estimates by computing the log-likelihood ratio restricted to the measured data. The problem of this technique is that it is still a point estimator without providing an efficient way to evaluate the uncertainty of the estimation results.

Expert systems, neural networks and genetic algorithms have been also used for parameter estimation purposes. Parlos et al²⁷ use artificial neural networks to perform the identification of a nonlinear model for a U-tube steam generator. The system identification consists of estimating unknown parameters and/or system variables which cannot be monitored in order to obtain a complete model for the system. A similar work by Patton, Lopez-Toribio and Uppal²⁸ applies artificial intelligence techniques for fault detection and identification in process systems. Marseguerra and Zio²⁹ use the genetic algorithm to optimize the objective function for estimating the effective nuclear parameters and the non-measured Xenon and Iodine initial conditions in a nuclear reactor. The authors state that the estimation results given by this approach are very close to the true values.

1.3 Objectives

The literature survey shows that the parameter/state estimation of nonlinear systems seems to have the following difficulties:

1. Heavy computational load and large computer memory are required if accurate estimates of the parameters are desired. This requirement usually reduces the capability of most methods for on-line implementation.
2. Noise is difficult to handle, especially when the noise is large.
3. Most methods cannot account for random variations in the parameters.
4. Most estimators are point estimators without providing measures on the uncertainty in the results.
5. Estimators need to be individually designed for given system.

The recent development of a parameter estimation technique at The Ohio State University may reduce these difficulties. The estimation algorithm, developed by Dinca^{1,2}, models the system evolution in terms of probability of transitions in time between sets of user defined parameter/state variable magnitude intervals (cells). These probabilities are determined from the user specified system model using quadratures. The most important feature of the methodology with regard to fault diagnosis is that it can account for uncertainties in the monitored system state, inputs and modeling uncertainties through the appropriate choice of the cells, as well as providing a measure to rank the likelihood of faults in view of these uncertainties. Other important features of this methodology are

1. It does not require linearization of the system.
2. It allows flexibility in system representation. Differential or difference equations as well as almost any type of input/output model (e.g. neural net,

response surface) can be used to generate the cell-to-cell transition probabilities.

3. The discrete-time nature of the methodology is directly compatible with a look-up table implementation, which is very convenient for the use of data that may be available from tests or actual incidents.
4. It does not require model inversion (which may lead to singularity problems) or inverse models (which usually have limited range of applicability).
5. It is both an interval and a point estimator. Subsequently, it yields the lower and upper bounds on the estimated values of state variables/parameters as well as their expected values. Knowledge of such bounds is particularly important in the determination of safety margins.

A software package called DSD^{30,31} (Dynamic System Doctor) has been developed for the on-line implementation of this methodology in a system independent manner and has been successfully tested on a variety of nuclear and mechanical dynamic systems^{32,33,34}. The estimation algorithm of DSD is based on the assumption that the measurement noise is uniformly distributed and the measured data are restricted to components of the state variable vector. Some current difficulties in the implementation of DSD are the following:

1. Current estimation algorithm in DSD is not able to handle model uncertainties and measurement noise with arbitrary probability distributions.
2. Arbitrary observers are not allowed.

3. Directly measured components of the state variable vector are used in the estimation process without further validation, which may cause convergence problems in the case that the measurement uncertainty is high.
4. Present theory does not indicate the relationship between this technique and other conventional techniques (e.g. least square estimator, general maximum likelihood estimator, Kalman filter).
5. The accuracy of the presently applied quadrature scheme in approximating the transition probability cannot be evaluated because of the lack of closed form presentation of the transition probability.

The objectives of this study are the following:

1. Develop the theoretical basis for DSD for generalized model/measurement uncertainty distributions and arbitrary observers.
2. Determine relationship between DSD and conventional techniques.
3. Derive a closed form of the transition probabilities.
4. Integrate measurement signal validation process (in the sense of reducing the uncertainty in the signal) into the estimation algorithm.

1.4 Dissertation Organization

The dissertation is organized as follows. Chapter 1 gives the literature survey on parameter/state estimation. Chapter 2 provides an overview of the previous work on DSD. Chapter 3 presents the new generalized theoretical basis for DSD. In Chapter 4, the equivalence between DSD, the general maximum likelihood estimator and least

square estimator is proved. Chapter 5 analyzes the relationship between the Kalman filter approach and the DSD. The conclusions of the study and recommendations for future work are given in Chapter 6.

CHAPTER 2

PREVIOUS WORK ON DSD

The DSD (Dynamic System Doctor) is a newly developed parameter/state estimation methodology which is based on the cell-to-cell mapping technique (CCMT) and can be used for parameter/state estimation in an arbitrary non-linear dynamic system. The previous work on DSD^{30,31} assumes: a) the measured data can be contained within the cells in the monitored state variable space, and, b) some of the state variables are directly measurable. Section 2.1 introduces the theoretical basis of the current DSD method based on these assumptions and Section 2.2 discusses possible algorithms for the implementation of the method presented in Section 2.1. The theory in this chapter is adopted from Ref [1].

2.1 Theoretical Basis of the Current DSD Algorithm

Consider a dynamic system described by a set of first order differential equations

$$\frac{dx}{dt} = f(x, a) \quad (2.1)$$

where $\mathbf{x}$ is a vector whose elements are the system state variables $x_l (l = 1, \dots, L)$, t is time, and $\boldsymbol{\alpha}$ is a vector whose elements are the model parameters $\alpha_m (m = 1, \dots, M)$. The system is assumed to be autonomous, i.e. the system dynamics do not explicitly depend on time. Since a non-autonomous system can be transformed into an autonomous system by making time t an additional system state variable, this assumption does not lead to a loss in generality.

The CCMT represents system evolution in terms of probability of transitions between user specified magnitude intervals or cells in the system parameter/state space during user specified time intervals $k\tau \leq t < (k+1)\tau$ ($k = 0, 1, 2, \dots$). These cells are similar to those used by the finite difference and finite element methods and the cell sizes may correspond to the desired estimation accuracy in unknown state variables or may be defined to contain the noise/uncertainty in the monitored variables.

Assume that system variable $\mathbf{x} = [x_1, x_2, \dots, x_L]$ falls within the range of interest $a_l \leq x_l \leq b_l (l = 1, \dots, L)$. This range of interest is divided into $J_l (l = 1, \dots, L)$ intervals

$$\Delta_{l,j_l} = \{x_l : a_{l,j_l} \leq x_l < a_{l,j_l+1}; a_{l,1} = a_l, a_{l,J_l} = b_l\}$$

$$j_l = 1, \dots, J_l; l = 1, \dots, L$$
(2.2)

The sets $V_j = \{\Delta_{1,j_1}, \Delta_{2,j_2}, \dots, \Delta_{L,j_L}\}$ ($j = 1, \dots, J = \prod_{l=1}^L J_l$) constitute the cells and partition the multi-dimensional state space. Fig. 2.1 illustrates such a possible partitioning for a second order system and indicates the possible cells the system can be in if one of the state variables is directly monitored.

Similarly partition the range of interest $\tilde{a}_m \leq \alpha_m \leq \tilde{b}_m$ ($m = 1, \dots, M$) for the model parameters α_m ($m = 1, \dots, M$) into N_m ($m = 1, \dots, M$) intervals:

$$\begin{aligned} \tilde{\Delta}_{m,n_m} &= \{\alpha_m : \tilde{\alpha}_{m,n_m} \leq \alpha_m < \tilde{\alpha}_{m,n_m+1}; \tilde{\alpha}_{m,1} = \tilde{a}_m, \alpha_{m,N_m} = \tilde{b}_m\} \\ n_m &= 1, \dots, N_m; m = 1, \dots, M \end{aligned} \quad (2.3)$$

The number of intervals for each monitored state variable can be chosen corresponding to the uncertainty level in the observed system state. Some other sufficient conditions for convergence regarding these numbers of intervals are given later in this section.

Assume that the system parameters are fixed at $\alpha \in \{\tilde{\Delta}_{m,n_m}, m = 1, \dots, M, n_m = 1, \dots, N_m\}$ during the time interval $[t, t + \tau)$ but possibly move to a new cell at $t + \tau$ (a discrete jump in the system parameter). During this time interval, the system will move in the state space from the location $x(t)$ to a new location $\tilde{x}(x(t))$ at time $t + \tau$, as can be uniquely determined from integration of system Equation (2.1), i.e.

$$\tilde{x}(x', \alpha, \tau) = \int_t^{t+\tau} f(x(t'), \alpha) dt' + x' \quad (x' = x(t)) \quad (2.4)$$

If the changes in the system parameters are determined by known rules which only depend on the current and future values of the system state variables and the system parameters and not on the history of the system operation, the system under consideration has the Markov property in the state space. Thus, the system evolution can be expressed in the probability domain using the Chapman-Kolmogorov equation

$$p(\mathbf{x}, \mathbf{a}, (k+1)\tau) = \int_{a_1}^{b_1} dx'_1 \dots \int_{a_L}^{b_L} dx'_L p(\mathbf{x}, \mathbf{a}, (k+1)\tau | \mathbf{x}', \mathbf{a}, k\tau) p(\mathbf{x}', \mathbf{a}, k\tau) \quad (2.5)$$

where $p(\mathbf{x}', \mathbf{a}, k\tau)$ is the probability that the system is located at $\mathbf{x}'$ in the state space and at $\mathbf{a}$ in the parameter space at the time $k\tau$ and $p(\mathbf{x}, \mathbf{a}, (k+1)\tau | \mathbf{x}', \mathbf{a}, k\tau)$ is the transition probability which represents the probability that the system will go from the previous location $\mathbf{x}'$ to the new location $\mathbf{x}$ during the time interval $k\tau \leq t < (k+1)\tau$ assuming that the parameter value does not change during this time interval.

Now assume that within each variable or parameter cell, the variables and parameters are uniformly distributed at any time, i.e.

$$\begin{aligned} p(\mathbf{x}', \mathbf{a}, k\tau) = & \sum_{j'_1=1}^{J_1} e_{j'_1}(x'_1) / (a_{1,j'} - a_{1,j'-1}) \dots \sum_{j'_L=1}^{J_L} e_{j'_L}(x'_L) / (a_{L,j'} - a_{L,j'-1}) \\ & \times \sum_{n'_1=1}^{N_1} e_{n'_1}(\alpha_1) / (\tilde{\alpha}_{1,n} - \tilde{\alpha}_{1,n-1}) \dots \sum_{n'_M=1}^{N_M} e_{n'_M}(\alpha_M) / (\tilde{\alpha}_{M,n} - \tilde{\alpha}_{M,n-1}) \\ & \times p(j'_1, \dots, j'_L, n_1, \dots, n_M, k\tau) \quad (k = 0, 1, \dots), \end{aligned} \quad (2.6)$$

where

$$p(j'_1, \dots, j'_L, n_1, \dots, n_M, k\tau) \equiv \quad (2.7)$$

$$\begin{aligned} & \int_{\tilde{\alpha}_{1,n}}^{\tilde{\alpha}_{1,n+1}} d\alpha_1 \dots \int_{\tilde{\alpha}_{M,n}}^{\tilde{\alpha}_{M,n+1}} d\alpha_M \int_{a_{1,j'}}^{a_{1,j'+1}} dx'_1 \dots \int_{a_{L,j'}}^{a_{L,j'+1}} dx'_L p(\mathbf{x}', \mathbf{a}, k\tau) \\ e_{j_i}(x_i) = & \begin{cases} 1 & \text{if } a_{i,j_i} \leq x_i < a_{i,j_i+1} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (2.8)$$

and

$$e_{n_m}(\alpha_m) = \begin{cases} 1 & \text{if } \tilde{\alpha}_{m,n_m} \leq \alpha_m < \tilde{\alpha}_{m,n_m+1} \\ 0 & \text{otherwise} \end{cases} \quad (2.9)$$

Substituting Equation (2.6) to Equation (2.5) and integrating over each Δ_{i,j_i} and $\tilde{\Delta}_{m,n_m}$ yields

$$\begin{aligned} p_{k+1}(j_1, \dots, j_L, n_1, \dots, n_M) = \\ \sum_{j'_1=1}^{j_1} \dots \sum_{j'_L=1}^{j_L} g(j_1, \dots, j_L \mid j'_1, \dots, j'_L, n_1, \dots, n_M, \tau) \\ \times p_{k+1}(j'_1, \dots, j'_L, n_1, \dots, n_M) \end{aligned} \quad (2.10)$$

where

$$p_{k+1}(j_1, \dots, j_L, n_1, \dots, n_M) \equiv p(j_1, \dots, j_L, n_1, \dots, n_M, (k+1)\tau)$$

and

$$g(j_1, \dots, j_L \mid j'_1, \dots, j'_L, n_1, \dots, n_M, \tau) = \quad (2.11)$$

$$\left\{ \begin{aligned} & \int_{a_{1,j'_1}}^{a_{1,j_1}+1} \frac{dx'_1}{a_{1,j'_1+1} - a_{1,j'_1}} e_{j_1}(\tilde{\mathbf{x}}(\mathbf{x}', \mathbf{a}_n, \tau)) \dots \int_{a_{L,j'_L}}^{a_{L,j_L}+1} \frac{dx'_L}{a_{L,j'_L+1} - a_{L,j'_L}} e_{j_L}(\tilde{\mathbf{x}}(\mathbf{x}', \mathbf{a}_n, \tau)) \\ & \text{if } a_l \leq x'_l \leq b_l \\ & 0 \quad \text{otherwise} \end{aligned} \right.$$

denotes the conditional probability that $x'_l((k+1)\tau) \in \Delta_{l,j_l}$ given that $x'_l(k\tau) \in \Delta_{l,j'_l}$

and $\alpha_n \in \tilde{\Delta}_{m,n_m}$ at time $k\tau$. The function $e_{j_l}(x)$ is as defined by Equation (2.8) and

$\tilde{\mathbf{x}}(\mathbf{x}', \mathbf{a}_n, \tau)$ is determined by Equation (2.4).

In general, not all the variables can be measured. If we assume that only $x_1, x_2, \dots, x_i$ are monitored, we can write the joint probability $p_k(j_1, \dots, j_L, n_1, \dots, n_M)$ as

$$p_k(j_1, \dots, j_L, n_1, \dots, n_M) = p_k(j_{i+1}, \dots, j_L, n_1, \dots, n_M | \hat{j}_1, \dots, \hat{j}_i) \delta_{\hat{j}_1, j_1} \dots \delta_{\hat{j}_i, j_i} \quad (2.12)$$

where $\delta_{\hat{j}_l, j_l}$ ($l = 1, \dots, i$) are Kronecker deltas and $p(j_{i+1}, \dots, j_L, n_1, \dots, n_M | \hat{j}_1, \dots, \hat{j}_i)$ is the pdf of the unmonitored variables $x_{i+1}, \dots, x_L$ and parameters $\alpha_1, \alpha_2, \dots, \alpha_M$, given that $x_l \in \Delta_{l, \hat{j}_l}$ ($l = 1, \dots, i$). Substituting Equation (2.12) into Equation (2.10) gives

$$p_{k+1}(j_{i+1}, \dots, j_L, n_1, \dots, n_M | \hat{j}_1, \dots, \hat{j}_i) = \sum_{j'_1=1}^{j_1} \dots \sum_{j'_L=1}^{j_L} g(\hat{j}_1, \dots, \hat{j}_i, j_{i+1}, \dots, j_L | \hat{j}'_1, \dots, \hat{j}'_i, j'_{i+1}, \dots, j'_L, n_1, \dots, n_M, \tau) \times p_k(j'_{i+1}, \dots, j'_L, n_1, \dots, n_M | \hat{j}'_1, \dots, \hat{j}'_i) \quad (2.13)$$

or, equivalently, in a more compact notation

$$p_{k+1}(\bar{j}, n) = \sum_{\bar{j}'} g_k(\bar{j} | \bar{j}', n, \tau) p_k(\bar{j}', n) \quad (2.14)$$

where

$$p_k(\bar{j}, n) \equiv p_k(j_{i+1}, \dots, j_L, n_1, \dots, n_M | \hat{j}_1, \dots, \hat{j}_i) \quad (k = 0, 1, 2, \dots) \quad (2.15)$$

$$g_k(\bar{j} | \bar{j}', n) \equiv g(\hat{j}_1, \dots, \hat{j}_i, j_{i+1}, \dots, j_L | \hat{j}'_1, \dots, \hat{j}'_i, j'_{i+1}, \dots, j'_L, n_1, \dots, n_M, \tau) \quad (2.16)$$

Note that in Equation (2.14), the transition probability $g_k(\bar{j} | \bar{j}', n)$ depends upon the observed data at times $(k+1)\tau$ and $k\tau$ (i.e. $\hat{j}_{k+1}$ and $\hat{j}_k$, respectively) as can be seen from its definition by Equation (2.16) and thus Equation (2.14) defines non-

homogeneous Markov chain. Given $p_0(\bar{j}, n)$, Equation (2.14) yields the probability of the unmonitored variables and parameters to be in the intervals sets $\bar{j}$ and n , respectively, at each time step $k\tau$ and hence allows a ranking of their possible values toward the ranking of possible system faults if the mapping from the parameter/state space to fault space is known.

It has been shown by Dinca^{1,2} that if

$$\begin{aligned} g_k(j_{s,k+1} | j_{s,k}, n_s, \tau) &> g_k(j_{s,k+1} | j_{s,k}, n, \tau) \text{ for all } n \neq n_s, \\ g_k(j_{s,k+1} | j_{s,k}, n_s, \tau) &> g_k(j_{s,k+1} | j_{s,k}, n, \tau) \text{ for all } n \neq n_s, \end{aligned} \quad (2.17)$$

$$g_k(\bar{j}_{k+1} | \bar{j}_{s,k}, n, \tau) > g_k(\bar{j}_{k+1} | \bar{j}_k, n, \tau) \text{ for all } \bar{j}_k \neq \bar{j}_{s,k}, \quad (2.18)$$

and

$$g_k(\bar{j}_{s,k+1} | \bar{j}_k, n, \tau) > g_k(\bar{j}_{k+1} | \bar{j}_k, n, \tau) \text{ for all } \bar{j}_{k+1} \neq \bar{j}_{s,k+1}, \quad (2.19)$$

where $j_{s,l}$ and $n_{s,m}$ denote the correct intervals the variables $x_l (l=i+1, \dots, L)$ and the parameters $\alpha_m (m=1, \dots, M)$ are in, respectively, then

$$\lim_{k \rightarrow \infty} \frac{p_k(\bar{j}_k, n)}{\sum_n \sum_{\bar{j}_k} p_k(\bar{j}_k, n)} \rightarrow \delta_{j_{i+1}, j_{i+1}} \dots \delta_{j_{i,L}, j_{i,L}} \delta_{n_1, n_{1,i}} \dots \delta_{n_M, n_{1,M}} \quad (2.20)$$

irrespective of the choice of $p_0(\bar{j}_0, n)$.

The theoretical basis described above shows that the parameter estimation process is conducted in the probability domain. Subsequently, all the results are expressed in the form of probability distributions, instead of fixed-point values. For the estimated parameters, the resulting probability distributions as well as the time variations of the estimated ranges are of interest. The former is useful to decide on the likelihood of

a given type of fault, the latter provides information on the state of the system (i.e. fault or non-fault) if a mapping provided from the parameter/state space to fault space.

2.2 The Current DSD Algorithm

Three possible algorithms (Algorithm A, B and C) for the implementation of the method described in Section 2.1 are presented in Refs [2] and [31].

The partitioning Δ_{l,j_l} and $\tilde{\Delta}_{m,n_m}$ can be defined through the specification of a_{l,j_l} and $\tilde{a}_{m,n_m}$ using Equations (2.2) and (2.3) off-line (Algorithm A) or on-line (Algorithm B). Then the monitored data are read from the monitors at each time step $t = k\tau$ ($k = 1, 2, 3, \dots$). The initial probability $p_0(\hat{j}_0, n)$ is also specified by the user to initialize the estimation process. The transition probabilities $g_k(\bar{j}_k | \bar{j}_{k-1}, n, \tau)$ can be found from Equation (2.11) and (2.16) using a quadrature scheme. Figure 2.1 illustrates this numerical technique graphically for a second-order system with both state variables monitored and one parameter, α_1 , to be determined. In this example, a 4-point quadrature scheme is used to approximate the integral in Equation (2.11). Since each x_l ($l = 1, 2, 3, \dots$) is assumed to be uniformly distributed over each interval Δ_{l,j_l} , these 4-points are selected as the center of the subcells which divide each cell in Fig. 2.1 into 4 equal parts. Then

$$g(j_1, j_2 | j'_1, j'_2, n_1, \tau) \approx \frac{I(j_1, j_2, j'_1, j'_2, n_1, \tau)}{4}$$

where $I(j_1, j_2, j'_1, j'_2, n_1, \tau)$ is the number of transitions into cell (j_1, j_2) at time $(k+1)\tau$ from cell (j'_1, j'_2) with $\alpha_1 \in \tilde{\Delta}_{1, n_1}$ at time $k\tau$. For example, Fig.2.1 shows that the system is located in cell (7,4) in the state-space at time $k\tau$. During the next time interval, starting from the center points of the 4 subcells, there are 3 transitions from cell (7, 4) to cell (5,2) for $\alpha_1 \in \tilde{\Delta}_{1, n_1}$ and hence $g(5,2 | 7,4, n_1, \tau) \approx 3/4 \approx 0.75$. Similarly, there are three transitions from cell (5,2) to cell (3,4), and hence $g(3,4 | 5,2, n_1, \tau) \approx 0.75$ as well. If all the state variables are monitored, this process is repeated for system locations $\hat{j}_k$ and all possible cells the parameter can be in at time $k\tau$ and the parameter distribution will be updated by the following equation

$$p_{k+1}(n) = \frac{p_k(n)}{\sum_{n'} p_k(n')} \quad (2.21)$$

In Algorithm A and B, the cell size is kept fixed. Algorithm C can enlarge the cell size by a user defined amount to accommodate the transition in order to avoid the so-called “boundary effects”³¹. If some x_i are not being monitored, then the process will also be carried over all possible cells the system might be in time $k\tau$ as well as n (e.g. the shaded cells in Fig. 2.1 if only x_1 is monitored). The probability distribution at time $(k+1)\tau$ is calculated by

$$p_{k+1}(\bar{j}, n) = \frac{p_k(\bar{j}, n)}{\sum_{n'} \sum_{\bar{j}'} p_k(\bar{j}', n')} \quad (2.22)$$

In the case that the parameters remain constant during $0 \leq t < k\tau$, this process will yield the $\tilde{\Delta}_{m,n_m}$ in which the parameter is actually in with increasing confidence level, reaching $\tilde{\Delta}_{m,n_m} = \delta_{n_m,\hat{n}_m}$ for some $\hat{n}_m$ as stated in Equation (2.20).

If the parameters change in time then this process will converge into the new parameter values instead of the old values. If the parameter changes are large enough, then all the transition probabilities $g_k(\tilde{j}_{k+1}|\tilde{j}_k, n, \tau)$ at a given $k\tau$ may be zero. Once the algorithm detects such a situation, the process will reinitialize itself by setting the probability distribution at current time step the same as the initial probability $p_0(\hat{j}_0, n)$ and estimate the new parameter by repeating the above process. The reinitialization will also occur if

- the number of subcells is not large enough such that the transition probability calculated from the quadrature scheme is not a good approximation to the real transition probability,
- if the monitored system location lies close to the boundary of the cell the system is in at time $t = k\tau$ (boundary effect), and,
- if the parameter to be estimated is not observable with the available monitored system variables.

CHAPTER 3

THE DERIVATION OF DSD FOR GENERALIZED NOISE DISTRIBUTION

In Chapter 2, we introduced the original version of the DSD algorithm with uniform noise assumption. In this chapter, a new theoretical basis for DSD is presented, which does not assume uniformly distributed measurement noise and allows arbitrary observers. The new theoretical basis is developed in two steps. In Section 3.1, a continuous Bayesian estimator is presented for an arbitrary dynamic system. Section 3.2 applies this Bayesian estimator to a discretized representation of the system used by CCMT and develops a generalized algorithm (the extended DSD algorithm) applicable under arbitrary stochastic variations in the dynamic system variables and parameters. The resulting improvements in comparison to the DSD algorithm described in Chapter 2 are illustrated in Section 3.3 using the van der Pol oscillator.

3.1 A Continuous Bayesian Estimator

Consider the dynamic system

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) + \mathbf{v} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{w}_k \quad (k = 1, 2, \dots)\end{aligned}\tag{3.1}$$

where $\mathbf{x}$ is defined as before and

$\mathbf{v}$: system noise (e.g. due to stochastic variation of system parameters) or a measure of modeling uncertainties, in general

$\mathbf{x}_k$: state vector at time step $k\tau$ ($k = 0, 1, 2, \dots$)

$\mathbf{y}_k$: M -dimensional vector whose elements y_k are the measured data at time $t = k\tau$ ($k = 0, 1, 2, \dots$)

$\mathbf{w}_k$: measurement uncertainty at time $t = k\tau$ ($k = 0, 1, 2, \dots$)

$\mathbf{h}$: M - dimensional vector whose elements are known nonlinear functions of $\mathbf{x}_k$

$\mathbf{f}$: L - dimensional vector whose elements are known nonlinear functions of $\mathbf{x}_k$

Equation (3.1) does not exclude parameter estimation problems because we can always define an unknown set of parameters $\boldsymbol{\theta}$ that are constant in time as a set of state variables which satisfy the equation

$$\dot{\boldsymbol{\theta}} = 0 \quad (3.2)$$

Similarly, systems whose dynamics explicitly depend on time can be described by Equation (3.1) by regarding time as another state variable satisfying

$$\dot{t} = 1 \quad (3.3)$$

System parameters may be known functions of time. In this research, we assume the parameter change rate is slow enough compared to the data sampling interval τ so that the parameter value can be assumed constant within $k\tau \leq t < (k+1)\tau$ ($k = 0, 1, 2, \dots$).

Then by defining the following new state variables

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{bmatrix} \quad (3.4)$$

Equation (3.1) can be written as

$$\begin{aligned} \dot{\mathbf{z}} &= \begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\mathbf{x}) + \mathbf{v} \\ \mathbf{0} \end{bmatrix} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{C} \mathbf{z}_k) + \mathbf{w}_k \quad (k = 1, 2, \dots) \end{aligned} \quad (3.5)$$

where $\mathbf{C} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and $\mathbf{I}$ is the $L \times L$ identity matrix.

Let $\mathbf{x}_k$ indicate the value of the state variables at time step $t = k\tau$ as before. The estimation problem is stated as the following:

Given an initial guess $p(\mathbf{x}_0 | \mathbf{y}_0)$ of the probability distribution function (pdf) of the unknown state variable vector $\mathbf{x}_0$ at time $t = 0$ and the information vector $\bar{\mathbf{y}}_k = [\mathbf{y}_k, \mathbf{y}_{k-1}, \dots, \mathbf{y}_0]$ which includes all the measurements from $t = \tau$ through $t = k\tau$ and the fictitious measurement $\mathbf{y}_0$, estimate the conditional pdf $p(\mathbf{x}_k | \mathbf{y}_k, \mathbf{y}_{k-1}, \dots, \mathbf{y}_0) = p(\mathbf{x}_k | \bar{\mathbf{y}}_k)$ of $\mathbf{x}_k$ at time $t = k\tau$.

While the main motivation for this problem statement is to lay the groundwork for a more generalized theoretical basis for DSD that will be developed in this chapter, the problem statement also addresses the issue of signal validation as will be also illustrated later in this chapter.

Proposition 3.1

Assume

1. $p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1})$ and $p(\mathbf{x}_{k+1} | \mathbf{x}_k)$ are known pdfs,
2. $\mathbf{w}_k$ are statistically independent for all k , and,
3. $\mathbf{v}$ does not depend on previous system history.

Then $p(\mathbf{x}_k | \bar{\mathbf{y}}_k)$ can be recursively determined from

$$p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1}) = \frac{\int p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k}{\int \int p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k d\mathbf{x}_{k+1}} \quad (3.6)$$

where $\bar{\mathbf{y}}_{k+1} = [\mathbf{y}_{k+1}, \mathbf{y}_k, \dots, \mathbf{y}_0]$.

Assumption 1 implies that the pdfs for $\mathbf{v}$ and $\mathbf{w}_k$ are known. A practically important situation that satisfies all these assumptions is when $\mathbf{v}$ represents random fluctuations in system parameters and $\mathbf{w}_k$ corresponds to white noise. However, these assumptions do not exclude correlated noise since correlated noise can be regarded as the transformation of a white noise process through a linear filter (when the power spectral density of the correlated noise is in rational form).³⁵

Proof

From the definition of conditional probability and the information vector $\bar{\mathbf{y}}_k$

$$p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1}) = \frac{p(\mathbf{y}_{k+1}, \mathbf{x}_{k+1} | \bar{\mathbf{y}}_k)}{\int p(\mathbf{y}_{k+1}, \mathbf{x}_{k+1} | \bar{\mathbf{y}}_k) d\mathbf{x}_{k+1}} \quad (3.7)$$

Now consider the following conditional pdfs

$$\begin{aligned}
p(y_{k+1} | x_{k+1}, x_k, \bar{y}_k) &= \frac{p(y_{k+1}, x_{k+1}, x_k, \bar{y}_k)}{p(x_{k+1}, x_k, \bar{y}_k)} \\
p(x_{k+1} | x_k, \bar{y}_k) &= \frac{p(x_{k+1}, x_k, \bar{y}_k)}{p(x_k, \bar{y}_k)} \\
p(x_k | \bar{y}_k) &= \frac{p(x_k, \bar{y}_k)}{p(\bar{y}_k)}
\end{aligned} \tag{3.8}$$

From Equation (3.8), the joint pdf $p(y_{k+1}, x_{k+1} | \bar{y}_k)$ can be written as

$$\begin{aligned}
&p(y_{k+1}, x_{k+1} | \bar{y}_k) \\
&= \int p(y_{k+1}, x_{k+1}, x_k | \bar{y}_k) dx_k \\
&= \int \frac{p(y_{k+1}, x_{k+1}, x_k, \bar{y}_k)}{p(\bar{y}_k)} dx_k \\
&= \int \frac{p(y_{k+1}, x_{k+1}, x_k, \bar{y}_k)}{p(x_{k+1}, x_k, \bar{y}_k)} \frac{p(x_{k+1}, x_k, \bar{y}_k)}{p(x_k, \bar{y}_k)} \frac{p(x_k, \bar{y}_k)}{p(\bar{y}_k)} dx_k \\
&= \int p(y_{k+1} | x_{k+1}, x_k, \bar{y}_k) p(x_{k+1} | x_k, \bar{y}_k) p(x_k | \bar{y}_k) dx_k
\end{aligned} \tag{3.9}$$

Note that the conditional probabilities in Equation (3.8) have the following properties:

$$\begin{aligned}
p(y_{k+1} | x_{k+1}, x_k, \bar{y}_k) &= p(y_{k+1} | x_{k+1}) \\
p(x_{k+1} | x_k, \bar{y}_k) &= p(x_{k+1} | x_k)
\end{aligned} \tag{3.10}$$

Equation (3.10) holds because y_{k+1} only depends on the value of x_{k+1} and w_{k+1} by Equation (3.1). When x_{k+1} is given, the probability of y_{k+1} will be determined by the value of w_{k+1} whose value does not depend on any previous system states by Assumption 2. Similarly, x_{k+1} only depends on x_k and v by Equation (3.1) and v is independent of system history by Assumption 3. Using Equation (3.9) and (3.10), Equation (3.7) can be written as

$$p(x_{k+1} | \bar{y}_{k+1}) = \frac{\int p(y_{k+1} | x_{k+1}) p(x_{k+1} | x_k) p(x_k | \bar{y}_k) dx_k}{\iint p(y_{k+1} | x_{k+1}) p(x_{k+1} | x_k) p(x_k | \bar{y}_k) dx_k dx_{k+1}}$$

which completes the proof.

Equation (3.6) is a Bayesian rule for recursive determination of $p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1})$.

As indicated earlier, a practically important situation is when the modeling uncertainties and the measurement uncertainties are represented by zero mean white Gaussian noises. In this situation,

$$\mathbf{x}_{k+1} = \tilde{\mathbf{x}}(\mathbf{x}_k) + \Delta \mathbf{B}_t \quad (3.11)$$

where

$$\Delta \mathbf{B}_t = \int_{k\tau}^{(k+1)\tau} d\mathbf{B}_t = \int_{k\tau}^{(k+1)\tau} \mathbf{v}(t) dt$$

$$\tilde{\mathbf{x}}(\mathbf{x}_k) = \int_{k\tau}^{(k+1)\tau} \mathbf{f}(\mathbf{x}(s)) ds + \mathbf{x}_k$$

and $\mathbf{B}_t$ denotes the Brownian motion³⁶ whose pdf is the joint normal distribution of the elements of $\Delta \mathbf{B}_t$. Then from Equation (3.11) we have

$$p(\mathbf{x}_{k+1} | \mathbf{x}_k) = \frac{1}{(2\pi\sigma_p^2)^{L/2}} e^{-\frac{(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k))'(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k))}{2\sigma_p^2}} \quad (3.12)$$

where σ_p is the standard deviation of modeling uncertainty. Similarly, when $\mathbf{w}_k$ is white Gaussian noise, from Equation (3.1) we can write

$$p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) = \frac{1}{(2\pi\sigma_m^2)^{M/2}} e^{-\frac{(\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1}))'(\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1}))}{2\sigma_m^2}} \quad (3.13)$$

where again σ_m is the standard deviation of the noise. Substituting Equations (3.12) and (3.13) into Equation(3.6) yields the Bayesian rule for the recursive estimation of the

conditional distribution $p(\mathbf{x}_k | \bar{\mathbf{y}}_k)$ when the modeling and measurement uncertainties are random white Gaussian noise

$$p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1}) = \frac{\int e^{-\left[\frac{(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k))'(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k))}{2\sigma_p^2} + \frac{(\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1}))'(\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1}))}{2\sigma_m^2} \right]} p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k}{\iint e^{-\left[\frac{(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k))'(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k))}{2\sigma_p^2} + \frac{(\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1}))'(\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1}))}{2\sigma_m^2} \right]} p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k d\mathbf{x}_{k+1}} \quad (3.14)$$

Note that if $\mathbf{h}(\mathbf{x}_k) = \mathbf{x}_k$, then Equation (3.14) can be used as a recursive rule for signal validation as illustrated in Section 3.3.2 and 3.3.3.

In the case there is no system noise/or modeling uncertainties and $\tilde{\mathbf{x}}(\mathbf{x}_k)$ (see Equation (3.11)) is invertible, then $\mathbf{v} = \mathbf{0}$ and

$$p(\mathbf{x}_{k+1} | \mathbf{x}_k) = \delta(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k)) = \delta(l(\mathbf{x}_{k+1}) - \mathbf{x}_k) \quad (3.15)$$

where $l(\mathbf{x}_{k+1})$ is the inverse of $\tilde{\mathbf{x}}(\mathbf{x}_k)$ and δ denotes the Dirac delta function, i.e.

$$\int_{\varepsilon} \delta(\mathbf{x} - \hat{\mathbf{x}}) d\mathbf{x} = 1 \quad (3.16)$$

with ε as an infinitesimally small ball around $\hat{\mathbf{x}}$ and $\delta(\mathbf{x} - \hat{\mathbf{x}}) = 0$ for $\mathbf{x} \neq \hat{\mathbf{x}}$. As an example for the inverse $l(\mathbf{x}_{k+1})$, consider the system

$$\dot{\mathbf{x}} = \mathbf{x}$$

Then

$$\mathbf{x}_{k+1} = \tilde{\mathbf{x}}(\mathbf{x}_k) = \frac{\mathbf{x}_k}{e^{t_k}} e^{t_{k+1}}$$

and

$$x_k = l(x_{k+1}) = \frac{x_{k+1}}{e^{t_k}} e^{t_k}$$

Continuing on with the case where there is no system noise/or modeling uncertainties and $l(x_{k+1})$ is invertible, from Equations (3.6) and (3.15) we have

$$\begin{aligned} p(x_{k+1} | \bar{y}_{k+1}) &= \frac{\int p(y_{k+1} | x_{k+1}) \delta(l(x_{k+1}) - x_k) p(x_k | \bar{y}_k) dx_k}{\iint p(y_{k+1} | x_{k+1}) \delta(l(x_{k+1}) - x_k) p(x_k | \bar{y}_k) dx_k dx_{k+1}} \\ &= \frac{p(y_{k+1} | x_{k+1}) p(l(x_{k+1}) | \bar{y}_k)}{\int p(y_{k+1} | x_{k+1}) p(l(x_{k+1}) | \bar{y}_k) dx_{k+1}} \end{aligned} \quad (3.17)$$

Equation (3.6) is the continuous form of the Bayesian estimator. The recursive scheme starts with the initial distribution $p(x_0 | y_0)$. The conditional probability distribution for the state variables at the following time steps, given the measurement from initial time step till present time step, are calculated recursively from Equation (3.6). If the Equation (3.1) has no closed form solution or the closed form solution is not invertible (i.e. $l(x_{k+1})$ cannot be obtained), Equation (3.6) needs to be evaluated numerically. The next Section 3.2 shows how such a numerical evaluation can be performed using the CCMT technique.

3.2 Derivation of the Discrete Form of the Bayesian Estimator

The CCMT describes the dynamic system evolution in terms of probability of transitions between user specified variable magnitude intervals or cells in the system parameter/state space during user specified time intervals $k\tau \leq t \leq (k+1)\tau$ ($k = 0, 1, 2, \dots$)

(see Section 2.1 in Chapter 2). In order to derive the discrete form of Equation (3.6) using the CCMT technique, let $\mathbf{j} = \{j_1, j_2, \dots, j_l, \dots, j_L\}$ indicate the cell location in the discretized state space (see Equation (2.2)). Integrating both sides of Equation (3.6) over the possible cells j_{k+1} where the system might be in at time $t = (k+1)\tau$, we get

$$p(j_{k+1}|\bar{y}_{k+1}) = \frac{\int d\mathbf{x}_{k+1} \int d\mathbf{x}_k p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k)}{\sum_{j_{k+1}} \int d\mathbf{x}_{k+1} \int d\mathbf{x}_k p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k)}. \quad (3.18)$$

In Equation (3.18)

$$p(j_{k+1}|\bar{y}_{k+1}) = \int_{j_{k+1}} p(\mathbf{x}_{k+1}|\bar{y}_{k+1}) d\mathbf{x}_{k+1} \quad (3.19)$$

is the probability that the system is in cell j_{k+1} at time $(k+1)\tau$. From Equation (3.19)

$$p(j_k|\bar{y}_k) = \int_{j_k} p(\mathbf{x}_k|\bar{y}_k) d\mathbf{x}_k$$

and we can write

$$\begin{aligned} & \int p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k \\ &= \sum_{j_k} \frac{\int_{j_k} p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k}{\int_{j_k} p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k} p(j_k | \bar{y}_k) \end{aligned} \quad (3.20)$$

Substituting Equation (3.20) into Equation (3.18) we get

$$\begin{aligned}
p(j_{k+1}|\bar{y}_{k+1}) &= \frac{\int d\mathbf{x}_{k+1} \int d\mathbf{x}_k p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k)}{\sum_{j_{k+1}} \int d\mathbf{x}_{k+1} \int d\mathbf{x}_k p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k)} \\
&= \frac{\int d\mathbf{x}_{k+1} \sum_{j_k} \frac{\int p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k}{\int p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k} p(j_k | \bar{y}_k) d\mathbf{x}_{k+1}}{\sum_{j_{k+1}} \int d\mathbf{x}_{k+1} \sum_{j_k} \frac{\int p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k}{\int p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k} p(j_k | \bar{y}_k) d\mathbf{x}_{k+1}} \\
&= \frac{\sum_{j_k} \iint p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) \frac{p(\mathbf{x}_k | \bar{y}_k)}{\int p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k} d\mathbf{x}_k d\mathbf{x}_{k+1} p(j_k | \bar{y}_k)}{\sum_{j_{k+1}} \sum_{j_k} \iint p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) \frac{p(\mathbf{x}_k | \bar{y}_k)}{\int p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k} d\mathbf{x}_k d\mathbf{x}_{k+1} p(j_k | \bar{y}_k)} \quad (3.21)
\end{aligned}$$

Define the cell-to-cell transition probability as

$$g(j_{k+1}|j_k) = \iint p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) \frac{p(\mathbf{x}_k | \bar{y}_k)}{\int p(\mathbf{x}_k | \bar{y}_k) d\mathbf{x}_k} d\mathbf{x}_k d\mathbf{x}_{k+1} \quad (3.22)$$

The $g(j_{k+1}|j_k)$ does not include y_{k+1} and $\bar{y}_k$ as arguments for simplicity and also because the y_{k+1} and $\bar{y}_k$ are known measured data points whereas j_{k+1} is a variable.

Then Equation (3.21) yields the discrete counterpart of Equation (3.6) as

$$p(j_{k+1}|\bar{y}_{k+1}) = \frac{\sum_{j_k} g(j_{k+1}|j_k) p(j_k | \bar{y}_k)}{\sum_{j_{k+1}} \sum_{j_k} g(j_{k+1}|j_k) p(j_k | \bar{y}_k)} \quad (3.23)$$

with $g(j_{k+1}|j_k)$ as defined by Equation (3.22).

Equations (3.22) and (3.23) constitute the extended DSD algorithm. The cell-to-cell transition probabilities (Equation (3.22)) can be calculated numerically from

$$\begin{aligned}
g(j_{k+1}|j_k) &= \iint_{j_{k+1}, j_k} f_w(y_{k+1} - h(x_{k+1})) f_{\Delta B}(x_{k+1} - \tilde{x}(x_k)) \frac{p(x_k|\bar{y}_k)}{\int_{j_k} p(x_k|\bar{y}_k) dx_k} dx_k dx_{k+1} \\
&= \sum_r \sum_q f_w(y_{k+1} - h(\hat{x}'_{k+1})) f_{\Delta B}(\hat{x}'_{k+1} - \tilde{x}(\hat{x}_k^q)) \frac{p(\hat{x}_k^q|\bar{y}_k)}{\int_{j_k} p(x_k|\bar{y}_k) dx_k} \Delta \hat{x}_k^q \Delta \hat{x}'_{k+1}
\end{aligned} \tag{3.24}$$

where $f_{\Delta B}$ and f_w are, respectively, the user specified pdfs for the system noise (e.g. see Equation (3.12)) and measurement noise (e.g. see Equation (3.13)). The $\hat{x}_k^q$ are points selected in the cell j_k with $\Delta \hat{x}_k^q$ denoting a small volume which surrounds $\hat{x}_k^q$. Similarly the $\hat{x}'_{k+1}$ are points selected in the cell j_{k+1} with $\Delta \hat{x}'_{k+1}$ denoting the small volume which surrounds $\hat{x}'_{k+1}$. The $\tilde{x}(\hat{x}_k^q)$ can be determined by any numerical integration technique. At this point it should be mentioned that the equations describing the evolution of the system do not have to be differential equations as assumed in Equation (3.1). All that is needed is a rule $\tilde{x}(x_k)$ that yields the system location in the state space at time $(k+1)\tau$ given its location at time $k\tau$, such as a set of difference equations, neural nets, response surfaces, tabular data or a set of algebraic equations.

In the case that there is no modeling uncertainty, Equation (3.24) can be simplified using Equation (3.15). In this situation

$$f_{\Delta B}(x_{k+1} - \tilde{x}(x_k)) = \delta(x_{k+1} - \tilde{x}(x_k))$$

and we get

$$\begin{aligned}
g(j_{k+1}|j_k) &= \iint_{j_{k+1}, j_k} p(y_{k+1} | \mathbf{x}_{k+1}) \delta(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}(\mathbf{x}_k)) \frac{p(\mathbf{x}_k | \bar{\mathbf{y}}_k)}{\int_{j_k} p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k} d\mathbf{x}_k d\mathbf{x}_{k+1} \\
&= \int_{j_k} f_w(y_{k+1} - h(\tilde{\mathbf{x}}(\mathbf{x}_k))) \frac{p(\mathbf{x}_k | \bar{\mathbf{y}}_k)}{\int_{j_k} p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k} e_{k+1}(\tilde{\mathbf{x}}(\mathbf{x}_k)) d\mathbf{x}_k \\
&\approx \sum_i f_w(y_{k+1} - h(\tilde{\mathbf{x}}(\hat{\mathbf{x}}_k^i))) \frac{p(\hat{\mathbf{x}}_k^i | \bar{\mathbf{y}}_k)}{\int_{j_k} p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k} e_{k+1}(\tilde{\mathbf{x}}(\hat{\mathbf{x}}_k^i)) \Delta \hat{\mathbf{x}}_{k+1}
\end{aligned} \tag{3.25}$$

where $e_{k+1}(\tilde{\mathbf{x}}(\mathbf{x}_k))$ is defined as

$$e_{k+1}(\tilde{\mathbf{x}}(\mathbf{x}_k)) = \begin{cases} 0 & \tilde{\mathbf{x}}(\mathbf{x}_k) \notin j_{k+1} \\ 1 & \tilde{\mathbf{x}}(\mathbf{x}_k) \in j_{k+1} \end{cases} \tag{3.26}$$

and $\hat{\mathbf{x}}_k^i$ ($i = 1, \dots, N$) are points selected in the cell j_k and $\Delta \hat{\mathbf{x}}_k^i$ is the small volume which surrounds $\hat{\mathbf{x}}_k^i$. Further simplification of Equation (3.25) can be made by assuming that the probability $p(\hat{\mathbf{x}}_k | \bar{\mathbf{y}}_k)$ is constant over the cell j_k and selecting the points $\hat{\mathbf{x}}_k$ equally spaced inside the cell, then

$$\frac{p(\hat{\mathbf{x}}_k | \bar{\mathbf{y}}_k)}{\int_{j_k} p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k} = \frac{p(\hat{\mathbf{x}}_k | \bar{\mathbf{y}}_k)}{N p(\hat{\mathbf{x}}_k | \bar{\mathbf{y}}_k) \Delta \hat{\mathbf{x}}_k} = \frac{1}{N \Delta \hat{\mathbf{x}}_k} = \frac{1}{V_k} \tag{3.27}$$

where V_k is the volume of cell j_k (i.e. the product $\Delta_{1,j_1} \Delta_{2,j_2} \dots \Delta_{L,j_L}$). Inserting Equation

(3.27) into Equation (3.25), we get

$$\begin{aligned}
g(j_{k+1}|j_k) &= \int_{j_k} f_w(y_{k+1} - h(\tilde{x}(x_k))) \frac{p(x_k|\bar{y}_k)}{\int_{j_k} p(x_k|\bar{y}_k) dx_k} e_{k+1}(\tilde{x}(x_k)) dx_k \\
&\approx \sum_i f_w(y_{k+1} - h(\tilde{x}(\hat{x}_k^i))) \frac{p(\hat{x}_k^i|\bar{y}_k)}{Np(\hat{x}_k^i|\bar{y}_k)\Delta\hat{x}_k} e_{k+1}(\tilde{x}(\hat{x}_k^i)) \Delta\hat{x}_k \\
&= \sum_i \frac{1}{N} f_w(y_{k+1} - h(\tilde{x}(\hat{x}_k^i))) e_{k+1}(\tilde{x}(\hat{x}_k^i))
\end{aligned} \tag{3.28}$$

The extended DSD algorithm defined by Equations (3.22) and (3.23) is similar to the recursive estimation rule reported in [1] that the software package DSD is based on. However, the definition of the cell-to-cell transition probabilities through Equation (3.22) differs from the rule in [1] in the following respects:

1. The rule in [1] assumes there is no model uncertainty (except possibly small random fluctuations in system parameters that are contained within the cells). Equation (3.22) represents the model uncertainties through the term $p(x_{k+1} | x_k)$.
2. The rule in [1] assumes that only the state variables are directly measured, i.e.

$$\begin{aligned}
y_k &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix}_k = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \end{bmatrix}_k + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_i \end{bmatrix}_k \quad k=1,2,\dots \\
\Rightarrow y_k &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \bar{x} \end{bmatrix}_k + w_k
\end{aligned}$$

where $\hat{x} = [x_1, x_2, \dots, x_i]$ is the vector whose elements are the monitored state variables and $\bar{x} = [x_{i+1}, x_{i+2}, \dots, x_L]$ is the vector whose elements are the unmonitored state

variables. Hence Equation (3.22) allows arbitrary observers whereas [1] is restricted to the directly measured components of the state variable vector.

3. The rule in [1] assumes that $\mathbf{w}_k$ is uniformly distributed, i.e

$$f_w(\mathbf{w}_k) = \begin{cases} 1 & |\mathbf{w}_k| \leq \lambda \\ 0 & \text{otherwise} \end{cases} \quad (3.29)$$

4. In [1], two sets of cells $\hat{j} = \{j_1 j_2 \dots j_i\}$ and $\bar{j} = \{j_{i+1} j_{i+2} \dots j_L\}$ are defined in the monitored and unmonitored state variable space (i.e. $j = \{\hat{j} \bar{j}\}$), respectively. The cells $\hat{j}$ contain the measured data point y as well as the measurement noise (i.e. $\hat{j}_k = \{\hat{\mathbf{x}}_k : |\hat{\mathbf{x}}_k - y_k| \leq \lambda\}$). The cell-to-cell transition probabilities are calculated from

$$g(j_{k+1}|j_k) \equiv g(\hat{j}_{k+1}, \bar{j}_{k+1} | \hat{j}_k, \bar{j}_k) = \int_{j_k} \frac{d\mathbf{x}_k}{V_k} e_{k+1}(\tilde{\mathbf{x}}(\mathbf{x}_k)) \quad (3.30)$$

where j_k is the cell containing $\mathbf{x}_k$ and V_k is the volume of j_k (i.e. the product $\Delta_{1,j_1} \Delta_{2,j_2} \dots \Delta_{L,j_L}$) as defined before.

Now we proceed to show that Equation (3.22) reduces to Equation (3.30) when Differences 1-4 are removed. First, using Equation (3.29), and noticing that $y_{k+1} = \mathbf{x}'_{k+1} + \mathbf{w}_{k+1}$ from Difference 2, we get

$$p(y_{k+1} | \mathbf{x}_{k+1}) = f_w(y_{k+1} - \mathbf{x}'_{k+1}) = \begin{cases} 1 & |\mathbf{x}'_{k+1} - y_{k+1}| \leq \lambda \\ 0 & \text{else} \end{cases} = \begin{cases} 1 & \mathbf{x}'_{k+1} \in \hat{j}_{k+1} \\ 0 & \mathbf{x}'_{k+1} \notin \hat{j}_{k+1} \end{cases} \quad (3.31)$$

Recalling that $j = \{\hat{j} \bar{j}\}$ and $\mathbf{x} = \{\hat{\mathbf{x}} \bar{\mathbf{x}}\}$, the transition probability $g(j_{k+1}|j_k)$ can be written from Equation (3.22) as the following

$$\begin{aligned}
g(j_{k+1}|j_k) &\equiv g(\hat{j}_{k+1}, \bar{j}_{k+1} | \hat{j}_k, \bar{j}_k) \\
&= \iint_{\hat{j}_{k+1}, \bar{j}_{k+1}} \iint_{\hat{j}_k, \bar{j}_k} p(y_{k+1} | \hat{x}_{k+1}, \bar{x}_{k+1}) p(\hat{x}_{k+1}, \bar{x}_{k+1} | \hat{x}_k, \bar{x}_k) \frac{p(\hat{x}_k, \bar{x}_k | \bar{y}_k)}{\int_{j_k} p(x_k | \bar{y}_k) dx_k} d\hat{x}_k d\bar{x}_k d\hat{x}_{k+1} d\bar{x}_{k+1} \\
&= \iint_{\hat{j}_{k+1}, \bar{j}_{k+1}} \iint_{\hat{j}_k, \bar{j}_k} f_w(y_{k+1} - \hat{x}_{k+1}) p(\hat{x}_{k+1}, \bar{x}_{k+1} | \hat{x}_k, \bar{x}_k) \frac{p(\hat{x}_k, \bar{x}_k | \bar{y}_k)}{\int_{j_k} p(x_k | \bar{y}_k) dx_k} d\hat{x}_k d\bar{x}_k d\hat{x}_{k+1} d\bar{x}_{k+1}
\end{aligned} \tag{3.32}$$

where $\hat{x}_k$ and $\bar{x}_k$ denote the location of the monitored and unmonitored system variables in their respective spaces. With Equation (3.31), Equation (3.32) becomes

$$\begin{aligned}
g(j_{k+1}|j_k) &\equiv g(\hat{j}_{k+1}, \bar{j}_{k+1} | \hat{j}_k, \bar{j}_k) \\
&= \iint_{\hat{j}_{k+1}, \bar{j}_{k+1}} \iint_{\hat{j}_k, \bar{j}_k} p(\hat{x}_{k+1}, \bar{x}_{k+1} | \hat{x}_k, \bar{x}_k) \frac{p(\hat{x}_k, \bar{x}_k | \bar{y}_k)}{\int_{j_k} p(\hat{x}_k, \bar{x}_k | \bar{y}_k) d\hat{x}_k d\bar{x}_k} d\hat{x}_k d\bar{x}_k d\hat{x}_{k+1} d\bar{x}_{k+1} \\
&= \iint_{j_{k+1}} p(x_{k+1} | x_k) \frac{p(x_k | \bar{y}_k)}{\int_{j_k} p(x_k | \bar{y}_k) dx_k} dx_{k+1} dx_k
\end{aligned} \tag{3.33}$$

where $j_{k+1} = \{\hat{j}_{k+1}, \bar{j}_{k+1}\}$ and $j_k = \{\hat{j}_k, \bar{j}_k\}$. If there is no modeling uncertainty (i.e. Difference 1 above), Equation (3.15) holds and Equation (3.33) becomes

$$\begin{aligned}
g(\hat{j}_{k+1}, \bar{j}_{k+1} | \hat{j}_k, \bar{j}_k) &\equiv g(j_{k+1} | j_k) \\
&= \iint_{j_{k+1}, j_k} p(x_{k+1} | x_k) \frac{p(x_k | \bar{y}_k)}{\int_{j_k} p(x_k | \bar{y}_k) dx_k} dx_{k+1} dx_k \\
&= \iint_{j_{k+1}, j_k} \delta(x_{k+1} - \tilde{x}(x_k)) \frac{p(x_k | \bar{y}_k)}{\int_{j_k} p(x_k | \bar{y}_k) dx_k} dx_{k+1} dx_k \\
&= \int_{j_k} \frac{p(x_k | \bar{y}_k)}{\int_{j_k} p(x_k | \bar{y}_k) dx_k} e_{k+1}(\tilde{x}(x_k)) dx_k
\end{aligned} \tag{3.34}$$

where $e_{k+1}(\tilde{\mathbf{x}}(\mathbf{x}_k))$ is as defined in Equation (3.26). Finally, by assuming that the system location is uniformly distributed over the cell j_k and using Equation (3.27)

$$\frac{p(\mathbf{x}_k|\bar{\mathbf{y}}_k)}{\int_{j_k} p(\mathbf{x}_k|\bar{\mathbf{y}}_k) d\mathbf{x}_k} = \frac{1}{V_k} \quad (3.35)$$

and we get

$$\begin{aligned} & g(\hat{j}_{k+1}, \bar{j}_{k+1} | \hat{j}_k, \bar{j}_k) \\ &= \frac{1}{V_k} \int_{j_k} e_{k+1}(\tilde{\mathbf{x}}(\mathbf{x}_k)) d\mathbf{x}_k \end{aligned} \quad (3.36)$$

which is the same as Equation(3.30).

3.3 Example Application

This section illustrates the implementation of the DSD algorithm using the well-known van der Pol oscillator. The van der Pol oscillator can be used to model a variety of mechanical systems including those used in nuclear reactors, such as vibration absorbers controlling the vibration in the fuel bundles. This section compares the extended DSD algorithm defined by Equation (3.22) and (3.23) to the rule reported in [1] and presents the results.

3.3.1 System Description

The dynamics of the van der Pol oscillator is described by the following set of first order differential equations

$$\begin{aligned}\frac{dx_1}{dt} &= \alpha(1 - x_2^2)x_1 - x_2 \\ \frac{dx_2}{dt} &= x_1\end{aligned}\tag{3.37}$$

where x_1 and x_2 are dynamic variables and α is the system parameter (all dimensionless). It is assumed that the measurement is corrupted by noise, i.e.

$$y_k = h(x_k) + w_k\tag{3.38}$$

where $y_k = [y_1, y_2, \dots, y_m]_{t=t_k}$ is the measurement, $x_k = [x_1(k\tau), x_2(k\tau)]$ is the system state vector at time $t = k\tau$ and w_k is the measurement noise with the probability distribution $f_w(w_k)$. We will assume that the w_k are statistically independent for all k and the pdf of the noise for each k is the same. Under the assumption that the parameter α does not change, Equation (3.37) can be rewritten as the following

$$\begin{aligned}\frac{dx_1}{dt} &= \alpha(1 - x_2^2)x_1 - x_2 \\ \frac{dx_2}{dt} &= x_1 \\ \frac{d\alpha}{dt} &= 0\end{aligned}\tag{3.39}$$

Three cases will be considered in this section

- I. Given state variables directly measured $y_k = x_k + w_k$ and the measurement noise to be white Gaussian noise, estimate the true value of x_1, x_2 and α .
- II. Given state variables directly measured $y_k = x_k + w_k$ and the measurement noise to be uniformly distributed, estimate the true value of x_1, x_2 and α .
- III. Given the observer $y_k = Cx_k + w_k, C = [1 \quad 1]$ and the measurement noise to be white Gaussian noise, estimate the true value of x_1, x_2 and α .

Sections 3.3.2, 3.3.3 and 3.3.4 describe the estimation results for Cases I, II and III, respectively.

3.3.2 Estimation Results With Gaussian Noise (Case I)

Table 3.1 shows the data used in Case I for the simulation of measured values of x_1 and x_2 . The simulated system evolution is shown in Figs.3.1 and 3.2. The results for $x_1(t)$ and $x_2(t)$ were obtained from the integration of Equation (3.37) using a fourth order Runge-Kutta scheme. The results for $y_1(t)$ and $y_2(t)$ were obtained by sampling from a normal distribution with mean 0 and standard deviation 0.1 (see Table 3.1) using a random number generator and algebraically adding the result to those obtained for $x_1(t)$ and $x_2(t)$. Since the noise is assumed to be Gaussian, its magnitude is always larger than the cell size and the procedure described in [1] does not converge. In order to apply the extended DSD algorithm described in Section 3.2, the partitioning data used are shown in Table 3.2.

Since there is no modeling uncertainty, the transition probabilities $g(j_{k+1}|j_k)$ are calculated numerically from Equation (3.28) with

$$h(\tilde{x}(x_k)) = \tilde{x}(x_k)$$

and f_w as given by Equation (3.13). The total number of points $\hat{x}_k^i$ (see Equation (3.28)) selected for the numerical evaluation of the transition probabilities $g(j_{k+1}|j_k)$ is $N = 3 \times 3 \times 3$.

α	Initial Value of x_1	Initial Value of x_2	Time Step τ	Mean of the noise	Standard Deviation of the noise
1.0	0.5	0.5	0.1	0	0.1

Table 3.1: Parameters Used for Simulation of System Evolution for Case I

Unknown Variables/Parameter	Possible Maximum Value	Possible Minimum Value	Number of Cells	Number of Points $\hat{x}_k^i, i = 1, \dots, N$ (See Equation (3.28))
x_1	3	-3	103	3
x_2	3	-3	103	3
α	3.5	-0.5	69	3

Table 3.2: Partitioning Data Used for Case I & II

Fig. 3.1: Simulated $x_1(t)$ and $y_1(t)$ for Case I

Fig. 3.2: Simulated $x_2(t)$ and $y_2(t)$ for Case I

The estimated 99.99% credibility intervals as a function of time for x_1, x_2 and α are shown, respectively, in Figs. 3.3 3.4 and 3.5. From Figs. 3.3 and 3.4, we can see that the estimated intervals for x_1 and x_2 contain the measured data y_1 and y_2 at all time points. Since the estimation process cross-correlates the measured data in view of the given pdfs of v and w_k (i.e. through $p(x_{k+1} | x_k)$ and $p(y_{k+1} | x_{k+1})$ in Equation (3.22), respectively) and the system model (i.e. through $p(x_k | \bar{y}_k)$) in the determination of $g(j_{k+1} | j_k)$, Figs. 3.3 and 3.4 imply that y_1 and y_2 are within the anticipated uncertainty margins. While the estimated interval for α is still fairly large after 20 time steps (see Fig.3.5), the mean of the estimated pdfs for x_1, x_2 are close to their true values at all times as shown in Figs.3.6 and 3.7, respectively. The expected value of α converges to its true value at 1.0 (see Table 1) after 30 time steps (see Fig.3.8). Figures 3.9 and 3.10 show the ratio of the posterior standard deviation over the noise standard deviation ($\sigma_{estimate} / \sigma_{noise}$). Together with Figs. 3.6 and 3.7 which show that the posterior mean converges to the true state variable value, these figures demonstrate that this new technique (the extended DSD algorithm) can reduce the uncertainty on the measurements by a factor of $\sigma_{estimate} / \sigma_{noise} = 37\%$ for x_1 and $\sigma_{estimate} / \sigma_{noise} = 52\%$ for x_2 at the time $t = 80\tau$.

Figure 3.11 shows the evolution of the posterior pdf $p(\alpha | \bar{y}_k)$ for Case I at time $t = \tau, 20\tau, 40\tau, 80\tau$, using $\tau = 0.1$ and the partitioning system used is shown in Table 3.2. The true parameter value is $\alpha = 1$ (see Table 3.1). At the first time step, the probability or the credibility for parameter α to be in the correct cell (i.e. 1-[3.5-(-

$0.5)/(2*69) \leq \alpha \leq 1+[3.5-(-0.5)]/(2*69)$ from the data in Table 3.1) is about 1.5% (Fig. 3.11 (a)). It improves to 7% at $t = 20\tau$ (Fig.3.11 (b)), to 30% at $t = 40\tau$ (Fig. 3.11 (c)) and finally increases to 43% at $t = 80\tau$ (Fig. 3.11 (d)). From another point of view, at the first time step, Fig. 3.11(a) estimates the 99.99% credibility interval as $-0.5 \leq \alpha \leq 3.5$ with almost no improvement from the original guess of the pdf of α . The estimated credibility interval reduces to $0.4855 \leq \alpha \leq 1.4710$ in Fig.3.11(c) and finally to $0.7174 \leq \alpha \leq 1.2971$ in Fig. 3.11(d).

Fig. 3.3: Estimated Interval of x_1 as a Function of Time for Case I
(solid line indicates the measured value of y_1)

Fig. 3.4: Estimated Interval of x_2 as a Function of Time for Case I
(solid line indicates the measured value of y_2)

Fig. 3.5: Estimated Interval of α as a Function of Time for Case I
(solid line indicates the true value of α)

Fig. 3.6: Mean Value of the Estimated Probability Distribution Function of x_1 for Case I

Fig. 3.7: Mean Value of the Estimated Probability Distribution Function of x_2 for Case I

Fig. 3.8: Mean Value of the Estimated Probability Distribution Function of α for Case I
(solid line indicates the true value of α)

Fig. 3.9: The Ratio of the Estimated Posterior Standard Deviation to the Standard Deviation of the Noise for x_1

Fig. 3.10: The Ratio of the Estimated Posterior Standard Deviation and the Standard Deviation of the Noise for x_2

(a)

(b)

(c)

(d)

Fig. 3.11: The Evolution of $p(\alpha|\bar{y}_k)$ for the van der Pol Oscillator for Case I:

a) $t = \tau$, b) $t = 20\tau$, c) $t = 40\tau$, d) $t = 80\tau$

3.3.3 Estimation Results With Uniform Noise (Case II)

Table 3.3 shows the data used in Case II for the simulation of measured values of x_1 and x_2 . The simulated system evolution is shown in Figs.3.12 and 3.13. The results for $x_1(t)$ and $x_2(t)$ were obtained from the integration of Equation (3.37) using a fourth order Runge-Kutta scheme. The results for $y_1(t)$ and $y_2(t)$ were obtained by sampling from a uniform distribution with mean 0 and uncertainty level $\lambda = 0.3$ (see Equation (3.29)) using a random number generator and algebraically adding the result to those obtained for $x_1(t)$ and $x_2(t)$.

α	Initial Value of x_1	Initial Value of x_2	Time Step τ	Mean of The Noise	Uncertainty Level λ (See Equation (3.29))
1.0	0.5	0.5	0.1	0	0.3

Table 3.3: Parameters Used for Simulation of System Evolution for Case II

The unknown parameter α is estimated by two methods: a) the extended DSD method (i.e. The transition probabilities $g(j_{k+1}|j_k)$ are calculated using Equation (3.28)), and, b) the previous DSD algorithm reported in [1]. Both methods use the same partitioning data listed in Table 3.2. The evolution of the estimated intervals for α , x_1 , x_2 through the extended DSD algorithm are shown in Figs. 3.14, 3.15 and 3.16, respectively, and the corresponding expected values are shown in Figs.3.17, 3.18 and 3.19. These figures show that again the measured data are within the expected

uncertainty margins (i.e. variations in $x_1(t)$ and $x_2(t)$ are within the estimated intervals) and convergence is rapid for α .

Figure 3.20 shows the estimation results using the DSD algorithm in [1] with same partitioning data listed in Table 3.2. Comparison of the estimated intervals in Fig.3.14 and Fig.3.20 shows better estimation results are obtained by using the extended algorithm than the previous DSD algorithm. The estimation process converges faster and provides smaller credibility intervals for the unknown parameter.

Figures 3.21 and 3.22 show the measured data and 100% credibility intervals at each time step and indicate that the credibility intervals contain both the actual values of $x_1(t)$ and $x_2(t)$ and the measured $y_1(t)$ and $y_2(t)$ as they should. The credibility interval can be determined directly from the measured data and the measurement uncertainty level from Equation (3.31) as shown below:

Since all the state variables are monitored $x_k = \hat{x}_k$ and Equation (3.31) can be rewritten as

$$p(y_k | x_k) = f_w(y_k - \hat{x}_k) = f_w(y_k - x_k) = \begin{cases} 1 & |x_k - y_k| \leq \lambda \\ 0 & \text{else} \end{cases}$$

Then

$$\begin{aligned} P(|x_k - y_k| \leq \lambda) &= 1 \\ \Rightarrow P(y_k - \lambda \leq x_k \leq y_k + \lambda) &= 1 \\ \Rightarrow y_k - \lambda \leq x_k \leq y_k + \lambda &\text{ with 100\% credibility} \end{aligned} \tag{3.40}$$

where P denotes probability and λ is the measurement uncertainty level with $\lambda = 0.3$ from Table 3.3.

Fig. 3.12: Simulated $x_1(t)$ and $y_1(t)$ for Case II

Fig. 3.13: Simulated $x_2(t)$ and $y_2(t)$ for Case II

Fig. 3.14: Estimated Interval of α as a Function of Time Using the Extended DSD Algorithm for Case II (solid line indicates the true value of α)

Fig.3.15: Estimated Interval of x_1 as Function of Time Using the Extended DSD Algorithm for Case II (solid line indicates the measured value of y_1)

Fig. 3.16: Estimated Interval of x_2 as a Function of Time Using the Extended DSD Algorithm for Case II (solid line indicates the measured value of y_2)

Fig.3.17: Mean Value of the Estimated Probability Distribution Function of α for Case II (solid line indicates the true value of α)

Fig.3.18: Mean Value of the Estimated Probability Distribution Function of x_1 for Case II

Fig.3.19: Mean Value of the Estimated Probability Distribution Function of x_2 for Case II

Fig. 3.20: Estimated Interval of α as a Function of Time Using the Rules in [1] for Case II (solid line indicates the true value of α)

Fig.3.21: Estimated Interval of x_1 as a Function of Time for Case II Using Equation (3.40) (solid line indicates the true value, asterisks indicate the measured data and bars indicate the uncertainty in the data)

Fig. 3.22: Estimated Interval of x_2 as a Function of Time for Case II Using Equation (3.40) (solid line indicates the true value, asterisks indicate the measured data and bars indicate the uncertainty in the data)

Figure 3.23 shows the estimated pdf for parameter α . Using the DSD algorithm in [1], Fig.3.23(a) estimates α as $0.2536 \leq \alpha \leq 2.2246$ at $t = 80\tau$. Figure 3.23(b) shows that the new procedure (extended DSD algorithm) reduces the estimated interval to $0.7754 \leq \alpha \leq 1.2391$ at the same time step to an interval which is $1 - (1.2391 - 0.7754) / (2.2246 - 0.2536) = 0.76$ (or 76%) smaller.

Further comparison made in Fig. 3.24 illustrates the signal validation capability of the new procedure. Figure 3.24 shows the estimated probability distributions of the state variables at time $t = 50\tau$ using the extended DSD algorithm developed in Section 3.2. In Figs.3.24(a) and 3.24(b), the state variables are, respectively, estimated as

$1.1942 \leq x_1 \leq 1.4272$ and $-0.9029 \leq x_2 \leq -0.6117$ with 100% credibility. These estimated intervals are, respectively, $1 - (1.4272 - 1.1942)/0.6 = 0.61$ (or 61%) and $1 - [(-0.6117) - (-0.9029)]/0.6 = 0.51$ (or 51%) smaller than the measurement uncertainty level which is $2\lambda = 0.6$ (See Equation (3.40)). These results imply that the extended DSD algorithm (i.e. Equations (3.22) and (3.23)) can be used to reduce the uncertainty on the measurements in view of noisy data.

(a)

(b)

Fig. 3.23: Estimated pdf of α for Case II at $t = 80\tau$

(a) DSD Algorithm Results

(b) Extended DSD Algorithm Results

(a)

(b)

Fig. 3.24: Estimated pdf for State Variables for Case II at $t = 50\tau$ Using the Extended DSD Algorithm: (a) $p(x_1|\bar{y}_{50})$, (b) $p(x_2|\bar{y}_{50})$

3.3.4 Estimation Results for Case III

The system used for illustration is the van der Pol oscillator with a linear observer

$$y_k = C\mathbf{x}_k + w_k, C = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad (3.41)$$

Table 3.4 shows the data used in Case III for the simulation of measured value of y . The simulation results for $y(t)$ are shown in Fig. 3.25. The system evolution (i.e. $x_1(t), x_2(t)$) is simulated by integrating Equation (3.37) using a 4th order Runge-Kutta numerical integrator with the initial values and parameter value listed in Table 3.4. Measured value of y is obtained by sampling from a normal distribution with mean 0 and standard deviation 0.3 (see Table 3.4) using a random number generator and algebraically adding x_1, x_2 and the obtained random number together at each time step.

α	Initial Value of x_1	Initial Value of x_2	Time Step	Mean of the noise	Standard Deviation of the noise
1.0	0.5	0.5	0.1	0	0.3

Table 3.4: Parameters Used for Simulation of System Evolution for Case III

Fig. 3.25: Simulated $y(t)$ for Case III

In order to apply the method introduced in Section 3.2, the partitioning data used is listed in Table 3.5 and the transition probabilities $g(j_{k+1}|j_k)$ are calculated by Equation (3.28) with

$$h(\tilde{x}(x_k)) = C\tilde{x}(x_k)$$

where $C = [1 \quad 1]$ and f_w as given by Equation (3.13). The total number of points $\tilde{x}_k^i$ (see Equation (3.28)) selected for the numerical evaluation of the transition probabilities $g(j_{k+1}|j_k)$ is $N = 3 \times 3 \times 3$.

Unknown Variables/Parameter	Possible Maximum Value	Possible Minimum Value	Number of Cells	Number of points $\hat{\mathbf{x}}_k^i, i = 1, \dots, N$ (see Equation (3.28))
x_1	3	-3	103	3
x_2	3	-3	103	3
α	3.5	-0.5	79	3

Table 3.5: Partitioning Data Used for Case III

Figures 3.26, 3.27, 3.28 show the estimated 99.99% credibility interval as a function of time for the unknown variables and parameter. With a linear observer as shown in Equation (3.41), the unknown state variables are estimated within a small interval. It only needs several time steps for the estimation process to converge. On the other hand, the parameter α is more difficult to estimate in this case (compared with Case I, the credibility intervals are larger) since that the variance of the measurement noise is larger ($\sigma = 0.3$ instead of 0.1 in Case I) and less information is available (instead of measuring x_1 and x_2 in Case I, only $y = x_1 + x_2$ is measured). But the mean of the estimated posterior distribution at each time step provides a close estimation of both the parameter value and the unknown state variables (see Figs. 3.29,3.30,3.31).

Fig. 3.26: Estimated Interval of x_1 as a Function of Time for Case III
(solid line indicates the true value of x_1)

Fig. 3.27: Estimated Interval of x_2 as a Function of Time for Case III
(solid line indicates the true value of x_2)

Fig. 3.28: Estimated Interval of α as a Function of Time for Case III
(solid line indicates the true value of α)

Fig. 3.29: Mean Value of the Estimated Probability Distribution Function of x_1 for
Case III

Fig. 3.30: Mean Value of the Estimated Probability Distribution Function of x_2 for Case III

Fig. 3.31: Mean Value of the Estimated Probability Distribution Function of α for Case III (solid line indicates the true value of α)

CHAPTER 4

RELATIONSHIP BETWEEN DSD, GENERALIZED MAXIMUM LIKELIHOOD ESTIMATOR AND THE LEAST SQUARES ESTIMATOR

Chapter 4 investigates the relationship between the DSD, Generalized Maximum Likelihood Estimator (GMLE) and the Least Squares Estimator (LSE). Theoretical analysis shows that when only unknown constant parameters are being estimated and the measurement noise is white Gaussian noise, these three estimators are equivalent.

In this chapter, we will show first that DSD is equivalent to GMLE (Section 4.1) and then prove that GMLE is equivalent to LSE (Section 4.2) and finally, a nuclear application will be used to illustrate the equivalency between these three estimators.

4.1 Relationship Between DSD and GMLE

Considering the case that only unknown parameters are being estimated and the parameters are not functions of time, the system can be described by

$$\begin{aligned}\dot{\theta} &= 0 \\ y_k &= h(\theta_k) + w_k \quad (k = 1, 2, \dots)\end{aligned}\tag{4.1}$$

where

θ : unknown parameter vector

- y_k : M -dimensional vector whose elements y_k are the measured data at time step $k\tau$ ($k = 0,1,2,\dots$)
- θ_k : parameter vector at time step $k\tau$ ($k = 0,1,2,\dots$)
- w_k : measurement noise, assuming w_k are statistically independent for all k and its probability distribution f_w is not a function of k .
- h : M -dimensional vector whose elements are known nonlinear functions

For Equation (4.1), the recursive Bayesian estimator given by Equation (3.6) can be written as

$$\begin{aligned}
 p(\theta_{k+1} | \bar{y}_{k+1}) &= \frac{p(\theta_{k+1}, y_{k+1} | \bar{y}_k)}{\int p(\theta_{k+1}, y_{k+1} | \bar{y}_k) d\theta_{k+1}} \\
 &= \frac{\int p(y_{k+1} | \theta_{k+1}) p(\theta_{k+1} | \theta_k) p(\theta_k | \bar{y}_k) d\theta_k}{\iint p(y_{k+1} | \theta_{k+1}) p(\theta_{k+1} | \theta_k) p(\theta_k | \bar{y}_k) d\theta_k d\theta_{k+1}}
 \end{aligned} \tag{4.2}$$

Since the parameter is assumed to be constant

$$\theta_{k+1} = \theta_k = \theta \tag{4.3}$$

and subsequently

$$p(\theta_{k+1} | \theta_k) = \delta(\theta_{k+1} - \theta_k) \tag{4.4}$$

Substituting Equation (4.4) into Equation (4.2) we get

$$\begin{aligned}
p(\theta_{k+1} | \bar{y}_{k+1}) &= \frac{p(\theta_{k+1}, y_{k+1} | \bar{y}_k)}{\int p(\theta_{k+1}, y_{k+1} | \bar{y}_k) d\theta_{k+1}} \\
&= \frac{\int p(y_{k+1} | \theta_{k+1}) \delta(\theta_{k+1} - \theta_k) p(\theta_k | \bar{y}_k) d\theta_k}{\iint p(y_{k+1} | \theta_{k+1}) \delta(\theta_{k+1} - \theta_k) p(\theta_k | \bar{y}_k) d\theta_k d\theta_{k+1}} \\
&= \frac{p(y_{k+1} | \theta_{k+1}) p(\theta_{k+1} | \bar{y}_k)}{\int p(y_{k+1} | \theta_{k+1}) p(\theta_{k+1} | \bar{y}_k) d\theta_{k+1}}
\end{aligned} \tag{4.5}$$

Considering Equation (4.3), Equation (4.5) can be rewritten as

$$p(\theta | \bar{y}_{k+1}) = \frac{p(y_{k+1} | \theta) p(\theta | \bar{y}_k)}{\int p(y_{k+1} | \theta) p(\theta | \bar{y}_k) d\theta} \tag{4.6}$$

Proposition 4.1

Assume an initial guess of the probability distribution $p(\theta | \bar{y}_0)$ is given. If the posterior pdf at each time step $p(\theta | \bar{y}_k)$ is recursively determined from Equation (4.6), then

maximizing the ratio $\frac{p(\theta | \bar{y}_{k+1})}{p(\theta | \bar{y}_0)}$ will also maximize the likelihood function.

Proof:

We first show that

$$p(\theta | \bar{y}_{k+1}) = \frac{p(y_{k+1} | \theta) p(y_k | \theta) \cdots p(y_1 | \theta) p(\theta | \bar{y}_0)}{\int p(y_{k+1} | \theta) p(y_k | \theta) \cdots p(y_1 | \theta) p(\theta | \bar{y}_0) d\theta} \tag{4.7}$$

At the first time step, from Equation (4.6), we can write

$$p(\theta | \bar{y}_1) = \frac{p(y_1 | \theta) p(\theta | \bar{y}_0)}{\int p(y_1 | \theta) p(\theta | \bar{y}_0) d\theta} \tag{4.8}$$

Substituting Equation (4.8) into the recursive rule for the second time step, we get

$$\begin{aligned}
p(\theta|\bar{y}_2) &= \frac{p(y_2|\theta)p(\theta|\bar{y}_1)}{\int p(y_2|\theta)p(\theta|\bar{y}_1)d\theta} \\
&= \frac{p(y_2|\theta) \frac{p(y_1|\theta)p(\theta|\bar{y}_0)}{\int p(y_1|\theta)p(\theta|\bar{y}_0)d\theta}}{\int p(y_2|\theta) \frac{p(y_1|\theta)p(\theta|\bar{y}_0)}{\int p(y_1|\theta)p(\theta|\bar{y}_0)d\theta} d\theta} \\
&= \frac{p(y_2|\theta)p(y_1|\theta)p(\theta|\bar{y}_0)}{\int p(y_2|\theta)p(y_1|\theta)p(\theta|\bar{y}_0)d\theta}
\end{aligned}$$

Assuming that Equation (4.7) holds for any k , we have

$$p(\theta|\bar{y}_k) = \frac{p(y_k|\theta)p(y_{k-1}|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)}{\int p(y_k|\theta)p(y_{k-1}|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)d\theta} \quad (4.9)$$

Substitute Equation (4.9) into Equation (4.6) to get

$$\begin{aligned}
p(\theta|\bar{y}_{k+1}) &= \frac{p(y_{k+1}|\theta) \frac{p(y_k|\theta)p(y_{k-1}|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)}{\int p(y_k|\theta)p(y_{k-1}|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)d\theta}}{\int p(y_{k+1}|\theta) \frac{p(y_k|\theta)p(y_{k-1}|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)}{\int p(y_k|\theta)p(y_{k-1}|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)d\theta} d\theta} \\
&= \frac{p(y_{k+1}|\theta)p(y_k|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)}{\int p(y_{k+1}|\theta)p(y_k|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)d\theta}
\end{aligned} \quad (4.10)$$

So, Equation (4.7) holds for all time steps by induction.

From Equation (4.7), the ratio $\frac{p(\theta|\bar{y}_{k+1})}{p(\theta|\bar{y}_0)}$ is calculated as

$$\frac{p(\theta|\bar{y}_{k+1})}{p(\theta|\bar{y}_0)} = \frac{p(y_{k+1}|\theta)p(y_k|\theta)\cdots p(y_1|\theta)}{\int p(y_{k+1}|\theta)p(y_k|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)d\theta}$$

Then

$$\max_{\theta} \frac{p(\theta|\bar{y}_{k+1})}{p(\theta|\bar{y}_0)} = \frac{\max_{\theta} \{p(y_{k+1}|\theta)p(y_k|\theta)\cdots p(y_1|\theta)\}}{\int p(y_{k+1}|\theta)p(y_k|\theta)\cdots p(y_1|\theta)p(\theta|\bar{y}_0)d\theta} \quad (4.11)$$

where $p(\theta|\bar{y}_0)$ is the initial guess of the probability distribution function of θ which is defined by the user.

Now we proceed to show that

$\prod_{i=1}^{k+1} p(y_i|\theta) \equiv p(y_{k+1}|\theta)p(y_k|\theta)\cdots p(y_1|\theta) = p(y_{k+1}, y_k, \cdots y_1|\theta)$ is the joint probability

that data $y_{k+1}, y_k, \cdots y_1$ will be observed given parameter θ :

From Equation (4.1) and considering that the parameters are constant (see Equation (4.3)), we get

$$y_k = h(\theta) + w_k \quad (k = 1, 2, \dots) \quad (4.12)$$

From the definition of conditional probability, we can write

$$p(y_k) = \int p(y_k, \beta) d\beta = \int p(y_k|\beta) p(\beta) d\beta \quad (4.13)$$

where β is the dummy variable representing the parameter vector to be estimated.

Since the parameters are constant, then

$$p(\beta) = \delta(\beta - \theta) \quad (4.14)$$

Substituting Equation (4.14) into (4.13) we get

$$p(y_k) = p(y_k|\theta) \quad (4.15)$$

In a similar manner, we can write

$$p(y_{k+1}, y_k, \cdots y_1) = p(y_{k+1}, y_k, \cdots y_1|\theta) \quad (4.16)$$

Recall that w_k are statistically independent for all k and y_k is a function of w_k and θ which are constant parameters. Hence the y_k are also statistically independent which implies

$$p(y_{k+1}, y_k, \dots, y_1) = \prod_{i=1}^{k+1} p(y_i) = \prod_{i=1}^{k+1} p(y_i | \theta) \quad (4.17)$$

Hence,

$$p(y_{k+1}, y_k, \dots, y_1 | \theta) = \prod_{i=1}^{k+1} p(y_i | \theta) \quad (4.18)$$

Equation (4.18) shows that the product of $p(y_i | \theta)$ is the joint probability that data $y_{k+1}, y_k, \dots, y_1$ being observed given parameter θ which is the likelihood function by definition³⁷.

End of Proof

If $p(\theta | \bar{y}_0)$ is the uniform distribution, Equation (4.11) can be simplified as

$$\max_{\theta} \frac{p(\theta | \bar{y}_{k+1})}{\text{contant}} = \max_{\theta} p(\theta | \bar{y}_{k+1}) = \frac{\max_{\theta} \{p(y_{k+1} | \theta) p(y_k | \theta) \dots p(y_1 | \theta)\}}{\int p(y_{k+1} | \theta) p(y_k | \theta) \dots p(y_1 | \theta) p(\theta | \bar{y}_0) d\theta} \quad (4.19)$$

which is a well-known relation used in system identification area³⁸.

From the derivation of DSD in Chapter 3, it is clear that when the cells are defined vanishingly small, DSD will converge to the continuous Bayesian estimator. Hence, when the cell size is chosen small enough, the maximum value of the ratio

$\frac{p(\theta | \bar{y}_{k+1})}{p(\theta | \bar{y}_0)}$ or when $p(\theta | \bar{y}_0)$ is uniform, the mode of the discrete type posterior

probability distribution obtained through DSD will approach to the GMLE estimate of θ .

4.2 Equivalence of GMLE and LSE

Assume that the pdf of w_k in Equation (4.1) is Gaussian for all k , i.e.

$$f_w(w_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{w_i w_i}{2\sigma^2}}, i = 1, 2, \dots, k \quad (4.20)$$

with mean $E(w_i) = 0$ and covariance matrix $Var(w_i) = \sigma^2 I$ where σ is the standard deviation of the noise and I is $M \times M$ identity matrix with M as the dimension of w_i .

Then given θ , the likelihood function $\prod_{i=1}^n p(y_i|\theta)$ can be determined in the following manner:

From Equation (4.1) and recalling that θ is constant (see Equation (4.3)), we get

$$y_k = h(\theta) + w_k \quad (k = 1, 2, \dots) \quad (4.21)$$

Considering Equations (4.20) and (4.21), the conditional probability $p(y_k|\theta)$ can be written as

$$p(y_k|\theta) = f_w(y_k - h(\theta)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_k - h(\theta))(y_k - h(\theta))}{2\sigma^2}} \quad (k = 1, 2, \dots) \quad (4.22)$$

Then, we get

$$\begin{aligned}
\prod_{i=1}^n p(y_i|\theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - h(\theta))(y_i - h(\theta))}{2\sigma^2}} \\
&= \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right]^n \exp \left[-\frac{\sum_{i=1}^n [(y_i - h(\theta))(y_i - h(\theta))]}{2\sigma^2} \right]
\end{aligned} \tag{4.23}$$

Proposition 4.2:

Consider the dynamic system described by Equation (4.1), Assume at each time step, the distribution of w_k is Gaussian. Then, the GMLE estimate of θ also minimize the Least Squares Error criterion.

Proof:

The Generalized Maximum Likelihood estimate of θ is the largest mode of the likelihood function

$$\hat{\theta} = \max_{\theta} \prod_{i=1}^n p(y_i|\theta) \tag{4.24}$$

Under the assumptions stated above, Equation (4.23) holds. Then from Equation (4.23) and (4.24) we get

$$\hat{\theta} = \max_{\theta} \prod_{i=1}^n p(y_i|\theta) = \max_{\theta} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right]^n \exp \left[-\frac{\sum_{i=1}^n [(y_i - h(\theta))(y_i - h(\theta))]}{2\sigma^2} \right] \tag{4.25}$$

Equation (4.25) indicates that the $\hat{\theta}$ which maximizes the likelihood function will minimize the sum of the squares term

$$\sum_{i=1}^n [(y_i - h(\theta))(y_i - h(\theta))].$$

End of Proof

Sections 4.1 and 4.2 show that in the case that only the unknown parameters are being estimated, the continuous Bayesian estimator (Equation 3.6) is equivalent to the GLME estimator, which also minimize the Least Squares Error criterion when the noise is white Gaussian noise. Hence the discrete type Bayesian estimator DSD is also equivalent to the GMLE and LSE under the condition that the cell size is chosen small enough. The equivalence is in the sense of the maximum value of the ratio $\frac{p(\theta|\bar{y}_{k+1})}{p(\theta|\bar{y}_0)}$ (or $p(\theta|\bar{y}_{k+1})$ when $p(\theta|\bar{y}_0)$ is uniform).

4.3 Illustration

In this section, the discrete type Bayesian estimator (the DSD method) will be compared with the GMLE and LSE through a nuclear application. The system description and the estimation results are given in Sections 4.3.1 and 4.3.2, respectively.

4.3.1. System Description

In a nuclear reactor operation, it is critical to meet the basic reactor safety thermal limits, which are fuel melt and fuel cladding failure. These two thermal limits are directly related to the local nuclear energy deposition and the local heat transfer rate. The fuel melt limits are determined by an energy deposition limit, while critical heat flux limits depend on coolant temperature, pressure and flow rate, which are measured globally instead of locally, as well as the estimated power. Traditionally, in order to

protect reactors from these limits, safety-grade power sensors, such as fission chambers, ion chambers or self-powered neutron detectors are used to measure neutron flux. This flux measurement is then related to the local energy deposition, often requiring the calculation of a peak-to-average power ratio that must bound all fuel cycle conditions, as well as the local material properties which introduces uncertainty in the predicted thermal margins. Subsequently, the direct measurement of the local nuclear energy deposition and the local heat transfer rate reduces the uncertainty in the predicted thermal margins, resulting in improvement in performance and economy of new reactor designs.

A recently developed in-core calorimetric instrument called constant temperature power sensor^{39,40} (CTPS) can directly measure the local nuclear energy deposition and heat transfer rate. This design concept is based on the idea of adding heat through resistive dissipation of input electrical energy, to a small mass of actual reactor fuel analogue (Mode 1). A feedback control loop is used to provide the exact amount of input electrical energy in order to keep the fuel mass at a constant temperature regardless of the nuclear energy generated in the mass. In steady state, the input electrical energy required will be related to the actual nuclear energy deposition, given the assumption that the external heat transfer rate remains constant. To measure the local heat transfer rate, the sensor can be temporarily taken out of the control loop (Mode 2), or practically, the supplied electrical current can be reduced to 1% of the steady state value in Mode 1. The local heat transfer rate can be obtained from the

dynamic characteristics of the sensor temperature decaying back from the steady state value in Mode 1 to its original level.

In this study, a two-node model⁴¹ is used to describe the decay process, where the inner node (node 1) consists of the heated wire and the fissile material and the outer node (node 2) consists of the insulator and the metal layer at the surface (Fig. 4.1)

Fig. 4.1: A Schematic Diagram of CTPS Sensor

Based on the two-node model, the sensor dynamics can be expressed as the following

$$C_1 M_1 \dot{T}_1 = q_{nuc} + i_c^2(t) R_0 (1 + \alpha(T_1 - T_{ref})) - U_1 A_1 (T_1 - T_2) \quad (4.26)$$

$$C_2 M_2 \dot{T}_2 = U_1 A_1 (T_1 - T_2) - U_2 A_2 (T_2 - T_{env}) \quad (4.27)$$

where

T_1, T_2 : Temperature for node 1 and 2 (K)

- T_{env} :Coolant bulk temperature (K),
- C_1, C_2 :Specific heat capacities for node 1 and 2 (J/(kg K)).
- M_1, M_2 :Mass of node 1 and 2 (kg).
- U_1, U_2 :Effective heat transfer coefficient for node 1 and 2 (J/(m²sK))
- A_1, A_2 :Effective heat transfer area for node 1 and node2 (m²)
- i_c :Input electric current (A)
- T_{ref} :Reference Temperature (K)
- R_0 :Reference resistance (Ω)
- α :Temperature coefficient of resistance (the term $R_0(1 + \alpha(T_1 - T_{ref}))$ in Equation (4.26) defines the heated wire resistance) (1/K)
- q_{nuc} :Local nuclear energy deposition (J)

The initial state of the decay process is determined by the steady state of Mode 1, which can be obtained by letting the derivative terms in Equation (4.26) and (4.27) to be zero and replace $i_c(t)$ by its steady state value in Mode 1,

$$0 = q_{nuc} + i_0^2 R_0 (1 + \alpha(T_{10} - T_{ref})) - U_1 A_1 (T_{10} - T_{20}) \quad (4.28)$$

$$0 = U_1 A_1 (T_{10} - T_{20}) - U_2 A_2 (T_{20} - T_{env}) \quad (4.29)$$

where

i_0 : Steady state value of the input electric current in mode 1

T_{10}, T_{20} : Steady state value of the temperature for node 1 and 2 in Mode 1

Subtracting Equation (4.28) from (4.26) and Equation (4.29) from (4.27) we get

$$\begin{aligned}
C_1 M_1 \dot{T}_1 &= i_c^2(t) R_0 (1 + \alpha(T_1 - T_{ref})) - i_0^2 R_0 (1 + \alpha(T_{10} - T_{ref})) - U_1 A_1 ((T_1 - T_{10}) - (T_2 - T_{20})) \\
C_2 M_2 \dot{T}_2 &= U_1 A_1 ((T_1 - T_{10}) - (T_2 - T_{20})) - U_2 A_2 (T_2 - T_{20})
\end{aligned} \tag{4.30}$$

By defining $\bar{T}_1 = T_{10} - T_1$ and $\bar{T}_2 = T_{20} - T_2$, Equation (4.30) can be rewritten as

$$\begin{aligned}
-C_1 M_1 \dot{\bar{T}}_1 &= (R_0 \alpha T_{ref} - R_0 - R_0 \alpha T_{10})(i_0^2 - i_c^2(t)) - i_c^2(t) R_0 \alpha \bar{T}_1 - U_1 A_1 (\bar{T}_2 - \bar{T}_1) \\
-C_2 M_2 \dot{\bar{T}}_2 &= U_1 A_1 (\bar{T}_2 - \bar{T}_1) + U_2 A_2 (\bar{T}_2)
\end{aligned} \tag{4.31}$$

Notice that both $(i_0^2 - i_c^2(t))$ and $i_c^2(t)$ are step functions

$$\begin{aligned}
(i_0^2 - i_c^2(t)) &= \begin{cases} 0 & t < 0 \\ (1 - 10^{-4}) i_0^2 & t > 0 \end{cases} \\
i_c^2(t) &= \begin{cases} i_0^2 & t < 0 \\ i_c^2 = 10^{-4} i_0^2 & t > 0 \end{cases}
\end{aligned}$$

Taking Laplace transform on both sides of Equation (4.31) we get

$$\begin{aligned}
\bar{\bar{T}}_1(s) &= \frac{(1 - 10^{-4}) i_0^2 R_0 (1 + \alpha(T_{10} - T_{ref})) (C_2 M_2 s + U_1 A_1 + U_2 A_2)}{s \{ C_1 M_1 C_2 M_2 s^2 + [C_1 M_1 (U_2 A_2 + U_1 A_1) + C_2 M_2 (-i_c^2 R_0 \alpha + U_1 A_1)] \} + F}
\end{aligned} \tag{4.32}$$

where $F = -i_c^2 R_0 \alpha U_2 A_2 + U_1 A_1 (-i_c^2 R_0 \alpha + U_2 A_2)$

The Laplace transform of $\bar{\bar{T}}_1$ indicates $\bar{T}_1(t)$ can be written as the sum of a constant and two exponential terms. Hence T_1 will be of the form

$$T_1(t) = T_\infty + \theta_1 e^{-\theta_2 t} + \theta_3 e^{-\theta_4 t} + w \tag{4.33}$$

where w is the noise term, which includes the modeling uncertainty and measurement noise and T_∞ is the final steady state value of T_1 when the exponential terms die out.

The $\theta_1, \theta_2, \theta_3, \theta_4$ are unknown parameters which are going to be estimated and θ_2, θ_4 are the roots of the following second order equation

$$C_1 M_1 C_2 M_2 S^2 + [C_1 M_1 (U_2 A_2 + U_1 A_1) + C_2 M_2 (-i_e^2 R_0 \alpha + U_1 A_1)] S + F = 0$$

After θ_2, θ_4 are estimated, the local heat transfer coefficients U_1, U_2 can be obtained by solving the following equations

$$\theta_2 + \theta_4 = -\frac{C_1 M_1 (U_2 A_2 + U_1 A_1) + C_2 M_2 (-i_e^2 R_0 \alpha + U_1 A_1)}{C_1 M_1 C_2 M_2}$$

$$\theta_2 \theta_4 = \frac{-i_e^2 R_0 \alpha U_2 A_2 + U_1 A_1 (-i_e^2 R_0 \alpha + U_2 A_2)}{C_1 M_1 C_2 M_2}$$

In the following sections, the LSE, GMLE and the extended DSD method (i.e Equations (3.22) and (3.23)) will be applied to estimate the unknown parameters $\theta_1, \theta_2, \theta_3, \theta_4$ in Equation (4.33). The estimation is based on the multi-node numerical model simulation data⁴² with sampling time step $\tau = 0.2s$ and 250 time steps have been simulated.

4.3.2 Estimation Results

The system considered in the estimation process is the discrete form of Equation (4.33), i.e.

$$T_{1,k} = T_\infty + \theta_1 e^{-\theta_2 t_k} + \theta_3 e^{-\theta_4 t_k} + w_k \quad (4.34)$$

where $T_{1,k} = T_1(t_k)$ is the measured temperature at time step k and w_k is the noise term,.

Assuming w_k are Gaussian and statistically independent for all k we have

$$f_w(w_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{w_i^2}{2\sigma^2}}, i = 1, 2, \dots, k \quad (4.35)$$

The Least Squares Estimator

The LSE results were obtained by minimizing the objective function

$$\sum_{k=1}^N (T_{1,k} - T_{\infty} - \theta_1 e^{-\theta_2 t_k} - \theta_3 e^{-\theta_4 t_k})^2$$

where the estimated parameters are denoted as $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$, $T_{\infty} = 710.98$ °F and $N = 250$ is the total number of data points measured. The LSE estimation results are shown in Table 4.1.

Parameter	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
Estimated Value	5.4854	0.2833	19.6994	2.7906

Table 4.1: Estimation Results of the LS Estimator

Figure 4.2 shows the comparison of the fitted data with the measured data and Fig. 4.3 shows the residuals, which are calculated as the following:

$$R(k) = T_{1,k} - T_{\infty} + \hat{\theta}_1 e^{-\hat{\theta}_2 t_k} + \hat{\theta}_3 e^{-\hat{\theta}_4 t_k}$$

where $T_{1,k}$ is the measured data at time step k and $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ are the estimated parameter values. The maximum residual is 0.8183 and the standard deviation of w_k is determined by the following unbiased estimation⁴³ of σ^2 (see Equation (4.35)):

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{k=1}^N (T_{1,k} - T_{\infty} + \hat{\theta}_1 e^{-\hat{\theta}_2 t_k} + \hat{\theta}_3 e^{-\hat{\theta}_4 t_k})^2 = 0.1 \quad (4.36)$$

Figures 4.2 and 4.3 indicate that fitting is good with the maximum difference between the measured temperatures and those calculated using $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ being less than $0.1^\circ F$ after $t = 5s$

Fig. 4.2: Comparison of the Measured Data and Fitting Results by the Least Square Method

Fig. 4.3: Plot of the Residuals

The Generalized Maximum Likelihood Estimator

Using Equation (4.23), the likelihood function is calculated as the product of a set of Gaussian distributions. Table 4.2 shows the estimation results of the GMLE estimator. Comparison of the data in Tables 4.1 and 4.2 shows that the relative errors between the GMLE and LSE results are $(5.4854-5.4825)/5.4854=0.05\%$, $(0.2833-0.2831)/0.2833=0.07\%$, $(19.6994-19.7017)/19.6994=-0.01\%$, $(2.7906-2.7897)/2.7906 = 0.03\%$, for $\theta_1, \theta_2, \theta_3, \theta_4$, respectively. The small difference between GMLE and LSE estimation results is due to the different numerical minimization techniques applied.

Parameter	θ_1	θ_2	θ_3	θ_4
Estimated Value	5.4825	0.2831	19.7017	2.7897

Table 4.2: Estimation Results of the GMLE Estimator

The Extended DSD Method

This problem (i.e. Equation (4.34)) will be fitted to the theory in Chapter 3 by rewriting it as dynamic system

$$\begin{aligned}\dot{\theta} &= 0 \\ T_{i,k} &= h(\theta) + w_k\end{aligned}\tag{4.37}$$

where $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$ and $h(\theta) = T_\infty + \theta_1 e^{-\theta_2 t_k} + \theta_3 e^{-\theta_4 t_k}$

The initial distribution $p(\theta_0 | T_{1,0})$ is assumed to be uniform over the unknown parameter space. The posterior pdf at each time step is updated from Equation (3.23) with

$$\begin{aligned}
p(j_{k+1}|\bar{T}_{1:k+1}) &= \frac{p(j_{k+1}, T_{1,k+1}|T_{1,k})}{\sum_{j_{k+1}} p(j_{k+1}, T_{1,k+1}|\bar{T}_{1,k})} \\
&= \frac{\sum_{j_k} g(j_{k+1}|j_k) p(j_k|\bar{T}_{1,k})}{\sum_{j_{k+1}} \sum_{j_k} g(j_{k+1}|j_k) p(j_k|\bar{T}_{1,k})}
\end{aligned} \tag{4.38}$$

where j denotes the cells which partition the parameter space, k is the time step index and $\bar{T}_{1,k}$ denote all the measurements from the initial time step to the current time step.

Since the system model is without uncertainty (Equation (4.37)), the transition probability will be calculated as the following (see Equation (3.28))

$$\begin{aligned}
g(j_{k+1}|j_k) &= \int_{j_k} f_w(T_{1,k+1} - h(\tilde{x}(\theta))) \frac{p(\theta|\bar{T}_{1,k})}{\int_j p(\theta|\bar{T}_{1,k}) d\theta} e(\tilde{x}(\theta)) d\theta \\
&\approx \sum_i f_w(T_{1,k+1} - h(\tilde{x}(\hat{\theta}^i))) \frac{p(\hat{\theta}^i|\bar{T}_{1,k})}{N p(\hat{\theta}^i|\bar{T}_{1,k}) \Delta\hat{\theta}^i} e(\tilde{x}(\hat{\theta}^i)) \Delta\hat{\theta}^i \\
&= \sum_i \frac{1}{N} f_w(T_{1,k+1} - h(\tilde{x}(\hat{\theta}^i))) e(\tilde{x}(\hat{\theta}^i))
\end{aligned} \tag{4.39}$$

with $\hat{\theta}^i$ denoting point values picked from cell j_k for the computation of the cell-to-cell transition probabilities (see Equation (3.28)) and $f_w(w)$ is the pdf of the noise which is assumed to be Gaussian, i.e.

$$f_w = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{w^2}{2\sigma^2}\right)$$

with the standard deviation $\sigma^2 = 0.1$ as given by Equation (4.36) and

$$\begin{aligned}
\tilde{x}(\hat{\theta}^i) &= \hat{\theta}^i \\
h(\tilde{x}(\hat{\theta}^i)) &= h(\hat{\theta}^i) = T_\infty + \theta_1' e^{-\theta_2' t_k} + \theta_3' e^{-\theta_4' t_k}
\end{aligned}$$

$$e_{k+1}(\tilde{x}(\theta^i)) = \begin{cases} 0 & \tilde{x}(\theta^i) \notin j_{k+1} \\ 1 & \tilde{x}(\theta^i) \in j_{k+1} \end{cases}$$

The partitioning data used is shown in Table 4.3.

Parameter	θ_1	θ_2	θ_3	θ_4
Max	7	0.6	22	3.5
Min	4	0.0	17	2.5
Cell Number	89	89	89	89
Number of Points N in Equation(4.39)	3	3	3	3

Table 4.3: Partitioning Data Used for the CTPS Sensor Model

The estimated 99.9% credibility interval as a function of time and the most probable values of $\theta_1, \theta_2, \theta_3, \theta_4$ are shown, respectively, in Figs.4.4, 4.5, 4.6 and 4.7. The transient nature of the estimation process can be explained by the fact that data were generated using a multi-node numerical model⁴² and subsequently DSD is regarding the difference between this model and $T_{l,k} - w_k$ in Equation (4.34) as model uncertainty. This fact is also illustrated in the residual plot on Fig. 4.2 where the residuals are much bigger during the first 10 seconds.

Fig. 4.4: Estimated Interval of θ_1 as a Function of Time for CTPS Sensor Model.
(asterisks indicate the most probable value)

Fig. 4.5: Estimated Interval of θ_2 as a Function of Time for CTPS Sensor Model
(asterisks indicate the most probable value)

Fig. 4.6: Estimated Interval of θ_3 as a Function of Time for CTPS Sensor Model (asterisks indicate the most probable value)

Fig. 4.7: Estimated Interval of θ_4 as a Function of Time for CTPS Sensor Model (asterisks indicate the most probable value)

The most probable value of the posterior distribution at the final time step is listed in Table 4.4.

Parameter	θ_1	θ_2	θ_3	θ_4
Most Probable Value at Final Time Step	5.4326	0.2798	19.6685	2.7865

Table 4.4: Estimated Most Probable Values at the Final Time Step for the CTPS Sensor Model

Comparing the most probable value at the final time step estimated by DSD (Table 4.4) with the estimation results by GMLE (Table 4.2) and LSE (Table 4.1) shows that DSD gives very close estimation results with the other two estimators. The relative difference between DSD and LSE estimation results are $(5.4854-5.4326)/5.4854=1\%$, $(0.2833-0.2798)/0.2833=1.2\%$, $(19.6994-19.6685)/19.6994=0.2\%$, $(2.7906-2.7865)/2.7906=0.1\%$ for $\theta_1, \theta_2, \theta_3, \theta_4$ respectively. The relative error between DSD and GMLE estimation results are $(5.4825-5.4326)/5.4825=0.9\%$, $(0.2831-0.2798)/0.2831=1.1\%$, $(19.7017-19.6685)/19.7017=0.2\%$, $(2.7897-2.7865)/2.7897=0.1\%$ for $\theta_1, \theta_2, \theta_3, \theta_4$ respectively. In addition, as an estimation algorithm designed for probability risk analysis, DSD yields additional information such as expected values, standard deviations and credibility intervals for the quantities to be estimated which are useful for risk-informed decision making and cannot be obtained by the LSE and GMLE methods. For example, Table 4.5 shows the 99.99% credibility interval at the final time step using the posterior

probability distributions $p(\theta_1|\bar{T}_{1,k})$, $p(\theta_2|\bar{T}_{1,k})$, $p(\theta_3|\bar{T}_{1,k})$, $p(\theta_4|\bar{T}_{1,k})$ of the unknown parameters at $t = 250 * \tau = 50s$ as given in Fig.4.8. The estimation uncertainty can be directly obtained through these probability distributions. It is clear that both the LSE and GMLE estimation results fall into the 99% credibility intervals and lie in the middle of the corresponding intervals. Using DSD, the probability for the LSE and GMLE results to be in selected intervals can also be calculated. Table 4.6 lists these probabilities for the intervals $[\hat{\theta}_i(1 - 2\%), \hat{\theta}_i(1 + 2\%)]$, $i = 1, 2, 3, 4$ where $\hat{\theta}_i$ is the estimated parameter value using LSE. Table 4.6 shows that if LSE is used to estimate parameters $\theta_1, \theta_2, \theta_3, \theta_4$, the probability that the real parameter values are within 2% around the estimated value are 79%, 87%, 99.5%, 88% for $\theta_1, \theta_2, \theta_3, \theta_4$ respectively, which can be used as a probabilistic measure of how close the system is to the thermal limits.

Parameter	θ_1	θ_2
Estimated 99.99% Credibility Interval at Final Time Step	[5.0112 6.0225]	[0.2629 0.3101]

Parameter	θ_3	θ_4
Estimated 99.99% Credibility Interval at Final Time Step	[19.1348 20.3146]	[2.6236 2.9831]

Table 4.5: Estimated Credibility Intervals for the CTPS Sensor Model

θ_1	Probability of the Selected Interval	θ_2	Probability of the Selected Interval
[5.3757 5.5951]	79%	[0.2776 0.2890]	87%

θ_3	Probability of the Selected Interval	θ_4	Probability of the Selected Interval
[19.3054 20.0933]	99.5%	[2.7348 2.8464]	88%

Table 4.6: Probability of Selected Intervals for the CTPS Sensor Model

(a)

(b)

(c)

(d)

Fig. 4.8: Estimated pdf of the Unknown Parameters at the Final Time Step $k = 250$:
 (a) $p(\theta_1|\bar{T}_{1,k})$, (b) $p(\theta_2|\bar{T}_{1,k})$, (c) $p(\theta_3|\bar{T}_{1,k})$, (d) $p(\theta_4|\bar{T}_{1,k})$

CHAPTER 5

THE DSD AND THE KALMAN FILTER

In Chapter 3, the extended DSD algorithm for generalized noise distribution is derived from the continuous Bayesian estimator and in Chapter 4 this extended algorithm is compared to the GMLE and LSE. This chapter shows that the continuous Bayesian estimator derived in Section 3.1 is equivalent to the well known conventional state/parameter estimation technique, the Kalman filter, for linear systems with white Gaussian noise (Section 5.1). Section 5.2 compares the discrete version of the Bayesian estimator (i.e. the extended DSD algorithm developed in Section 3.2) to the Kalman filter approach using a simple discrete linear system.

5.1 The Continuous Bayesian Estimator and the Kalman Filter

Consider the following system,

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}_k \mathbf{x}_k + \mathbf{v}_k \\ \mathbf{y}_k &= \mathbf{C}_k \mathbf{x}_k + \mathbf{w}_k \end{aligned} \tag{5.1}$$

where $\mathbf{A}_k$ and $\mathbf{C}_k$ are $L \times L$ matrices and $\mathbf{v}_k, \mathbf{w}_k$ represent white Gaussian noise with

$$f_{\mathbf{w}_k}(\mathbf{w}_k) = \frac{1}{(2\pi)^{m/2} \sqrt{|\mathbf{R}_k|}} \exp\left\{-\frac{1}{2} \mathbf{w}_k' \mathbf{R}_k^{-1} \mathbf{w}_k\right\} \tag{5.2}$$

$$f_{v_k}(v_k) = \frac{1}{(2\pi)^{n_k/2} \sqrt{|\mathbf{M}_k|}} \exp\left\{-\frac{1}{2} v_k' \mathbf{M}_k^{-1} v_k\right\} \quad (5.3)$$

$$\begin{aligned} E\{w_k\} &= E\{v_k\} = 0 \\ \mathbf{R}_k &= E\{w_k w_k'\} \quad \mathbf{M}_k = E\{v_k v_k'\} \\ E\{w_k v_i'\} &= 0 \quad \text{for all } k, i \end{aligned} \quad (5.4)$$

In this section we will prove that if

$$p(x_0|y_0) = \frac{1}{(2\pi)^{n_2/2} \sqrt{|\Sigma_{0|0}|}} \exp\left\{-\frac{1}{2} (x_0 - \hat{x}_0)' \Sigma_{0|0}^{-1} (x_0 - \hat{x}_0)\right\}$$

where $\hat{x}_0, \Sigma_{0|0}$ are the mean and covariance of the initial guess of the pdf of x at the initial time step, then Equation (3.6) is equivalent to the Kalman filter in the sense that the mean and covariance of the posterior pdf is the same as the results obtained from the Kalman filter. The proof will proceed in three steps through Propositions 5.1, 5.2 and 5.3.

Proposition 5.1:

Assume that the posterior probability distribution $p(x_k|\bar{y}_k)$ is Gaussian, i.e.

$$p(x_k|\bar{y}_k) = \frac{1}{(2\pi)^{n_k/2} \sqrt{|\Sigma_{k|k}|}} \exp\left\{-\frac{1}{2} (x_k - \hat{x}_k)' \Sigma_{k|k}^{-1} (x_k - \hat{x}_k)\right\} \quad (5.5)$$

where $\hat{x}_k, \Sigma_{k|k}$ is the mean and covariance of x_k at time step k . Then, $p(x_{k+1}|\bar{y}_k)$ is

Gaussian with the mean $A_k \hat{x}_k$ and covariance matrix $\Sigma_{k+1|k} = A_k \Sigma_{k|k} A_k' + M_k$

Proof:

From Equation (3.8)

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k, \bar{\mathbf{y}}_k) = \frac{p(\mathbf{x}_{k+1}, \mathbf{x}_k, \bar{\mathbf{y}}_k)}{p(\mathbf{x}_k, \bar{\mathbf{y}}_k)}$$

$$p(\mathbf{x}_k|\bar{\mathbf{y}}_k) = \frac{p(\mathbf{x}_k, \bar{\mathbf{y}}_k)}{p(\bar{\mathbf{y}}_k)}$$

we have

$$\begin{aligned} p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_k) &= \int p(\mathbf{x}_{k+1}, \mathbf{x}_k|\bar{\mathbf{y}}_k) d\mathbf{x}_k \\ &= \int \frac{p(\mathbf{x}_{k+1}, \mathbf{x}_k, \bar{\mathbf{y}}_k)}{p(\mathbf{x}_k, \bar{\mathbf{y}}_k)} \frac{p(\mathbf{x}_k, \bar{\mathbf{y}}_k)}{p(\bar{\mathbf{y}}_k)} d\mathbf{x}_k \\ &= \int p(\mathbf{x}_{k+1}|\mathbf{x}_k, \bar{\mathbf{y}}_k) p(\mathbf{x}_k|\bar{\mathbf{y}}_k) d\mathbf{x}_k \end{aligned} \quad (5.6)$$

Note that $\mathbf{x}_{k+1}$ only depends on the value of $\mathbf{x}_k$ and $\mathbf{v}_k$ from Equation (5.1). When $\mathbf{x}_k$ is given, the probability of $\mathbf{x}_{k+1}$ will be determined by the value of $\mathbf{v}_k$ whose value does not depend on any previous system states by the white noise assumption. Then

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k, \bar{\mathbf{y}}_k) = p(\mathbf{x}_{k+1}|\mathbf{x}_k) \quad (5.7)$$

Substituting Equation (5.7) into Equation (5.6), we get

$$p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_k) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{x}_k|\bar{\mathbf{y}}_k) d\mathbf{x}_k \quad (5.8)$$

Note that from the system equations (Equation 5.1) and the probability distribution of $\mathbf{v}_k$ (Equation 5.3), we get

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k) = \frac{1}{(2\pi)^{\frac{n_x}{2}} \sqrt{|\mathbf{M}_k|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_{k+1} - \mathbf{A}_k \mathbf{x}_k)' \mathbf{M}_k^{-1} (\mathbf{x}_{k+1} - \mathbf{A}_k \mathbf{x}_k) \right\} \quad (5.9)$$

Equations (5.5), (5.8) and (5.9) yield

$$\begin{aligned}
p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_k) &= \int p(\mathbf{x}_{k+1}|\mathbf{x}_k)p(\mathbf{x}_k|\bar{\mathbf{y}}_k)d\mathbf{x}_k \\
&= \int \frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{M}_k|}} \exp\left\{-\frac{1}{2}(\mathbf{x}_{k+1} - \mathbf{A}_k\mathbf{x}_k)' \mathbf{M}_k^{-1}(\mathbf{x}_{k+1} - \mathbf{A}_k\mathbf{x}_k)\right\} \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma_{k|k}|}} \exp\left\{-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{x}}_k)' \Sigma_{k|k}^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_k)\right\} d\mathbf{x}_k \\
&= \int \frac{1}{(2\pi)^n \sqrt{|\mathbf{M}_k| |\Sigma_{k|k}|}} \exp\left\{-\frac{1}{2}\left[\mathbf{x}_{k+1}' \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \mathbf{x}_k' \mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{A}_k \mathbf{x}_k - 2\mathbf{x}_k' \mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \mathbf{x}_k' \Sigma_{k|k}^{-1} \mathbf{x}_k + \hat{\mathbf{x}}_k' \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k - 2\mathbf{x}_k' \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k\right]\right\} d\mathbf{x}_k \\
&= \int \frac{1}{(2\pi)^n \sqrt{|\mathbf{M}_k| |\Sigma_{k|k}|}} \exp\left\{-\frac{1}{2}\left[\mathbf{x}_{k+1}' \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}_k' \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k + \mathbf{x}_k' (\mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{A}_k + \Sigma_{k|k}^{-1}) \mathbf{x}_k - 2\mathbf{x}_k' (\mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k)\right]\right\} d\mathbf{x}_k
\end{aligned} \tag{5.10}$$

Equation (5.10) makes use of the fact that $\mathbf{M}_k^{-1}$ and $\Sigma_{k|k}^{-1}$ are symmetric matrices. Then since $\mathbf{x}_k' \mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{x}_{k+1}$ is a scalar,

$\mathbf{x}_k' \mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{x}_{k+1} = (\mathbf{x}_k' \mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{x}_{k+1})' = (\mathbf{M}_k^{-1} \mathbf{x}_{k+1})' (\mathbf{x}_k' \mathbf{A}_k')' = \mathbf{x}_{k+1}' \mathbf{M}_k^{-1} \mathbf{A}_k \mathbf{x}_k = \mathbf{x}_{k+1}' \mathbf{M}_k^{-1} \mathbf{A}_k \mathbf{x}_k$. We can similarly show that $\mathbf{x}_k' \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k' \Sigma_{k|k}^{-1} \mathbf{x}_k$. Now denote $\mathbf{H} = (\mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{A}_k + \Sigma_{k|k}^{-1})$ and $\mathbf{z} = (\mathbf{A}_k' \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k)$ for notational simplicity. Then

Equation (5.10) can be rewritten in quadratic form of $\mathbf{x}_k$ as the following,

$$\begin{aligned}
 p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_k) &= \int \frac{1}{(2\pi)^n \sqrt{|\mathbf{M}_k| |\boldsymbol{\Sigma}_{k|k}|}} \exp \left\{ -\frac{1}{2} [\mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}'_k \boldsymbol{\Sigma}_{k|k}^{-1} \hat{\mathbf{x}}_k + \mathbf{x}'_k \mathbf{H} \mathbf{x}_k - 2 \mathbf{x}'_k \mathbf{z}] \right\} d\mathbf{x}_k \\
 &= \int \frac{1}{(2\pi)^n \sqrt{|\mathbf{M}_k| |\boldsymbol{\Sigma}_{k|k}|}} \exp \left\{ -\frac{1}{2} [\mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}'_k \boldsymbol{\Sigma}_{k|k}^{-1} \hat{\mathbf{x}}_k + (\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z}) - \mathbf{z}' \mathbf{H}^{-1} \mathbf{z}] \right\} d\mathbf{x}_k \\
 &= \int \frac{\sqrt{|\mathbf{H}^{-1}|}}{(2\pi)^n \sqrt{|\mathbf{M}_k| |\boldsymbol{\Sigma}_{k|k}|} \sqrt{|\mathbf{H}^{-1}|}} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z})] \right\} \exp \left\{ -\frac{1}{2} [\mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}'_k \boldsymbol{\Sigma}_{k|k}^{-1} \hat{\mathbf{x}}_k - \mathbf{z}' \mathbf{H}^{-1} \mathbf{z}] \right\} d\mathbf{x}_k
 \end{aligned} \tag{5.11}$$

Again, Equation (5.11) makes use of the facts: a) since the covariance matrices $\mathbf{M}_k, \boldsymbol{\Sigma}_{k|k}$ are symmetric, $\mathbf{H}$ is also symmetric,

and, b) $\mathbf{H}$ and $\mathbf{z}$ do not depend on $\mathbf{x}_k$. Also since $\frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{H}^{-1}|}} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z})] \right\}$ is Gaussian

$$\int \frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{H}^{-1}|}} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_k - \mathbf{H}^{-1} \mathbf{z})] \right\} d\mathbf{x}_k = 1 \tag{5.12}$$

Hence, Equation (5.11) can be written as

$$\begin{aligned}
& p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_k) \\
&= \int \frac{\sqrt{|H^{-1}|}}{(2\pi)^n \sqrt{|\mathbf{M}_k|} |\Sigma_{k|k}| \sqrt{|H^{-1}|}} \exp\left\{-\frac{1}{2}[(\mathbf{x}_k - H^{-1}\mathbf{z})' H(\mathbf{x}_k - H^{-1}\mathbf{z})]\right\} \exp\left\{-\frac{1}{2}[\mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}'_k \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k - \mathbf{z}' H^{-1} \mathbf{z}]\right\} d\mathbf{x}_k \\
&= \frac{\sqrt{|H^{-1}|}}{(2\pi)^{n/2} \sqrt{|\mathbf{M}_k|} |\Sigma_{k|k}|} \exp\left\{-\frac{1}{2}[\mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}'_k \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k - \mathbf{z}' H^{-1} \mathbf{z}]\right\} \int \frac{1}{(2\pi)^{n/2} \sqrt{|H^{-1}|}} \exp\left\{-\frac{1}{2}[(\mathbf{x}_k - H^{-1}\mathbf{z})' H(\mathbf{x}_k - H^{-1}\mathbf{z})]\right\} d\mathbf{x}_k \\
&= \frac{\sqrt{|H^{-1}|}}{(2\pi)^{n/2} \sqrt{|\mathbf{M}_k|} |\Sigma_{k|k}|} \exp\left\{-\frac{1}{2}[\mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}'_k \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k - \mathbf{z}' H^{-1} \mathbf{z}]\right\} \\
& \quad (5.13)
\end{aligned}$$

99

Resubstituting the definition of H and $\mathbf{z}$ back into Equation (5.13), we get

$$\begin{aligned}
& p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_k) \\
&= \frac{\sqrt{(\mathcal{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \Sigma_{k|k}^{-1})^{-1}}}{(2\pi)^{n/2} \sqrt{|\mathbf{M}_k|} |\Sigma_{k|k}|} \exp\left\{-\frac{1}{2}[\mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}'_k \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k - (\mathcal{A}'_k \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \Sigma_{k|k}^{-1})^{-1} (\mathcal{A}'_k \mathbf{M}_k^{-1} \mathbf{x}_{k+1} + \Sigma_{k|k}^{-1} \hat{\mathbf{x}}_k)]\right\} \\
&= \alpha \exp\left\{-\frac{1}{2}[\beta + \xi - \psi]\right\} \\
& \quad (5.14)
\end{aligned}$$

where

$$\alpha = \frac{\sqrt{|(A'_k M_k^{-1} A_k + \Sigma_{k|k}^{-1})^{-1}|}}{(2\pi)^{\frac{n'}{2}} \sqrt{|M_k|} \sqrt{|\Sigma_{k|k}|}} \quad (5.15)$$

$$\beta = \hat{x}'_k (\Sigma_{k|k}^{-1} - \Sigma_{k|k}^{-1} (A'_k M_k^{-1} A_k + \Sigma_{k|k}^{-1})^{-1} \Sigma_{k|k}^{-1}) \hat{x}_k \quad (5.16)$$

$$\xi = x'_{k+1} (M_k^{-1} - M_k^{-1} A_k (A'_k M_k^{-1} A_k + \Sigma_{k|k}^{-1})^{-1} A'_k M_k^{-1}) x_{k+1} \quad (5.17)$$

$$\psi = 2x'_{k+1} M_k^{-1} A_k (A'_k M_k^{-1} A_k + \Sigma_{k|k}^{-1})^{-1} \Sigma_{k|k}^{-1} \hat{x}'_k \quad (5.18)$$

Let $E = \Sigma_{k|k}^{-1}$, $F = A'_k$, $G = M_k^{-1} A_k$. Using the Matrix Inversion Lemma⁴⁴ (or the

Sherman-Morrison-Woodbury formula)

$$(E + FG)^{-1} = E^{-1} - E^{-1} F (I + GE^{-1} F)^{-1} GE^{-1}$$

we have

$$\begin{aligned} (A'_k M_k^{-1} A_k + \Sigma_{k|k}^{-1})^{-1} &= \Sigma_{k|k} - \Sigma_{k|k} A'_k (I + M_k^{-1} A_k \Sigma_{k|k} A'_k)^{-1} M_k^{-1} A_k \Sigma_{k|k} \\ &= \Sigma_{k|k} - \Sigma_{k|k} A'_k (M_k + M_k M_k^{-1} A_k \Sigma_{k|k} A'_k)^{-1} A_k \Sigma_{k|k} \\ &= \Sigma_{k|k} - \Sigma_{k|k} A'_k (M_k + A_k \Sigma_{k|k} A'_k)^{-1} A_k \Sigma_{k|k} \end{aligned}$$

Then

$$\begin{aligned} \beta &= \hat{x}'_k (\Sigma_{k|k}^{-1} - \Sigma_{k|k}^{-1} (A'_k M_k^{-1} A_k + \Sigma_{k|k}^{-1})^{-1} \Sigma_{k|k}^{-1}) \hat{x}_k \\ &= \hat{x}'_k (\Sigma_{k|k}^{-1} - \Sigma_{k|k}^{-1} (\Sigma_{k|k} - \Sigma_{k|k} A'_k (M_k + A_k \Sigma_{k|k} A'_k)^{-1} A_k \Sigma_{k|k}) \Sigma_{k|k}^{-1}) \hat{x}_k \\ &= \hat{x}'_k (\Sigma_{k|k}^{-1} - (\Sigma_{k|k}^{-1} - A'_k (M_k + A_k \Sigma_{k|k} A'_k)^{-1} A_k)) \hat{x}_k \\ &= \hat{x}'_k A'_k (M_k + A_k \Sigma_{k|k} A'_k)^{-1} A_k \hat{x}_k \end{aligned} \quad (5.20)$$

Similarly, let $E = M_k^{-1}$, $G = (A_k \Sigma_{k|k} A'_k)^{-1}$, $F = I$. Then

$$\begin{aligned} (M_k^{-1} + (A_k \Sigma_{k|k} A'_k)^{-1})^{-1} &= M_k - M_k (I + (A_k \Sigma_{k|k} A'_k)^{-1} M_k)^{-1} (A_k \Sigma_{k|k} A'_k)^{-1} M_k \\ &= M_k - M_k (A_k \Sigma_{k|k} A'_k + (A_k \Sigma_{k|k} A'_k)(A_k \Sigma_{k|k} A'_k)^{-1} M_k)^{-1} M_k \\ &= M_k - M_k (A_k \Sigma_{k|k} A'_k + M_k)^{-1} M_k \end{aligned}$$

Hence

$$\begin{aligned}
\xi &= \mathbf{x}'_{k+1} (\mathbf{M}_k^{-1} - \mathbf{M}_k^{-1} \mathbf{A}_k (\mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \Sigma_{k|k}^{-1})^{-1} \mathbf{A}'_k \mathbf{M}_k^{-1}) \mathbf{x}_{k+1} \\
&= \mathbf{x}'_{k+1} (\mathbf{M}_k^{-1} - \mathbf{M}_k^{-1} ((\mathbf{A}'_k)^{-1} \mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k \mathbf{A}_k^{-1} + (\mathbf{A}'_k)^{-1} \Sigma_{k|k}^{-1} \mathbf{A}_k^{-1})^{-1} \mathbf{M}_k^{-1}) \mathbf{x}_{k+1} \\
&= \mathbf{x}'_{k+1} (\mathbf{M}_k^{-1} - \mathbf{M}_k^{-1} (\mathbf{M}_k^{-1} + (\mathbf{A}_k \Sigma_{k|k} \mathbf{A}'_k)^{-1})^{-1} \mathbf{M}_k^{-1}) \mathbf{x}_{k+1} \\
&= \mathbf{x}'_{k+1} (\mathbf{M}_k^{-1} - \mathbf{M}_k^{-1} (\mathbf{M}_k - \mathbf{M}_k (\mathbf{A}_k \Sigma_{k|k} \mathbf{A}'_k + \mathbf{M}_k)^{-1} \mathbf{M}_k) \mathbf{M}_k^{-1}) \mathbf{x}_{k+1} \\
&= \mathbf{x}'_{k+1} (\mathbf{A}_k \Sigma_{k|k} \mathbf{A}'_k + \mathbf{M}_k)^{-1} \mathbf{x}_{k+1}
\end{aligned} \tag{5.21}$$

In a similar manner, ψ and α in Equation (5.14) can be simplified as

$$\begin{aligned}
\psi &= 2 \mathbf{x}'_{k+1} \mathbf{M}_k^{-1} \mathbf{A}_k (\mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \Sigma_{k|k}^{-1})^{-1} \Sigma_{k|k}^{-1} \hat{\mathbf{x}}'_k \\
&= 2 \mathbf{x}'_{k+1} \mathbf{M}_k^{-1} (\mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k \mathbf{A}_k^{-1} + \Sigma_{k|k}^{-1} \mathbf{A}_k^{-1})^{-1} \Sigma_{k|k}^{-1} \hat{\mathbf{x}}'_k \\
&= 2 \mathbf{x}'_{k+1} \mathbf{M}_k^{-1} (\Sigma_{k|k} \mathbf{A}'_k \mathbf{M}_k^{-1} + \mathbf{A}_k^{-1})^{-1} \hat{\mathbf{x}}'_k \\
&= 2 \mathbf{x}'_{k+1} \mathbf{M}_k^{-1} (\Sigma_{k|k} \mathbf{A}'_k \mathbf{M}_k^{-1} + \mathbf{A}_k^{-1})^{-1} \hat{\mathbf{x}}'_k \\
&= 2 \mathbf{x}'_{k+1} (\Sigma_{k|k} \mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{M}_k + \mathbf{A}_k^{-1} \mathbf{M}_k)^{-1} \hat{\mathbf{x}}'_k \\
&= 2 \mathbf{x}'_{k+1} (\Sigma_{k|k} \mathbf{A}'_k + \mathbf{A}_k^{-1} \mathbf{M}_k)^{-1} \hat{\mathbf{x}}'_k \\
&= 2 \mathbf{x}'_{k+1} (\mathbf{A}_k^{-1} \mathbf{A}_k \Sigma_{k|k} \mathbf{A}'_k + \mathbf{A}_k^{-1} \mathbf{M}_k)^{-1} \hat{\mathbf{x}}'_k \\
&= 2 \mathbf{x}'_{k+1} (\mathbf{A}_k \Sigma_{k|k} \mathbf{A}'_k + \mathbf{M}_k)^{-1} \mathbf{A}_k \hat{\mathbf{x}}'_k
\end{aligned} \tag{5.22}$$

and

$$\begin{aligned}
\alpha &= \frac{\sqrt{|(\mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \boldsymbol{\Sigma}_{k|k}^{-1})^{-1}|}}{(2\pi)^{\frac{n_k}{2}} \sqrt{|\mathbf{M}_k|} \sqrt{|\boldsymbol{\Sigma}_{k|k}|}} = \frac{1}{(2\pi)^{\frac{n_k}{2}} \sqrt{|\mathbf{M}_k|} \sqrt{|\boldsymbol{\Sigma}_{k|k}|} \sqrt{|(\mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \boldsymbol{\Sigma}_{k|k}^{-1})|}} \\
&= \frac{1}{(2\pi)^{\frac{n_k}{2}} \sqrt{|\boldsymbol{\Sigma}_{k|k} (\mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \boldsymbol{\Sigma}_{k|k}^{-1}) \mathbf{M}_k|}} \\
&= \frac{1}{(2\pi)^{\frac{n_k}{2}} \sqrt{|(\boldsymbol{\Sigma}_{k|k} \mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \mathbf{I}) \mathbf{M}_k|}} \\
&= \frac{1}{(2\pi)^{\frac{n_k}{2}} \sqrt{|(\mathbf{A}_k^{-1} \mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}'_k \mathbf{M}_k^{-1} \mathbf{A}_k + \mathbf{A}_k^{-1} \mathbf{A}_k) \mathbf{M}_k|}} \\
&= \frac{1}{(2\pi)^{\frac{n_k}{2}} \sqrt{|\mathbf{A}_k^{-1}| |\mathbf{A}_k^{-1} (\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}'_k \mathbf{M}_k^{-1} + \mathbf{I}) \mathbf{A}_k \mathbf{M}_k| |\mathbf{A}_k|}} \\
&= \frac{1}{(2\pi)^{\frac{n_k}{2}} \sqrt{|(\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}'_k \mathbf{M}_k^{-1} + \mathbf{I}) \mathbf{M}_k|}} \\
&= \frac{1}{(2\pi)^{\frac{n_k}{2}} \sqrt{|(\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}'_k + \mathbf{M}_k)|}}
\end{aligned} \tag{5.23}$$

Substituting Equations (5.20) through (5.23) into Equation (5.14), we get

$$\begin{aligned}
 p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_k) &= \frac{1}{(2\pi)^{n_2/2} \sqrt{|(\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k)|}} \exp \left\{ -\frac{1}{2} \left[\hat{\mathbf{x}}_k' \mathbf{A}_k' (\mathbf{M}_k + \mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k')^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{x}_{k+1}' (\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k)^{-1} \mathbf{x}_{k+1} - 2 \mathbf{x}_{k+1}' (\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k)^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k \right] \right\} \\
 &= \frac{1}{(2\pi)^{n_2/2} \sqrt{|(\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k)|}} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k)' (\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k)^{-1} (\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k)] \right\}
 \end{aligned} \tag{5.24}$$

which is a Gaussian distribution with mean $\mathbf{A}_k \hat{\mathbf{x}}_k$ and variance $\boldsymbol{\Sigma}_{k+1|k} = \mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k$.

End of Proof

Proposition 5.2:

Given that the conditional probability $p(\mathbf{x}_k|\bar{\mathbf{y}}_k)$ is Gaussian, the posterior probability distribution $p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_{k+1})$ obtained through Equation (3.6) is also Gaussian and the mean $\hat{\mathbf{x}}_{k+1}$ and covariance matrix is $\boldsymbol{\Sigma}_{k+1|k+1}$ given by

$$\begin{aligned}
 \hat{\mathbf{x}}_{k+1} &= \mathbf{A}_k \hat{\mathbf{x}}_k + \boldsymbol{\Sigma}_{k+1|k+1} \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} (\mathbf{y}_{k+1} - \mathbf{C}_{k+1} \mathbf{A}_k \hat{\mathbf{x}}_k) \\
 \boldsymbol{\Sigma}_{k+1|k+1} &= (\mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \boldsymbol{\Sigma}_{k+1|k}^{-1})^{-1} = \boldsymbol{\Sigma}_{k+1|k} - \boldsymbol{\Sigma}_{k+1|k} \mathbf{C}_{k+1}' (\mathbf{C}_{k+1} \boldsymbol{\Sigma}_{k+1|k} \mathbf{C}_{k+1}' + \mathbf{R}_{k+1})^{-1} \mathbf{C}_{k+1} \boldsymbol{\Sigma}_{k+1|k}
 \end{aligned} \tag{5.25}$$

where $\boldsymbol{\Sigma}_{k+1|k} = \mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k$ (see Proposition 5.1)

Proof:

From Equation (3.6),

$$p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1}) = \frac{\int p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k}{\int \int p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k d\mathbf{x}_{k+1}} \quad (5.26)$$

From Equations (5.1) and (5.2), we have

$$\begin{aligned} \mathbf{y}_{k+1} &= \mathbf{C}_{k+1} \mathbf{x}_{k+1} + \mathbf{w}_{k+1} \\ \Rightarrow p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) &= \frac{1}{(2\pi)^{m/2} \sqrt{|\mathbf{R}_{k+1}|}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_{k+1} - \mathbf{C}_{k+1} \mathbf{x}_{k+1})' \mathbf{R}_{k+1}^{-1} (\mathbf{y}_{k+1} - \mathbf{C}_{k+1} \mathbf{x}_{k+1}) \right\} \end{aligned} \quad (5.27)$$

From Equation (5.8) and Proposition 5.1,

$$\begin{aligned} p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_k) &= \int p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \bar{\mathbf{y}}_k) d\mathbf{x}_k \\ &= \frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k|}} \exp \left\{ -\frac{1}{2} [\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k]' (\mathbf{A}_k \boldsymbol{\Sigma}_{k|k} \mathbf{A}_k' + \mathbf{M}_k)^{-1} (\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k) \right\} \\ &= \frac{1}{(2\pi)^{n/2} \sqrt{|\boldsymbol{\Sigma}_{k+1|k}|}} \exp \left\{ -\frac{1}{2} [\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k]' \boldsymbol{\Sigma}_{k+1|k}^{-1} (\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k) \right\} \end{aligned} \quad (5.28)$$

Substituting Equations (5.27) and (5.28) into Equation (5.26), we get

$$\begin{aligned}
p(\mathbf{x}_{k+1}|\bar{\mathbf{y}}_{k+1}) &= \frac{p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}) \int p(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{x}_k|\bar{\mathbf{y}}_k) d\mathbf{x}_k}{\int p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}) \int p(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{x}_k|\bar{\mathbf{y}}_k) d\mathbf{x}_k d\mathbf{x}_{k+1}} \\
&= \frac{\frac{1}{(2\pi)^{m+n/2}} \frac{1}{\sqrt{\|\mathbf{R}_{k+1}\| \|\Sigma_{k+1}\|_A}} \exp\left\{-\frac{1}{2}[(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1})' \mathbf{R}_{k+1}^{-1}(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1}) + (\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)' \Sigma_{k+1|k}^{-1}(\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)]\right\}}{\int \frac{1}{(2\pi)^{m+n/2}} \frac{1}{\sqrt{\|\mathbf{R}_{k+1}\| \|\Sigma_{k+1}\|_A}} \exp\left\{-\frac{1}{2}[(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1})' \mathbf{R}_{k+1}^{-1}(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1}) + (\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)' \Sigma_{k+1|k}^{-1}(\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)]\right\} d\mathbf{x}_{k+1}} \\
&= \frac{\exp\left\{-\frac{1}{2}[(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1})' \mathbf{R}_{k+1}^{-1}(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1}) + (\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)' \Sigma_{k+1|k}^{-1}(\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)]\right\}}{\int \exp\left\{-\frac{1}{2}[(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1})' \mathbf{R}_{k+1}^{-1}(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1}) + (\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)' \Sigma_{k+1|k}^{-1}(\mathbf{x}_{k+1} - \mathbf{A}_k\hat{\mathbf{x}}_k)]\right\} d\mathbf{x}_{k+1}}
\end{aligned}$$

(5.29)

Equation. (5.29) can be rewritten as

$$\begin{aligned}
 p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1}) &= \frac{\exp\left\{-\frac{1}{2}\left[(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1})' \mathbf{R}_{k+1}^{-1}(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1}) + (\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k)' \Sigma_{k+1|k}^{-1}(\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k)\right]\right\}}{\int \exp\left\{-\frac{1}{2}\left[(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1})' \mathbf{R}_{k+1}^{-1}(\mathbf{y}_{k+1} - \mathbf{C}_{k+1}\mathbf{x}_{k+1}) + (\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k)' \Sigma_{k+1|k}^{-1}(\mathbf{x}_{k+1} - \mathbf{A}_k \hat{\mathbf{x}}_k)\right]\right\} d\mathbf{x}_{k+1}} \\
 &= \frac{\exp\left\{-\frac{1}{2}\left[\mathbf{y}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \mathbf{x}_{k+1}' \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} \mathbf{x}_{k+1} - 2\mathbf{x}_{k+1}' \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \mathbf{x}_{k+1}' \Sigma_{k+1|k}^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}_k' \mathbf{A}_k' \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k - 2\mathbf{x}_{k+1}' \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k\right]\right\}}{\int \exp\left\{-\frac{1}{2}\left[\mathbf{y}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \mathbf{x}_{k+1}' \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} \mathbf{x}_{k+1} - 2\mathbf{x}_{k+1}' \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \mathbf{x}_{k+1}' \Sigma_{k+1|k}^{-1} \mathbf{x}_{k+1} + \hat{\mathbf{x}}_k' \mathbf{A}_k' \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k - 2\mathbf{x}_{k+1}' \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k\right]\right\} d\mathbf{x}_1} \\
 &= \frac{\exp\left\{-\frac{1}{2}\left[\mathbf{y}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}_k' \mathbf{A}_k' \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{x}_{k+1}' (\mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1}) \mathbf{x}_{k+1} - 2\mathbf{x}_{k+1}' (\mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k)\right]\right\}}{\int \exp\left\{-\frac{1}{2}\left[\mathbf{y}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}_k' \mathbf{A}_k' \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{x}_{k+1}' (\mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1}) \mathbf{x}_{k+1} - 2\mathbf{x}_{k+1}' (\mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k)\right]\right\} d\mathbf{x}_1}
 \end{aligned} \tag{5.30}$$

Equation (5.30) makes use of the fact that $\mathbf{R}_{k+1}^{-1}$ and $\Sigma_{k+1|k}^{-1}$ are symmetric matrices. Then since $\mathbf{x}_{k+1}' \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1}$ is a scalar,

$\mathbf{x}_{k+1}' \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} = (\mathbf{x}_{k+1}' \mathbf{C}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1})' = \mathbf{y}_{k+1}' \mathbf{R}_{k+1}^{-1} (\mathbf{x}_{k+1}' \mathbf{C}_{k+1}')' = \mathbf{y}_{k+1}' \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} \mathbf{x}_{k+1}$. We can similarly show that

$$\mathbf{x}_{k+1}' \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k = (\mathbf{A}_k \hat{\mathbf{x}}_k)' \Sigma_{k+1|k}^{-1} \mathbf{x}_{k+1}.$$

Define notation $\mathbf{H} = (\mathbf{C}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{C}_{k+1} + \Sigma^{-1}_{k+1|k})$ and $\mathbf{z} = (\mathbf{C}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{y}_{k+1} + \Sigma^{-1}_{k+1|k} \mathbf{A}_k \hat{\mathbf{x}}_k)$. Then Equation (5.30) can be written as

$$\begin{aligned}
 p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1}) &= \frac{\exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}'_k \mathbf{A}'_k \Sigma^{-1}_{k+1|k} \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{x}'_{k+1} \mathbf{H} \mathbf{x}_{k+1} - 2 \mathbf{x}'_{k+1} \mathbf{z} \right] \right\}}{\int \exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}'_k \mathbf{A}'_k \Sigma^{-1}_{k+1|k} \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{x}'_{k+1} \mathbf{H} \mathbf{x}_{k+1} - 2 \mathbf{x}'_{k+1} \mathbf{z} \right] \right\} d\mathbf{x}_{k+1}} \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}'_k \mathbf{A}'_k \Sigma^{-1}_{k+1|k} \mathbf{A}_k \hat{\mathbf{x}}_k + (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) - \mathbf{z}' \mathbf{H}^{-1} \mathbf{z} \right] \right\}}{\int \exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}'_k \mathbf{A}'_k \Sigma^{-1}_{k+1|k} \mathbf{A}_k \hat{\mathbf{x}}_k + (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) - \mathbf{z}' \mathbf{H}^{-1} \mathbf{z} \right] \right\} d\mathbf{x}_{k+1}} \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}'_k \mathbf{A}'_k \Sigma^{-1}_{k+1|k} \mathbf{A}_k \hat{\mathbf{x}}_k - \mathbf{z}' \mathbf{H}^{-1} \mathbf{z} \right] \right\} \exp \left\{ -\frac{1}{2} \left[(\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) \right] \right\}}{\exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_{k+1} \mathbf{R}^{-1}_{k+1} \mathbf{y}_{k+1} + \hat{\mathbf{x}}'_k \mathbf{A}'_k \Sigma^{-1}_{k+1|k} \mathbf{A}_k \hat{\mathbf{x}}_k - \mathbf{z}' \mathbf{H}^{-1} \mathbf{z} \right] \right\} \int \exp \left\{ -\frac{1}{2} \left[(\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) \right] \right\} d\mathbf{x}_{k+1}} \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[(\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) \right] \right\}}{\int \exp \left\{ -\frac{1}{2} \left[(\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) \right] \right\} d\mathbf{x}_{k+1}} \tag{5.31}
 \end{aligned}$$

From Equation (5.12), we have

$$\begin{aligned}
 &\int \frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{H}^{-1}|}} \exp \left\{ -\frac{1}{2} \left[(\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) \right] \right\} d\mathbf{x}_{k+1} = 1 \\
 &\Rightarrow \int \exp \left\{ -\frac{1}{2} \left[(\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1} \mathbf{z}) \right] \right\} d\mathbf{x}_{k+1} = (2\pi)^{n/2} \sqrt{|\mathbf{H}^{-1}|}
 \end{aligned}$$

Hence, Equation (5.30) becomes

$$\begin{aligned}
 p(\mathbf{x}_{k+1} | \bar{\mathbf{y}}_{k+1}) &= \frac{\exp\left\{-\frac{1}{2}[(\mathbf{x}_{k+1} - \mathbf{H}^{-1}\mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1}\mathbf{z})]\right\}}{\int \exp\left\{-\frac{1}{2}[(\mathbf{x}_{k+1} - \mathbf{H}^{-1}\mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1}\mathbf{z})]\right\} d\mathbf{x}_{k+1}} \\
 &= \frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{H}^{-1}|}} \exp\left\{-\frac{1}{2}[(\mathbf{x}_{k+1} - \mathbf{H}^{-1}\mathbf{z})' \mathbf{H} (\mathbf{x}_{k+1} - \mathbf{H}^{-1}\mathbf{z})]\right\}
 \end{aligned} \tag{5.32}$$

This is a Gaussian distribution with the mean as

$$\begin{aligned}
 \hat{\mathbf{x}}_{k+1} &= \mathbf{H}^{-1}\mathbf{z} = (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1})^{-1} (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k) \\
 &= (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1})^{-1} \mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1})^{-1} \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k
 \end{aligned} \tag{5.33}$$

And covariance

$$\begin{aligned}
 \Sigma_{k+1|k+1} &= \mathbf{H}^{-1} = (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1})^{-1} \\
 &= \Sigma_{k+1|k} - \Sigma_{k+1|k} \mathbf{C}'_{k+1} (\mathbf{C}_{k+1} \Sigma_{k+1|k} \mathbf{C}'_{k+1} + \mathbf{R}_{k+1})^{-1} \mathbf{C}_{k+1} \Sigma_{k+1|k}
 \end{aligned} \tag{5.34}$$

Note that

$$\begin{aligned}
 \Sigma_{k+1|k+1} &= (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1})^{-1} \\
 \Rightarrow \Sigma_{k+1|k}^{-1} &= \Sigma_{k+1|k+1}^{-1} - \mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1}
 \end{aligned} \tag{5.35}$$

Substitute Equation (5.35) into (5.33) to get

$$\begin{aligned}
 \hat{\mathbf{x}}_{k+1} &= \mathbf{H}^{-1}\mathbf{z} \\
 &= (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1})^{-1} \mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + (\mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1} + \Sigma_{k+1|k}^{-1})^{-1} \Sigma_{k+1|k}^{-1} \mathbf{A}_k \hat{\mathbf{x}}_k \\
 &= \Sigma_{k+1|k+1} \mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{y}_{k+1} + \Sigma_{k+1|k+1} (\Sigma_{k+1|k+1}^{-1} - \mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{C}_{k+1}) \mathbf{A}_k \hat{\mathbf{x}}_k \\
 &= \mathbf{A}_k \hat{\mathbf{x}}_k + \Sigma_{k+1|k+1} \mathbf{C}'_{k+1} \mathbf{R}_{k+1}^{-1} (\mathbf{y}_{k+1} - \mathbf{C}_{k+1} \mathbf{A}_k \hat{\mathbf{x}}_k)
 \end{aligned} \tag{5.36}$$

Equations (5.34) and (5.36) are the same as the expressions given in Equation (5.25).

End of proof

Proposition 5.3: Assume that the initial guess of the joint pdf of the unknown state variable is

$$p(\mathbf{x}_0 | \mathbf{y}_0) = \frac{1}{(2\pi)^{n/2} \sqrt{|\Sigma_{0|0}|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_0 - \hat{\mathbf{x}}_0)' \Sigma_{0|0}^{-1} (\mathbf{x}_0 - \hat{\mathbf{x}}_0) \right\} \quad (5.37)$$

where $\hat{\mathbf{x}}_0, \Sigma_{0|0}$ are the mean and covariance of the initial guess of the joint pdf. Then, the posterior probability $p(\mathbf{x}_k | \bar{\mathbf{y}}_k)$ at each time step $t = k\tau$ will be Gaussian and the mean and covariance matrix, respectively, given by

$$\begin{aligned} \hat{\mathbf{x}}_k &= \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1} + \Sigma_{k|k} \mathbf{C}_k' \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{C}_k \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1}) \\ \Sigma_{k|k} &= (\mathbf{C}_k' \mathbf{R}_k^{-1} \mathbf{C}_k + \Sigma_{k|k-1}^{-1})^{-1} = \Sigma_{k|k-1} - \Sigma_{k|k-1} \mathbf{C}_k' (\mathbf{C}_k \Sigma_{k|k-1} \mathbf{C}_k' + \mathbf{R}_k)^{-1} \mathbf{C}_k \Sigma_{k|k-1} \end{aligned} \quad (5.38)$$

where $\Sigma_{k|k-1} = \mathbf{A}_{k-1} \Sigma_{k-1|k-1} \mathbf{A}_{k-1}' + \mathbf{M}_{k-1}$ and $\hat{\mathbf{x}}_{k-1}, \Sigma_{k-1|k-1}$ are the mean and covariance matrix for the previous time step.

Proof

For the initial time step, $k = 0$, $p(\mathbf{x}_0 | \mathbf{y}_0)$ is Gaussian from the assumption. Proposition 5.2 with $k = 0$ implies that $p(\mathbf{x}_1 | \mathbf{y}_1)$ is also Gaussian with

$$\begin{aligned} \hat{\mathbf{x}}_1 &= \mathbf{A}_0 \hat{\mathbf{x}}_0 + \Sigma_{1|1} \mathbf{C}_1' \mathbf{R}_1^{-1} (\mathbf{y}_1 - \mathbf{C}_1 \mathbf{A}_0 \hat{\mathbf{x}}_0) \\ \Sigma_{1|1} &= (\mathbf{C}_1' \mathbf{R}_1^{-1} \mathbf{C}_1 + \Sigma_{1|0}^{-1})^{-1} = \Sigma_{1|0} - \Sigma_{1|0} \mathbf{C}_1' (\mathbf{C}_1 \Sigma_{1|0} \mathbf{C}_1' + \mathbf{R}_1)^{-1} \mathbf{C}_1 \Sigma_{1|0} \end{aligned}$$

where $\hat{\mathbf{x}}_1, \Sigma_{1|1}$ denote the covariance matrix of $p(\mathbf{x}_1 | \mathbf{y}_1)$ and $\Sigma_{1|0} = \mathbf{A}_0 \Sigma_{0|0} \mathbf{A}_0' + \mathbf{M}_0$.

Now assume that Proposition 5.3 is true for any $k-1 > 0$. From Proposition 5.2 we have that $p(\mathbf{x}_k | \bar{\mathbf{y}}_k)$ is Gaussian and the mean and covariance matrix, respectively, given by

$$\begin{aligned} \hat{\mathbf{x}}_k &= \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1} + \Sigma_{k|k} \mathbf{C}_k' \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{C}_k \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1}) \\ \Sigma_{k|k} &= (\mathbf{C}_k' \mathbf{R}_k^{-1} \mathbf{C}_k + \Sigma_{k|k-1}^{-1})^{-1} = \Sigma_{k|k-1} - \Sigma_{k|k-1} \mathbf{C}_k' (\mathbf{C}_k \Sigma_{k|k-1} \mathbf{C}_k' + \mathbf{R}_k)^{-1} \mathbf{C}_k \Sigma_{k|k-1} \end{aligned}$$

which is the same as Equation (5.38). So, Proposition 5.3 is true for all $k \geq 0$ by induction.

End of Proof

Proposition 5.3 shows that if the initial guess of probability distribution is Gaussian, the posterior distribution $p(\mathbf{x}_k | \bar{\mathbf{y}}_k)$ at every time step is also Gaussian with the posterior mean and covariance are recursively determined by Equation (5.38)

Equation (5.38) is known as Riccati equation⁴⁵ which is implemented in Kalman filter approach to estimate unknown state variables. Hence, for discrete linear system with Gaussian noise, the continuous Bayesian estimator given by Equation (3.6) is equivalent to Kalman filter which implies that, the extended DSD algorithm, which is the discrete form of Equation (3.6), is also equivalent to Kalman filter when the cell-size is sufficiently small.

5.2 Numerical Examples

Section 5.1 shows that in the case of discrete linear system with measurement and model uncertainty as white Gaussian noise, Kalman filter is equivalent to the continuous Bayesian estimator. As a discrete type Bayesian estimator, the DSD uses cell-to-cell mapping technique to approximate the continuous probability distribution function by a discrete type distribution function. In this section, DSD is compared with Kalman filter through a discrete linear application. Two concerns are considered: a) How is the discrete posterior probability distribution compared with the continuous pdf? b) Is the approximation error stable with respect to time?

Consider the following linear system,

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{v}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{w}_k \end{aligned} \quad (5.39)$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are zero mean Gaussian noise with covariance matrixes $\mathbf{R}$ and $\mathbf{M}$, respectively. In order to address Concern (a) above, first visually compare the pdfs. Fig. 5.1 shows both the discrete and continuous pdfs at time step $k=10$ for $\mathbf{A}=1, \mathbf{C}=1, \mathbf{M}=1, \mathbf{R}=4$ which correspond to a simple oscillating system. The continuous pdf is obtained from Kalman filter using Equation (5.38) and the discrete pdf is obtained through the extended DSD method described in Chapter 3 (Equations (3.23) and (3.24)) using the partitioning data shown in Table 5.1.

Unknown Variables	Possible Maximum Value	Possible Minimum Value	Number of Cells	Number of Points Selected In the Cell (see Equation (3.24))
x	45	-5	200	1

Table 5.1: Partitioning Data Used for Case $\mathbf{A}=1, \mathbf{C}=1, \mathbf{M}=1, \mathbf{R}=4$

Fig. 5.1: Actual Vs. Approximate Posterior Probability Distribution at Time Step $k = 10$. (solid line indicates continuous pdf, dashed line indicates discrete pdf)

From Fig. 5.1, it is shown that the discrete pdf closely approximates the continuous pdf. Checks on other time steps show that the closeness in approximation is consistent and the approximation error is stable with respect to time steps.

For a linear system with Gaussian noise, the posterior probability distribution $p(x_k | \bar{y}_k)$ is also Gaussian (Proposition 5.3), which is uniquely determined by its mean and variance. Hence, the obvious quantitative measure of the approximation error is the difference in posterior mean and variance between the continuous and discrete distribution functions at each time step. Fig. 5.2 shows the plot of the relative error
$$\left(\frac{\text{Posterior Mean/Variance By DSD} - \text{Posterior Mean/Variance By Kalman Filter}}{\text{Posterior Mean/Variance By Kalman Filter}} \times 100\% \right)$$

in posterior mean and variance as a function of time step. Again, it is shown that approximation is good and approximation error is stable.

(a)

(b)

Fig. 5.2: Relative Error In Posterior Mean and Variance Between the Actual and Approximate Probability Distribution Function At Each Time Step
($A = 1, C = 1, M = 1, R = 4$):

(a) Relative Error in Posterior Mean, (b) Relative Error in Posterior Variance

To explore the behavior of the DSD algorithm under different system dynamics and different noise variances, other cases has also been run for stable and unstable systems with different noise variance. Table 5.2 and 5.3 lists the partitioning data for stable ($A = 0.9, C = 1, M = 9, R = 4$) and unstable ($A = 1.05, C = 1, M = 4, R = 1$) systems, respectively. Figures 5.3 and 5.4 shows the comparison results. Once again, it is concluded that the approximation is good and reasonably stable for both stable and unstable system dynamics. The peaks in Fig. 5.3 (a) and Fig. 5.4 (a) are due to the fact that at some time steps, the mean value of the posterior probability distribution is very close to zero.

The above numerical examples show that DSD can approximate the continuous posterior distribution very well for linear discrete system with Gaussian noise. The approximation error is surprisingly small (relative error is less than 10^{-6}) even with small number of cells. On the other hand, Kalman filter (or extended Kalman filter for nonlinear systems) only provide the posterior mean and variance, and not the distribution itself. Since the posterior probability distribution function is not necessarily Gaussian for non-linear systems or linear systems with non-Gaussian noise, an advantage of the DSD approach is that it provides additional information that the Kalman filter approach cannot provide.

Unknown Variables	Possible Maximum Value	Possible Minimum Value	Number of Cells	Number of Points Selected In the Cell (see Equation (3.24))
x	45	-25	100	1

Table 5.2: Partitioning Data Used for Case $A = 0.9, C = 1, M = 9, R = 4$

Unknown Variables	Possible Maximum Value	Possible Minimum Value	Number of Cells	Number of Points Selected In the Cell (see Equation (3.24))
x	280	-20	300	1

Table 5.3: Partitioning Data Used for Case $A = 1.05, C = 1, M = 4, R = 1$

(a)

(b)

Fig. 5.3: Relative Error In Posterior Mean and Variance Between the Actual and Approximate Probability Distribution Function At Each Time Step
($A = 0.9, C = 1, M = 9, R = 4$):

(a) Relative Error in Posterior Mean, (b) Relative Error in Posterior Variance

(a)

(b)

Fig. 5.4: Relative Error In Posterior Mean and Variance Between the Actual and Approximate Probability Distribution Function At Each Time Step
($A = 1.05, C = 1, M = 4, R = 1$):

(a) Relative Error in Posterior Mean, (b) Relative Error in Posterior Variance

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This dissertation achieved all its objectives stated in Section 1. A new theoretical basis is developed for DSD in Chapter 3 which

- indicates a close relationship between DSD and other conventional parameter/state variable estimation techniques, which is presented in Chapter 4 and 5,
- leads to a improved, more general estimation procedure, and ,
- provides a general closed form expression of transition probability which lays the groundwork for future study of posterior probability approximation error analysis.

The theoretical basis presented in Chapter 3 leads to the development of a new estimation procedure, the extended DSD algorithm. Compared with the original form of DSD, the extended algorithm achieved the following improvements:

1. It allows arbitrary noise distributions in the measured data as well as arbitrary uncertainties in the system model.
2. It can handle arbitrary non-linear observers.
3. It achieves higher estimation speed and accuracy.
4. It enables DSD to be capable of validating the noisy measurements.

Chapters 4 and 5 present the relationship between DSD and several other conventional parameter/unknown state estimators. Compared with conventional parameter/state estimation techniques, it is shown that the DSD can provide more useful information about the unknown parameters and unknown state variables by providing the joint pdf for the quantities to be determined. The joint pdf can then be used to get all the statistical properties of these quantities such as expected values, variances and credibility intervals. Theoretical and numerical studies in these chapters show that DSD is closely related to General Maximum Likelihood Estimator, Least Square Estimator and Kalman Filter. More specifically:

- When only unknown constant parameters are being estimated and the measurement noise is white Gaussian noise, the DSD is equivalent to the GMLE and LSE under the condition that the cell size is chosen small enough. The equivalency is in the sense that the parameter values maximizing the ratio

$$\frac{p(\theta|\bar{y}_{k+1})}{p(\theta|\bar{y}_0)} \text{ (or when } p(\theta|\bar{y}_0) \text{ is uniform, the mode of the discrete posterior pdf)}$$

obtained through DSD) will approach the GMLE estimate of θ and also minimize the Least Square Error criterion.

- For linear discrete system, DSD is equivalent to Kalman filter.

6.2 Recommendations for Future Work

For future research, the following are recommended,

- *Firm Guidelines for Selecting Partitioning Data*

Currently, there are no firm guidelines to select the best partitioning data. A practical approach is to start with a small number of cells and increase it until reasonable results is obtained. Systemic ways of choosing optimum partitioning data could be obtained through approximation error analysis in calculating the transition probability.

- *Transition Probability Approximation Error Analysis*

The transition probability defined in Chapter 3 is evaluated through numerical techniques. The approximation error comes from the fact that we are partitioning the state space and also we are using quadrature scheme to approximate the integral in Equation (3.22). These errors will affect the convergence property and the accuracy of the final estimation results. A transition error analysis will not only be able to predict the convergence property of the estimation procedure, but also can be used in partitioning data selection.

REFERENCES

1. L.Dinca, T.Aldemir, "Parameter Estimation Towards Fault Diagnosis in Non-Linear Systems Using a Markov Model of System Dynamics," *Nuclear Science and Engineering*, **127**, 199-219 (1997).
2. Laurian G. Dinca, "A Probabilistic Approach to Parameter Estimation Towards Fault Diagnosis in Non-linear Dynamic Systems", PhD Dissertation, The Ohio State University (1997).
3. D. J. Mook, "Estimation and Identification of Nonlinear Dynamic Systems", *AIAA Journal*, **27**, 968 (1989).
4. J.S.Albuquerque, L.T.Biegler, "Decomposition Algorithm for On-line Estimation with Nonlinear Models", *Computers & Chemical Engineering*, **19**, No.10, 1031-1039 (1995).
5. M. Guay, D. D. McLean, "Optimization and Sensitivity Analysis for Multiresponse Parameter Estimation in Systems of Ordinary Differential Equations," *Computers & Chemical Engineering*, **19**, 1271, (1995).
6. Box G.E., N.Draper, "The Bayesian Estimation of Common Parameters", *Biometrika*, **52**, 355-365 (1962).
7. K. D. Hjelmstad, M. R. Banan, "Time-domain Parameter Estimation Algorithm for Structures I: Computational Aspects." *Journal of Engineering Mechanics*, **121**, 424, (1995).
8. Zheng-She Liu, "QR Methods of $O(N)$ Complexity in Adaptive Parameter Estimation," *IEEE Transactions on Signal Processing*, **43**, 720 (1995).
9. D.Dimogianopoulos, R.Lozano, "Adaptive Control for Linear Slowly Time Varying Systems Using Direct Least-Squares Estimation", *Automatica*, **37**, 251-256 (2001).
10. Reza Ghanadan, G.L.Blakenship, "Adaptive Control of Nonlinear Systems via Approximate Linearization," *IEEE Transactions on Automatic Control*, **41**, 618 (1996).

11. G. E. Young, K. S. Suresh, "Block -recursive Identification of Parameters and Delay in the Presence of Noise," *Journal of Dynamic Systems, Measurement and Control*, **117**, 600 (1995).
12. M. A. Hopkins, H.F. VanLandingham, "Optimal Nonlinear Estimation of Linear Stochastic Systems," *Journal of Dynamic Systems, Measurement and Control*, **116**, 529 (1994).
13. M. A. Hopkins, H. F. VanLandigham, "Optimal Nonlinear Estimation of Linear Stochastic Systems: the Multivariate Extension," *Journal of Dynamic Systems, Measurement and Control*, **118**, 350 (1996).
14. J. L. Douce, Q. M. Zhu, "Modeling a Class of Nonlinear Systems", *IEEE Proceedings*, **137**, 385 (1990).
15. Hai-Wen Chen, "Modeling and Identification of Parallel Nonlinear Systems: Structural Classification and Parameter Estimation Methods," *Proceedings of the IEEE*, **83**, 39 (1995).
16. Sergey Lyashevskiy, Yaobin Chen, "Identification of Nonlinear System Parameters Using Describing Function Approach", *Proceedings of the American Control Conference*, Albuquerque, New Mexico (1997).
17. M. Ricardo, P. Tomei, "Adaptive Observers with Arbitrary Exponential Rate of Convergence for Nonlinear Systems," *IEEE Transactions on Automatic Control*, **40**, 1300 (1995).
18. Chien S. Hsieh, Fu C. Chen, "Modified Stochastic Luenberger Observers", *Automatica*, **36**, 1847-1854 (2000).
19. Roy A. McCann. Mohammad S. Islam, Iqbal Husain, "Application of a Sliding-Mode Observer for Position and Speed Estimation in Switched Reluctance Motor Drives", *IEEE Transactions on Industry Applications*, Vol. **37**, No. 1 (2001).
20. P.Wang, T.Aldemir, Vadim I. Utkin, "Estimation of Xenon concentration and Reactivity in Nuclear Reactor Using Sliding Mode Observer", *40th IEEE Conference on Decision and Control*, Orlando, Florida, USA (Dec, 2001).
21. Yong-Wha Kim, Giorgio Rizzoni, Vadim Utkin, "Automotive Engine Diagnosis and Control via Nonlinear Estimation", *IEEE Control Systems* (October 1998).

22. C.E.D'Attellis, E.Cortina, "Simultaneous Estimation of Neutron Density and Reactivity in a Nuclear Reactor Using a Bank of Kalman Filters", *Nuclear Science and Engineering*, **105**, 297 (1990).
23. Luigi Glielmo, Roberto Setola, Francesco Vasca, "An Interlaced Extended Kalman Filter", *IEEE Transactions on Automatic Control*, Vol. **44**, No. 8, (1999).
24. M.Boutayeb, "Identification of Nonlinear Systems in the Presence of Unknown but Bounded Disturbances", *IEEE Transactions on Automatic Control*, Vol. **45**, No. 8 1503-1507(2000).
25. Erik Bolviken, Peter J.Acklam, Nils Christophersen, John-Mikal Stordal, "Monte Carlo Filters for Non-linear State Estimation", *Automatica*, **37**, 177-183 (2001).
26. Charalambos D. Charalambous, Andrew Logothetis, "Maximum Likelihood Parameter Estimation from Incomplete Data via Sensitivity Equations: The Continuous-Time Case", *IEEE Transactions on Automatic Control*, Vol. **45**, No.5 (2000).
27. A. G. Parlos, A. F. Atiya, "Nonlinear Identification of Process Dynamic Using Neural Networks," *Nuclear Technology*, **97**, 79 (1992).
28. R.J.Patton, C.J.Lopez-Toribio, F.J.Uppal, "Artificial Intelligence Approaches To Fault Diagnosis", *The Institute of Electrical Engineers*, IEEE, Savoy Place, London, UK(1999).
29. Marzio Marseguerra, Enrico Zio, "Genetic Algorithm for Estimating Effective Parameters In a Lumped Reactor Model For Reactivity Predictions", *Nuclear Science and Engineering*, **139**, 66-81 (2001).
30. X. M.Chen, P.Wang, T.Aldemir, "DSD; A Parameter/State Estimation Tool For Model-based Fault Diagnosis in Non-linear Dynamic Systems", *Dynamic Reliability: Future Directions, International Workshop Series on Advanced Topics in Reliability And Risk Analysis*, Center for Reliability Engineering, University of Maryland (1999).
31. P.Wang, "DSD:A Parameter/State Estimation Tool for Model-Based Fault Diagnosis in Non-linear Dynamic Systems", Master Thesis, The Ohio State University (2000).
32. P.Wang, T.Aldemir, D.W.Miller, " Reactivity Ramp Rate Estimation Using A Probabilistic Approach", *Transactions of American Nuclear Society*, **78**, 121-123 (1998).

33. P.Wang, T.Aldemir, "Real-Time Xenon Estimation in Nuclear Power Plants", *Transactions of American Nuclear Society*, **81**,154-156 (1999).
34. Peng Wang, T.Aldemir, E.Zio, "An On-Line Tool for Risk-based Accident Management", *Safety and Reliability*, G.I.Schueller, P. Kafka (Eds), 439-444, A.A.Balkema, Rotterdam (1999).
35. Henry Stark, John W. Woods, *Probability, Random Process and Estimation Theory For Engineers*, 2nd Edition, Chap.11.6, Prentice-Hall Inc, Upper Saddle River, New Jersey (1994).
36. C.W.Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Chap. 1.2, Springer-Verlag, New York (1983).
37. George Casella, Roger L. Berger, *Statistical Inference*, Page 265, Duxbury Press, Belmont, California (1990).
38. Pieter Eykhoff, *System Identification, Parameter and State Estimation*, Page 145, John Wiley & Sons, London (1974).
39. Thomas. D. Radcliff, Don. W. Miller, Andrew. C. Kauffman, "Modeling Of A Constant-Temperature Power Sensor", *Proceedings of the 8th International Conference on Nuclear Engineering*, Baltimore, MD, USA (April 2000).
40. Thomas D. Radcliff, Don. W. Miller, Andrew. C. Kauffman, "Constant-Temperature Calorimetry For In-Core Power Measurement", *Nuclear Technology*, Vol. **132**, 240-255 (Nov. 2000).
41. Hanying Liu, *Temperature Compensation of the CTPS In-core Power Sensor*, Research Report, Nuclear Engineering Program, Ohio State University (2001).
42. Jeffrey R. Fluhrer, "The System Dynamics And Control Of A Constant Temperature Local Power Sensor", Master Thesis, The Ohio State University (1997).
43. George Casella, Roger L. Berger, *Statistical Inference*, Page 567, Duxbury Press, Belmont, California (1990).
44. Harold W. Sorenson, *Parameter Estimation*, Page 57, Marcel Dekker, INC., New York (1980).
45. Pieter Eykhoff, *System Identification, Parameter and State Estimation*, John Wiley & Sons, London (1974).