9-10-2015

Particle filtering for EEG source localization and constrained state spaces

Bradley Ebinger

PARTICLE FILTERING FOR EEG SOURCE LOCALIZATION AND CONSTRAINED STATE SPACES

by

Bradley Ebinger

A Thesis

Submitted to the
Department of Electrical and Computer Engineering
College of Engineering
In partial fulfillment of the requirement
For the degree of
Master of Science in Electrical and Computer Engineering
at
Rowan University
August 31, 2015

Thesis Chair: Nidhal Bouaynaya, Ph.D.

Dedications

To my mother, father, and brother who have supported me throughout my journey.

Acknowledgments

“What you need, above all else, is a love for your subject, whatever it is. You've got to be so deeply in love with your subject that when curve balls are thrown, when hurdles are put in place, you've got the energy to overcome them."

- Neil deGrasse Tyson

I would like to thank my graduate advisor Dr. Nidhal Bouaynaya for her constant guidance and support over the last three years. When one puts their faith in another, it can become a tremendous pressure that can either strengthen or crush them. In this case, Dr. Bouaynaya has put pressure on me to become a better researcher, a better thinker, and has instilled a confidence that has allowed me to overcome all of the hurdles that have been put in my way. I would also like to thank Dr. Robi Polikar, Dr. Ravi Ramachandran, Dr. Lyudmila Mihaylova, and Dr. Petia Georgieva for agreeing to be part of my committee and extending me help throughout my graduate studies and research. Finally, I would like to thank my family, who has supported me throughout this entire process. This thesis is not only my work, but a collaboration; I would not be where I am today without the support of you all.

Thank you.

Abstract

Bradley Ebinger
PARTICLE FILTERING FOR EEG SOURCE LOCALIZATION AND CONSTRAINED STATE SPACES
2014-2015
Nidhal Bouaynaya, Ph.D.
Master of Science in Electrical and Computer Engineering

Particle Filters (PFs) have a unique ability to perform asymptotically optimal estimation for non-linear and non-Gaussian state-space models. However, the numerical nature of PFs cause them to have major weakness in two important areas: (1) handling constraints on the state, and (2) dealing with high-dimensional states. In the first area, handling constraints within the PF framework is crucial in dynamical systems, which are often required to satisfy constraints that arise from basic physical laws or other considerations. The current trend in constrained particle filtering is to enforce the constraints on all particles of the PF. We show that this approach leads to more stringent conditions on the posterior density that can cause incorrect state estimates. We subsequently describe a novel algorithm that restricts the mean estimate without restricting the posterior pdf, thus providing a more accurate state estimate. In the second area, we tackle the "curse of dimensionality," which causes the PF to require an exponential increase in computational complexity as the dimension of the state increases. The application of interest is localization of the brain neural generators that create the Electroencephalogram (EEG) signal. Specifically, we describe a state-space model that tracks the position and moments of multiple dynamic dipoles and apply the marginalized PF, which alleviates the "curse of dimensionality" for tracking multiple dynamic dipoles. This modified framework allows us to consider dynamic dipoles, which were historically considered time-invariant.

Table of Contents

Abstract v

List of Figures viii

List of Tables x

Chapter 1: Introduction 1

1.1 Motivation, Background, and Problem Statement 1

1.1.1 The EEG Source Localization Problem. 4

1.1.2 Particle Filtering in Constrained State-Spaces. 5

1.2 Research Contributions 6

1.3 Organization 8

Chapter 2: Literature Review 10

2.1 Problem Statement 10

2.2 Optimal State Estimation in Linear Models 13

2.3 Approximate Solutions in Nonlinear Models 15

2.3.1 The Extended Kalman Filter. 15

2.3.2 The Unscented Kalman Filter. 17

2.4 The Particle Filter Framework 18

2.4.1 Monte Carlo Sampling. 18

2.4.2 Importance Sampling. 19

2.4.3 Sequential Importance Sampling. 20

2.4.4 The Particle Filter. 24

Chapter 3: The Constrained Particle Filter 29

3.1 Introduction 29

Table of Contents (Continued)

3.2 The Unconstrained Particle Filter 30

3.3 The Constrained Particle Filter 32

3.3.1 Pointwise Density Truncation (PDT) 33

3.3.2 Mean Density Truncation (MDT). 34

3.4 Simulation Results 36

Chapter 4: The Marginalized Particle Filter
Application to EEG Dynamic Source Localization 40

4.1 Introduction 40

4.2 EEG Source Localization Model 43

4.3 The Particle Filter 45

4.4 The Marginalized Particle Filter 47

4.5 Results and Discussion 51

4.5.1 Simulation Results on Synthetic Data 51

4.5.2 Application to Real EEG Data 52

Chapter 5: Conclusion and Future Perspectives 64

References 67

List of Figures

Figure Page

Figure 1. A comparison of the three major methods used for tracking the state of a dynamical system: Particle Filter, Extended Kalman Filter, and Unscented Particle Filter [1]. 3

Figure 2. A graphical representation of the model in (2.1) and (2.2). Note the following two assumptions on the system: the current observation $\boldsymbol{y}_k$ depends only on the current state $\boldsymbol{x}_k$ and the state $\boldsymbol{x}_k$ is a Markov process, i.e., it depends only on the previous state $\boldsymbol{x}_{k-1}$.....11

Figure 3. An approximation of the PDF $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ using samples $\boldsymbol{x}^{(i)}$ generated from $\boldsymbol{x}^{(i)} \sim q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}, \boldsymbol{Y}_k)$. These samples are weighted and the result is an estimate of the posterior pdf $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$. 24

Figure 4. The generic particle filter algorithm. 27

Figure 5. Illustration of the PDT approach for an interval-constrained system, $\boldsymbol{x}_k \in [70, 170]$ for all k. The true posterior density (green curve) is multimodal with mean 110 (green x-mark). If all particles are constrained to be within the interval [70, 170], then the estimated posterior density (red curve) will be a truncated exponential density that is dramatically different from the true posterior distribution. 33

Figure 6. Constrained state-estimation of the nonlinear dynamic system in (3.11). State estimation for 1000 Monte Carlo simulations. The shading represents a two standard deviations band. Top row: 1st-order MDT (MSE=3.58, σ = 0.83); bottom row: PDT (MSE=5.36, σ = 2.03. 38

Figure 7. Constrained state-estimation of the nonlinear dynamic system in (3.11). State posterior densities evolving over time for (a) 1st-order MDT and (b) PDT. 39

Figure 8. The marginalized particle filter algorithm. 50

Figure 9. Tracking of a one dipole: Trajectory (top) and moment (bottom). Top figure: marginalized PF and bottom figure: classical PF. The solid lines represent the true state and the circles represent the estimated state for each dimension. The x-dimension is in red, the y-dimension in green, and the z-dimension in blue. 55

Figure 10. Mean Squared Error of tracking one dipole: Location (top) and moment (bottom). The Marginalized Particle filter is in blue and the classical Particle Filter is in red. 56

Figure 11. Tracking of two dipoles: first dipole trajectory (top) and moments (bottom). Top figure: marginalized PF and bottom figure: classical PF. The solid lines represent the true state and the circles represent the estimated state for each dimension. The x-dimension is in red, the y-dimension in green, and the z-dimension in blue. 57

Figure 12. Tracking of two dipoles: second dipole trajectory (top) and moments (bottom). Top figure: marginalized PF and bottom figure: classical PF. The solid lines represent the true state and the circles represent the estimated state for each dimension. The x-dimension is in red, the y-dimension in green, and the z-dimension in blue. 58

Figure 13. Mean Squared Error of Tracking the position (top) and moment (bottom) of the first dipole (top figure) and second dipole (bottom figure). The Marginalized PF tracking is in blue and and the classical PF tracking is in red. 59

Figure 14. The head model: (a) Depiction of a realistic EEG experiment; (b) Spatial scalp location of the EEG electrodes. 60

Figure 15. A set of images shown to the subjects in the experimental setup. 60

Figure 16. Dipoles and moments tracking over 3 trials for subject one. The first row shows tracking of the locations of the dipoles. The second row shows the dipole moments over time. The average location of dipole 1 is ($\bar{x}_1^1 = 0.83$, $\bar{y}_1^1 = -1.67$, $\bar{z}_1^1 = 1$) and dipole 2 is ($\bar{x}_2^1 = 0.37$, $\bar{y}_2^1 = -2$, $\bar{z}_2^1 = 0.9$). 61

Figure 17. Dipoles and moments tracking over 3 trials for subject two. The first row shows tracking of the locations of the dipoles. The second row shows the dipole moments over time. The average location of dipole 1 is ($\bar{x}_1^2 = 0.2$, $\bar{y}_1^2 = -1.25$, $\bar{z}_1^2 = 2.3$) and dipole 2 is ($\bar{x}_2^2 = -0.07$, $\bar{y}_2^2 = -1$, $\bar{z}_2^2 = 2.2$). 62

Figure 18. Axial view of primary visual cortex zone. The arrows point at the estimated source locations.....63

List of Tables

Table Page

Table 1. Moving dipoles starting and ending positions (in meters) in the brain 52

Table 2. Moving dipoles moment amplitude and frequency 53

Chapter 1

Introduction

1.1 Motivation, Background, and Problem Statement

Many problems in engineering require estimation of the state of a system that changes over time given noisy measurements made on the system. Examples include tracking targets, e.g., aircraft and missiles using radar, and robot localization and map building from range sensors. The state of the system is also called *hidden state* because it is not directly measurable. In this thesis, we use the state-space approach to model discrete-time dynamical systems because this approach is convenient for handling multivariate data. In order to analyze and make an accurate inference about a dynamic system, at least two models are required: First, a model describing the a priori evolution of the state with time (system model), and second, a model relating the measurements to the state of the system (observation model). In addition, since the evolution of the state is known approximately and the measurements are noisy, the state-space model (comprised of the system model and the observation model) is probabilistic in nature. The state-space model provides a rigorous general approach for state estimation in dynamical systems.

In a probabilistic framework, all information about the state $\boldsymbol{x}_k$ at time instant k given the history of observations up to time k, $\boldsymbol{Y}^k = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k\}$, is contained in the posterior distribution $p(\boldsymbol{x}_k|\boldsymbol{Y}^k)$. If the system is linear (both the system and measurement models are linear) and the noise is Gaussian, then the Kalman filter provides an optimal estimate of the state. The optimality condition minimizes the mean square error of the estimated state. The Kalman filter tracks the state at every time step by updating the mean and covariance of the posterior distribution. If the system is linear but the noise is not Gaussian, then the

Kalman filter provides the *optimal linear estimator* of the state. However, better non-linear estimators may exist. For systems with non-linear dynamics, several approximations can be used. The Extended Kalman Filter (EKF) uses a first-order Taylor series approximation to linearize the model around its current estimate [1]. EKF works well for nearly-linear state space models, but fails in "highly" non-linear systems, where a first order approximation does not reasonably describe the dynamics of the system. The second approximation that can be used in non-linear state-space models is the Unscented Kalman Filter (UKF). The UKF uses a technique called the *Unscented Transform*, which utilizes a deterministic set of samples, called *sigma points*, to propagate the posterior mean and covariance [2]. This technique is very useful because there are no Jacobians that need to be computed at every step. The problem, however, is that the UKF will only perform better than the EKF in cases of non-linearity or non-Gaussianity when certain arbitrary free parameters are tuned, which can make it hard to use in a general case [2].

The great descriptive power of state-space models comes at the expense of intractability: it is impossible to obtain analytic solutions to the estimation problems except for few special cases, including the linear and Gaussian case. A class of numerical algorithms, called *Particle Filters*, has become a very popular class of numerical methods for the solution of optimal estimation problems in non-linear and non-Gaussian dynamical systems. The Particle Filter (PF) is a Monte Carlo technique that uses *Sequential Importance Sampling* to estimate the posterior distribution of the state at every time step. The PF uses a set of *particles* to sample the state-space of the system. These particles are then weighted using the observation model to provide an estimate of the state posterior density. It can be shown that the estimation converges, in the mean-square error, to the true posterior density of the

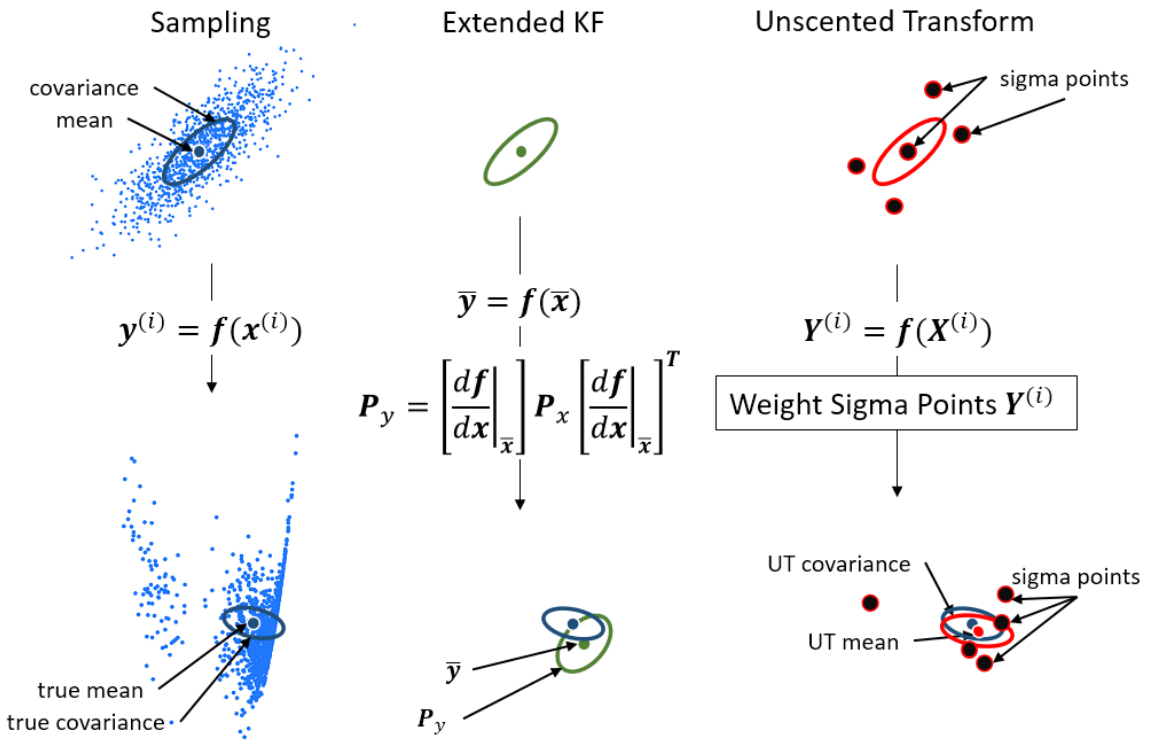


Figure 1. A comparison of the three major methods used for tracking the state of a dynamical system: Particle Filter, Extended Kalman Filter, and Unscented Particle Filter [4].

state [3]. The main idea of PFs may sound similar to the procedure of the UKF; however the UKF uses a deterministic set of sigma points to sample the state-space (whereas the PF uses randomly generated particles) and a different method to weight the sigma points. A comparison of the approximation and numerical approaches are provided in Figure 1.

While the Particle Filter can infer the state of non-linear dynamical systems with arbitrary noise statistics and without relying on any assumptions or approximations, its main drawback is computational. The PF is computationally expensive. However, thanks to the availability of increasing computational hardware, the PF has been used in real-time in many applications, including target tracking [5, 6], robotics [7], financial econometrics [8, 9] and chemical engineering [10]. This being said, the PF remains ineffective in

high-dimensional state-spaces. Specifically, in order for the Particle Filter to converge to the correct result, the number of particles used must increase super-exponentially with the dimension of the state [11]. This problem is commonly referred to as the "curse of dimensionality" in the literature, and makes the particle filter restrictively expensive in high-dimensional state-spaces. If computationally inexpensive solutions must be made, it can be very hard to implement a Particle Filter solution that solves the problem in real time.

1.1.1 The EEG Source Localization Problem. Electroencephalography (EEG) is a widely used technology in neurology because it is non-invasive, portable, low cost, and has a high temporal resolution. Some of the most recent clinical applications using EEG are the Brain Computer Interface (BCI) technologies, which have the potential to interface the brain directly using the EEG measurements as inputs. Another potential application that uses EEG is the diagnosis of specific electro-clinical systems, such as epilepsy. In order to better understand how the brain works, we must first understand how the EEG measurements are generated at the scalp. The main sources of EEG potentials, which are measured at the scalp, emanate from the simultaneous current flows of many neighboring neurons in the same direction. The total electrical current in an activated region of the brain is often modeled as a mathematical current dipole with an adequate dipole moment at that location. Also many of these current dipoles representing current flows with the same orientation can be replaced by an equivalent current dipole [12].

Currently two major research areas exist in modeling neural generators [13]. The first modeling technique involves utilizing imaging models, which explain the data using a dense set of current dipoles distributed at fixed locations. The second method is a para-

metric approach that takes advantage of the fact that these dense sets of current dipoles can be replaced using one equivalent current dipole. Although the imaging-based techniques can create a detailed map of the brain's neuronal activity, the parametric approach provides a direct mapping of the EEG measurements to a small number of parameters. By using the parametric approach, the equivalent current models can provide more intuitive interpretations that explain the electrical activity in the brain and can be fostered in emerging technologies, such as BCI systems [14].

An important challenge of the parametric approach is the estimation of the locations of the equivalent dipole sources in the 3D volume of the brain using EEG measurements recorded from the scalp. Most of the previous work in EEG source localization assumes fixed dipoles in the brain and cannot handle the case of moving EEG sources. Neurological studies, however, suggest that sources vary in the brain (in terms of number, location and signal) depending on various internal and external stimuli. This thesis presents a shift in the current paradigm by tackling the problem of estimation of dynamic sources in the brain. Specifically, we show that the particle filter framework can be used to track the locations and signals of moving EEG dipoles in the brain. Moreover, our study considers low signal-to-noise ratio (less than 5 dB) scenarios, in contrast to the high SNR values (100 dB) assumed in [15]. Moreover, we present a marginalized particle filtering framework, whereby we can handle the increased dimension of the state due to considering multiple dipole sources in the brain.

1.1.2 Particle Filtering in Constrained State-Spaces. In this thesis, we also extend the particle filtering framework to include constraints on the state that are not accounted for

in the state-space model. The state of many dynamical systems is often required to satisfy certain constraints arising from basic physical laws, mathematical properties or geometric considerations, e.g., maximum power or transmission capacity, energy conservation laws and bounded parameters.

The very numerical nature of the particle filters, which constitutes their strength for multidimensional numerical integration, becomes their major weakness in handling constraints on the state, however. The main difficulty of the constrained PF problem stems from the fact that every particle in the particle approximation of the state posterior density is a local representation of the density, and thus cannot characterize global properties of the density, such as constraints on the conditional mean or any other functional expectation. The current trend in constrained particle filtering simply enforces the constraints on all particles of the PF. This approach, however, constrains the posterior density of the state rather than its mean, which leads to more stringent conditions and possibly a completely different condition than the original constraints. In this thesis, we present an approach that relaxes the constraint on the posterior density of the state, while still keeping the original constraints on the mean estimate intact.

1.2 Research Contributions

This thesis presents two main contributions that are related to the field of Particle Filtering. The first contribution is in the field of constrained Particle Filtering. While there has been some work in constraining the mean estimate of the state at each time step, there has yet to be a method that does not impose stronger constraints on the posterior distribution of the state or the noise model of the system. We propose a new method for constrained Par-

ticle Filtering, called Mean Density Truncation, which relaxes the constraint imposed on the posterior distribution by previous methods and instead only places the constraint on the mean estimate as desired in the original formulation of the problem. Further, we show that in certain cases the previous methods of constraining the state estimate can actually lead to incorrect estimates. The second contribution is in the field of neurotechnology. We propose a solution that uses the Particle filter to track the brain sources of EEG measurement neural generators. While previous methods involved using methods that required either prior knowledge of the location of the neural generators or assumed stationary neural generators, the proposed method has neither of these constraints and tracks the brain sources as they move inside the brain depending on the activity and stimuli imposed on the subject. By exploiting the linear substructures that are in the state-space model, a technique called *Marginalized Particle Filtering* can be used to estimate the 3D position and moments of each dipole even though the dimension of the state-space would normally require a much larger number of particles. Specific contributions of this work include:

- Proposing a new method of constrained particle filtering, called *Mean Density Truncation*, that computes the estimate of the state that satisfies the constraint without imposing stronger conditions on the posterior distribution of the state. We also show that the proposed mean density truncation techniques leads to a smaller estimation error than the widely used *pointwise density truncation* technique.

- Formulating a new approach to EEG source localization using a non-linear state-space model and using the particle filter to track the moving dipoles and their corresponding moments in the 3D volume of the brain.

- Using the Marginalized Particle Filter (MPF), which takes advantage of the linear substructures in the EEG state-space model in order to reduce the dimension of the state estimated by the particle filter (PF). It is also shown that The MPF optimally estimates the linear part of the state using the Kalman Filter, while still tracking the non-linear components using the PF.

- Applying the MPF algorithm to real EEG data and showing that the obtained results correspond to the regions of the brain that are expected to be active in the experiment.

1.3 Organization

This thesis is organized as follows.

In Chapter 2, we provide a tutorial on Bayesian state estimation, as well as the fundamentals of particle filtering.

In Chapter 3, we tackle the constrained state-space estimation problem and propose a novel method in constrained Particle Filtering, called *Mean Density Truncation*, which provides an estimate that satisfies the desired constraints on the state estimate without imposing further conditions on the dynamical system.

In Chapter 4, we formulate the EEG source localization problem as a state-space model and consider simultaneous estimation of the dipoles locations and moments. In addition, we consider non-stationary dipoles that can move inside the brain. We then elaborate on the Marginalized Particle Filtering approach to estimate the high-dimensional state of the system with a reasonable computational cost. The application to real EEG data is shown and discussed.

Chapter 2 was written as a tutorial and contains all proofs of the particle filter algorithm. Chapters 3 and 4 were written to be self-sufficient. In particular, a brief review (no proofs provided) of the particle filter was presented at the introduction of every chapter. This thesis is concluded in Chapter 5 with a summary of this work and its main contributions as well as a priming of future research perspectives.

Chapter 2

Literature Review

"You do not really understand something unless you can explain it to your grandmother."

– Albert Einstein

In this chapter, we provide a comprehensive review of optimal state estimation in non-linear, non-Gaussian state space models. Specifically, we present the fundamental concepts and algorithm of particle filtering methods to solve these estimation problems numerically in an online manner.

2.1 Problem Statement

We consider a discrete-time state-space model defined by the state and measurement equations:

$$\boldsymbol{x}_k = f_k(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_k, \tag{2.1}$$

$$\boldsymbol{y}_k = h_k(\boldsymbol{x}_k) + \boldsymbol{v}_k, \tag{2.2}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{n_x}$ and $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$ represent the system state and the system output, respectively; f_k and h_k are known and possibly non-linear function mappings; $\boldsymbol{w}_k$ is a realization from the zero-mean process noise with known probability density function (PDF), $\boldsymbol{w}_k \sim p(\boldsymbol{w})$; $\boldsymbol{v}_k$ is a realization of the the zero-mean measurement noise with known PDF, $\boldsymbol{v}_k \sim p(\boldsymbol{v})$; and $k = 0, 1, ...$ is the current time-step that is being evaluated. The process and measurement noise sequences are assumed to be uncorrelated and the initial PDF of the state $\boldsymbol{x}_0$ is $p(\boldsymbol{x}_0)$. A graphical representation of the system shown in Eqs (2.1) and (2.2)

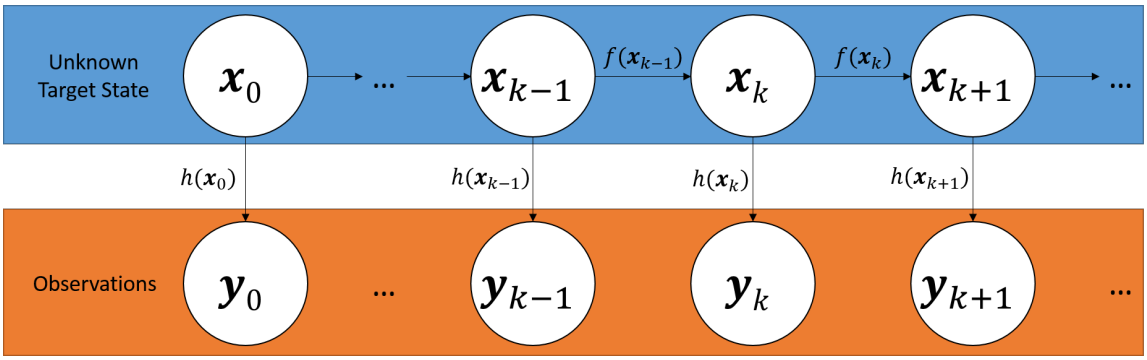


Figure 2. A graphical representation of the model in (2.1) and (2.2). Note the following two assumptions on the system: the current observation $\boldsymbol{y}_k$ depends only on the current state $\boldsymbol{x}_k$ and the state $\boldsymbol{x}_k$ is a Markov process, i.e., it depends only on the previous state $\boldsymbol{x}_{k-1}$.

is shown in Figure 2.

The dynamical system described in Eqs. (2.1) and (2.2) is also known as a Hidden Markov Model (HMM). More specifically, an HMM is a first-order Markov process with "hidden" or unobserved states. This class includes many models of interest. Thus, we have the following identities:

$$p(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}), \tag{2.3}$$

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{Y}_{k-1}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k), \tag{2.4}$$

where $\boldsymbol{Y}_k = \{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_k\}$ and $\boldsymbol{X}_k = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k\}$. In particular, the observations are conditionally independent given the states.

We are interested in estimating $\{\boldsymbol{x}_k\}_{k\geq 1}$ but only have access to the history of observations $\boldsymbol{Y}_k$. Hence, all information about the current state $\boldsymbol{x}_k$ can be found in the posterior distribution $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$. The optimal state estimate is then given by a point estimate of this

posterior density. In particular, the minimum mean square estimate is given by the conditional mean $E_{p(\boldsymbol{x}_k|\boldsymbol{Y}_k)}[\boldsymbol{x}_k]$ [16]. Using Bayes Theorem, the posterior distribution of the state can be written as:

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \frac{p(\boldsymbol{x}_k, \boldsymbol{Y}_k)}{p(\boldsymbol{Y}_k)} = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{Y}_{k-1})p(\boldsymbol{x}_k, \boldsymbol{Y}_{k-1})}{p(\boldsymbol{Y}_k)} = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{Y}_{k-1})p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1})}. \tag{2.5}$$

Using identity (2.4), we can simplify the expression of the posterior density shown in (2.5) as follows:

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1})}. \tag{2.6}$$

By applying the Total Probability Theorem, we can compute recursively the density $p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})$ using the previous posterior:

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) &= \frac{p(\boldsymbol{x}_k, \boldsymbol{Y}_{k-1})}{p(\boldsymbol{Y}_{k-1})} = \frac{\int p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1})d\boldsymbol{x}_{k-1}}{p(\boldsymbol{Y}_{k-1})} \\ &= \frac{\int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1})p(\boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1})d\boldsymbol{x}_{k-1}}{p(\boldsymbol{Y}_{k-1})} \\ &= \frac{p(\boldsymbol{Y}_{k-1})\int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})d\boldsymbol{x}_{k-1}}{p(\boldsymbol{Y}_{k-1})} \\ &= \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})d\boldsymbol{x}_{k-1}. \end{aligned}$$

Using identity (2.3), we can further simplify this expression.

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})d\boldsymbol{x}_{k-1} \tag{2.7}$$

Put simply, there are two equations: Eq. (2.7) constitutes the *prediction step*, which predicts

the current state from past observations; and Eq. (2.6) is the *correction step*, which updates the prediction by taking into account the latest observation. These two step equations are also known as the *Chapman-Kolmogorov equations*:

$$\begin{cases} p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) & = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1}) d\boldsymbol{x}_{k-1} \\ p(\boldsymbol{x}_k|\boldsymbol{Y}_k) & = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1})} \end{cases} \tag{2.8}$$

Unfortunately, the recursive relations in (2.8) are only a conceptual solution because the integrals involved are generally intractable and can only be solved in few special cases [3].

2.2 Optimal State Estimation in Linear Models

When the state-space model is linear, a closed form solution can be obtained analytically using the Kalman filter [17, 18]. This solution is optimal, in the mean square sense, if the noise is Gaussian. In this case, the posterior density $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ is Gaussian at every time step, therefore being fully parameterized by its mean and covariance matrix. In the linear Gaussian model, we have

- $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are both drawn from Gaussian distributions of known parameters.
- $f_k(\boldsymbol{x}_{k-1})$ is a known linear function of $\boldsymbol{x}_{k-1}$.
- $h_k(\boldsymbol{x}_k)$ is a known linear function of $\boldsymbol{x}_k$.

The linear system and measurement equations in (2.1) and (2.2) can be written as

$$\boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k, \tag{2.9}$$

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k., \tag{2.10}$$

Where $\boldsymbol{F}_k$ and $\boldsymbol{H}_k$ are both known matrices defining the linear functions. The noise realizations are $\boldsymbol{w}_k \sim \mathcal{N}(0, \boldsymbol{Q}_k)$ and $\boldsymbol{v}_k \sim \mathcal{N}(0, \boldsymbol{R}_k)$. In general, the system and measurement matrices, $\boldsymbol{F}_k$ and $\boldsymbol{H}_k$, as well as the covariance matrices, $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$, are time-varying and can change with k.

From the linearity and Gaussianity assumptions, we have

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1}) = \mathcal{N}(\boldsymbol{x}_{k-1}; \boldsymbol{m}_{k-1|k-1}, \boldsymbol{P}_{k-1|k-1}), \tag{2.11}$$

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) = \mathcal{N}(\boldsymbol{x}_k; \boldsymbol{m}_{k|k-1}, \boldsymbol{P}_{k|k-1}), \tag{2.12}$$

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \mathcal{N}(\boldsymbol{x}_k; \boldsymbol{m}_{k|k}, \boldsymbol{P}_{k|k}) \tag{2.13}$$

where

$$\boldsymbol{m}_{k|k-1} = \boldsymbol{F}_k \, \boldsymbol{m}_{k-1|k-1},$$

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{Q}_{k-1} + \boldsymbol{F}_k \boldsymbol{P}_{k-1|k-1} \boldsymbol{F}_k^T,$$

$$\boldsymbol{m}_{k|k} = \boldsymbol{m}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k \boldsymbol{m}_{k|k-1}),$$

$$\boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{P}_{k|k-1},$$

$\mathcal{N}(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{P})$ denotes a Gaussian distribution of the multidimensional variable $\boldsymbol{x}$ with mean

$\boldsymbol{m}$ and covariance matrix $\boldsymbol{P}$, and

$$\boldsymbol{S}_k = \boldsymbol{H}_k \boldsymbol{P}_{k|k-1} \boldsymbol{H}_k^T + \boldsymbol{R}_k,$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1} \boldsymbol{H}_k^T \boldsymbol{S}_k^{-1},$$

Are, respectively, the covariance matrix of the innovation term $\boldsymbol{y}_k - \boldsymbol{H}_k \boldsymbol{m}_{k|k-1}$ and the Kalman Gain [17].

The above equations define the familiar form of the Kalman Filter. Because of these highly restrictive assumptions on the state-space model (linearity and Gaussianity), it is sufficient to propagate only the mean and covariance of the posterior distribution to fully characterize the state. The estimate for the state, given by $\boldsymbol{m}_{k|k}$, is the optimal estimate of the hidden state.

2.3 Approximate Solutions in Nonlinear Models

2.3.1 The Extended Kalman Filter. If the system is non-linear, then we can use a local linearization to approximate the functions f_k and h_k in (2.1) and (2.2). This is the essence of the Extended Kalman Filter (EKF). If the non-linear functions are "nearly" linear, then the EKF would be appropriate as the approximation error would be small.

The EKF approximations are given by

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1}) \approx \mathcal{N}(\boldsymbol{x}_{k-1}; \boldsymbol{m}_{k-1|k-1}, \boldsymbol{P}_{k-1|k-1}), \tag{2.14}$$

$$p(\boldsymbol{x}_{k}|\boldsymbol{Y}_{k-1}) \approx \mathcal{N}(\boldsymbol{x}_{k}; \boldsymbol{m}_{k|k-1}, \boldsymbol{P}_{k|k-1}), \tag{2.15}$$

$$p(\boldsymbol{x}_{k}|\boldsymbol{Y}_{k}) \approx \mathcal{N}(\boldsymbol{x}_{k}; \boldsymbol{m}_{k|k}, \boldsymbol{P}_{k|k}), \tag{2.16}$$

where

$$\begin{aligned}
\boldsymbol{m}_{k|k-1} &= f_k(\boldsymbol{m}_{k-1|k-1}), \\
\boldsymbol{P}_{k|k-1} &= \boldsymbol{Q}_{k-1} + \hat{\boldsymbol{F}}_k \boldsymbol{P}_{k-1|k-1} \hat{\boldsymbol{F}}_k^T, \\
\boldsymbol{m}_{k|k} &= \boldsymbol{m}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - h_k(\boldsymbol{m}_{k|k-1})), \\
\boldsymbol{P}_{k|k} &= \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k \hat{\boldsymbol{H}}_k \boldsymbol{P}_{k|k-1},
\end{aligned}$$

where $\hat{\boldsymbol{F}}_k$ and $\hat{\boldsymbol{H}}_k$ are matrices that are local linearizations of the non-linear functions f_k and h_k, respectively,

$$\begin{aligned}
\hat{\boldsymbol{F}}_k &= \frac{df_k(x)}{dx}\Big|_{x=\boldsymbol{m}_{k-1|k-1}}, \\
\hat{\boldsymbol{H}}_k &= \frac{dh_k(x)}{dx}\Big|_{x=\boldsymbol{m}_{k|k-1}}, \\
\boldsymbol{S}_k &= \hat{\boldsymbol{H}}_k \boldsymbol{P}_{k|k-1} \hat{\boldsymbol{H}}_k^T + \boldsymbol{R}_k, \\
\boldsymbol{K}_k &= \boldsymbol{P}_{k|k-1} \hat{\boldsymbol{H}}_k^T \boldsymbol{S}_k^{-1}.
\end{aligned}$$

The Extended Kalman Filter uses the first-order Taylor approximation of the non-linear

system and observation functions. It is possible to expand this formulation to a higher order Taylor approximation for severely non-linear functions, however this can lead to prohibitively complex calculations [2, 19].

2.3.2 The Unscented Kalman Filter. In the realm of Kalman filtering, there is one more technique that is widely used in estimation problems. This approach, called the *Unscented Kalman Filter* (UKF), is yet another approximation for non-linear and potentially non-Gaussian state-space models. The UKF is based on a technique called the *Unscented Transform* (UT) which calculates the statistics of a random variable that undergoes a non-linear transformation [2]. To explain the UT, let us look at an example where we are propagating a random variable $\boldsymbol{x}$ of dimension L through a non-linear function $\boldsymbol{y} = g(\boldsymbol{x})$ and $\boldsymbol{x}$ has a mean $\bar{\boldsymbol{x}}$ and covariance $\boldsymbol{P}_{\boldsymbol{x}}$.

In order to calculate the statistics of $\boldsymbol{y}$, we generate a series $i = \{1, ..., 2L + 1\}$ of deterministic *sigma points* $\S_i$ according to the statistics of $\boldsymbol{x}$ and other free parameters. These free parameters allow us to control the spread of the points around $\bar{\boldsymbol{x}}$ as well as incorporate prior knowledge about the distribution of $\boldsymbol{x}$. These sigma points are then transformed using the function $\dagger_i = g(\S_i)$. The transformed points $\dagger_i$ are then weighted and summed to form a weighted sample mean $\bar{\boldsymbol{y}}$ and covariance $\boldsymbol{P}_y$ of the sigma points, which constitute estimates of the statistics of $\boldsymbol{y}$.

It is relatively simple to incorporate the UT into the recursive framework known as the UKF [2]. It is important to note that the UT technique is different from the Monte Carlo approaches, which will be explored later, because the small number of sigma points that are used in the UT are deterministically chosen, whereas Monte Carlo approaches rely

on a large number of random "points", called "particles". Even though the estimates are found using only $2L + 1$ samples (L being the dimension of the state), the UT can result in approximations that are accurate to the third order for Gaussian inputs and accurate to at least the second order for non-Gaussian inputs [2]. Thus, the UKF outperforms the EKF, which relies on a first-order approximation. The UKF also allows us to transform the sigma points directly without computing the Jacobian or the Hessian of the state. Overall, the UKF has been proven more accurate with the same computational complexity as the EKF [2].

2.4 The Particle Filter Framework

In comparison with the EKF and the UKF, the principal advantage of particle filters is that they do noy rely on any local linearization technique or any crude functional approximation. The Particle Filter is a numerical algorithm that is based on Monte Carlo sampling. We present below the underlying theory and computational framework of particle filters.

2.4.1 Monte Carlo Sampling. The need for Monte Carlo Sampling is not only within the span of particle filtering. Its applications actually span a broader class of problems. One of the most useful applications of Monte Carlo sampling is estimating an integral numerically. For instance, consider an integral of the form

$$\rho_x = \int_{x \in \mathcal{X}} f(x)p(x)dx, \tag{2.17}$$

where $\mathcal{X}$ is the sample space and p is a probability distribution function. We can approximate the result, $\hat{\rho}_x$, using Monte Carlo sampling. It is interesting to note that Eq. (2.17) represents the expected value $E_{p(x)}[f(x)]$. Given N random variables $x^{(1)}, x^{(2)}, ..., x^{(N)}$, all drawn from $p(x)$, $x^{(i)} \sim p(x)$, we can estimate $\hat{\rho}_x$ as

$$\hat{\rho}_x = \frac{1}{N} \sum_{i=1}^{N} f(x^{(i)}). \tag{2.18}$$

Equation (2.18) can be easily shown to be an unbiased estimator of ρ_x, i.e., $E[\hat{\rho}_x] = \rho_x$. We can also show that the variance of the estimate decreases as the number of samples N increases, i.e., as $N \to \infty$, the variance $Var[\hat{\rho}_x] \to 0$. By the law of large numbers, as $N \to \infty$, $\hat{\rho}_x \to \rho_x$, almost surely [20]. This is the well known result for the empirical calculation of the expected mean.

2.4.2 Importance Sampling. Using Monte Carlo sampling to approximate an integral is a very powerful technique, however, it does not work for distributions that are "hard" to sample from. In this case, we cannot form an empirical approximation of the mean $\hat{\rho}_x$ using Eq. (2.18) because we do not know how to sample from the distribution $p(x)$.

What if there was another distribution, $q(x)$, that we could sample from directly? By finding a procedure that uses a sampling distribution q to approximate samples of p, we could estimate the expected value of any arbitrary pdf without having to sample from it directly. This technique is called *Importance Sampling* and is commonly used for the estimation of non-standard pdfs. The density $q(x)$ is commonly referred to as the *importance distribution* or *biasing distribution*. By weighting the samples generated from $q(x)$ so that

they "appear" to have been taken from $p(x)$, we can perform a modified empirical mean calculation. In order to use this technique, we must first rewrite Eq. (2.17) to incorperate $q(x)$ as follows:

$$\rho_x = \int_{x \in \mathcal{X}} f(x) \frac{p(x)}{q(x)} q(x) dx = \int_{x \in \mathcal{X}} f(x) w(x) q(x) dx, \tag{2.19}$$

where $w(x) = \frac{p(x)}{q(x)}$. By dividing and multiplying by $q(x)$, we have modified the emprical approximation as well by computing the expected value with respect to $q(x)$ instead of $p(x)$. Given N random variables $x^{(1)}, x^{(2)}, ..., x^{(N)}$, all drawn from $q(x)$, i.e., $x^{(i)} \sim q(x)$, we can estimate the result $\hat{\rho}_x$ as

$$\hat{\rho}_x = \frac{1}{N} \sum_{i=1}^{N} f(x^{(i)}) w(x^{(i)}). \tag{2.20}$$

As before, the law of large numbers states that as $N \to \infty$, $\hat{\rho}_x \to \rho_x$, almost surely [20].

2.4.3 Sequential Importance Sampling. Importance sampling has been proven to be a valuable tool for estimating the expected value of many types of arbitrary distributions. It is also very important in approximating not only expected values but also pdfs. For instance, consider a pdf, which is "hard" to compute

$$p(x) = \frac{\pi(x)}{\int_{x \in \mathcal{X}} \pi(x) dx},$$

Where $\pi(x)$ is the un-normalized pdf. We may know how to compute $\pi(x)$, but the integrals in the denominator may be impossible to compute. By using importance sampling, we can

sample from a known distribution $q(x)$, such as a Gaussian distribution, and weight the particles to represent the distribution $p(x)$. Given N random variables $x^{(1)}, x^{(2)}, ..., x^{(N)}$ such that $x^{(i)} \sim q(x)$, we can formulate $p(x)$ as a pointwise approximation [1] defined by

$$p(x) \approx \sum_{i=1}^{N} w^{(i)} \delta(x - x^{(i)}), \tag{2.21}$$

where the unnormalized weights are computed as

$$\bar{w}^{(i)} = \frac{\pi(x^{(i)})}{q(x^{(i)})} \tag{2.22}$$

and the weights are normalized using the following

$$w^{(i)} = \frac{\bar{w}^{(i)}}{\sum_{j=1}^{N} w^{(j)}} \tag{2.23}$$

This has a direct application to the estimation problem that we are considering! Recall that we are trying to compute the posterior pdf of the state given past observations $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ in the non-linear and non-Gaussian problem described in (2.6). This pdf can be of any arbitrary shape and is not bound by any assumptions, such as Gaussianity. We would also like to be able to estimate $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ sequentially as our system evolves over k. *Sequential Importance Sampling* solves this problem by taking into account the evolution of a system over time. By estimating $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ as a weighted train of impulses, as described in Eq. (2.21), using N samples, or particles, drawn from a known importance distribution $q(x)$, the samples can be weighted using $w^{(i)}$ to form an estimate of $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$. In order to use Sequential Importance Sampling, we must first derive the expression of $w^{(i)}$. We will

then be able to find our mean estimate with respect to the posterior pdf, which is known to be the optimal estimate of the state $\boldsymbol{x}_k$ at time k.

In order to derive the recursive solution to the posterior density, we will first look at the joint posterior distribution $p(\boldsymbol{X}_k|\boldsymbol{Y}_k)$. Using Bayes theorem, we can write the recursion

$$p(\boldsymbol{X}_k|\boldsymbol{Y}_k) = \frac{p(\boldsymbol{X}_k, \boldsymbol{Y}_k)}{p(\boldsymbol{Y}_k)} = \frac{p(\boldsymbol{y}_k|\boldsymbol{X}_k, \boldsymbol{Y}_{k-1})p(\boldsymbol{X}_k|\boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1})}.$$

Using the Markov assumption in Eq. (2.4) and the fact that the observations are conditionally independent given the state, the above formulation becomes

$$\begin{aligned} p(\boldsymbol{X}_k|\boldsymbol{Y}_k) &= \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}, \boldsymbol{Y}_{k-1})p(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1})} && (2.24) \\ &= \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1})}. && (2.25) \end{aligned}$$

We have $p(\boldsymbol{X}_k|\boldsymbol{Y}_k) \propto p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_{k-1})$ and this quantity has a recursive component $p(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_{k-1})$, which allows us to sequentially update the posterior distribution.

The denominator in (2.25), however, cannot be computed analytically as it involves a complex multi-dimensional integral of arbitrary densities,

$$p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}) = \int_{\boldsymbol{X}_{k-1} \in \mathcal{X}} p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_{k-1})dx.$$

This normalization term is very hard to compute because it involves terms that are analytically intractable! The dimension of the integral also increases with the time k. Luckily,

we have already described a procedure that can approximate pdfs with an analytically intractable normalization term, such as $p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1})$.

Using Sequential Importance Sampling, Eq. (2.21) can be used to estimate $p(\boldsymbol{X}_k|\boldsymbol{Y}_k)$. We first sample N random variables $x^{(1)}, x^{(2)}, ..., x^{(N)}$ from the importance distribution $q(x)$, $x^{(i)} \sim q(\boldsymbol{X}_k|\boldsymbol{Y}_k)$. Reformulating the importance distribution as

$$q(\boldsymbol{X}_k|\boldsymbol{Y}_k) = q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}, \boldsymbol{Y}_k)q(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_k)$$

and utilizing the assumption that $q(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_{k-1}) = q(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_k)$ because $\boldsymbol{X}_{k-1}$ is independent of $\boldsymbol{y}_k$,

$$q(\boldsymbol{X}_k|\boldsymbol{Y}_k) = q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}, \boldsymbol{Y}_k)q(\boldsymbol{X}_{k-1}|\boldsymbol{Y}_{k-1}). \tag{2.26}$$

We can now have a formula for the weights $w_k^{(i)}$ as

$$w_k^{(i)} \propto \frac{p(\boldsymbol{X}_k^{(i)}|\boldsymbol{Y}_k)}{q(\boldsymbol{X}_k^{(i)}|\boldsymbol{Y}_k)} = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)})p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{q(\boldsymbol{x}_k^{(i)}|\boldsymbol{X}_{k-1}^{(i)}, \boldsymbol{Y}_k)} \frac{p(\boldsymbol{X}_{k-1}^{(i)}|\boldsymbol{Y}_{k-1})}{q(\boldsymbol{X}_{k-1}^{(i)}|\boldsymbol{Y}_{k-1})}.$$

Since $\frac{p(\boldsymbol{X}_{k-1}^{(i)}|\boldsymbol{Y}_{k-1})}{q(\boldsymbol{X}_{k-1}^{(i)}|\boldsymbol{Y}_{k-1})}$ is the weight estimate at time $k-1$, the un-normalized weight estimate is given by

$$\bar{w}_k^{(i)} = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)})p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{q(\boldsymbol{x}_k^{(i)}|\boldsymbol{X}_{k-1}^{(i)}, \boldsymbol{Y}_k)} w_{k-1}^{(i)}. \tag{2.27}$$

The weights are then normalized using Eq. (2.23). This process is illustrated in Figure 3.

In summary, at every time step k, N samples are drawn from the importance distribution $q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}^{(i)}, \boldsymbol{Y}_k)$. These samples are then weighted using Eq. (2.27) to form the estimate of $p(\boldsymbol{X}_k|\boldsymbol{Y}_k)$ as described in Eq. (2.21). As $N \to \infty$, the approximation in (2.21) approaches

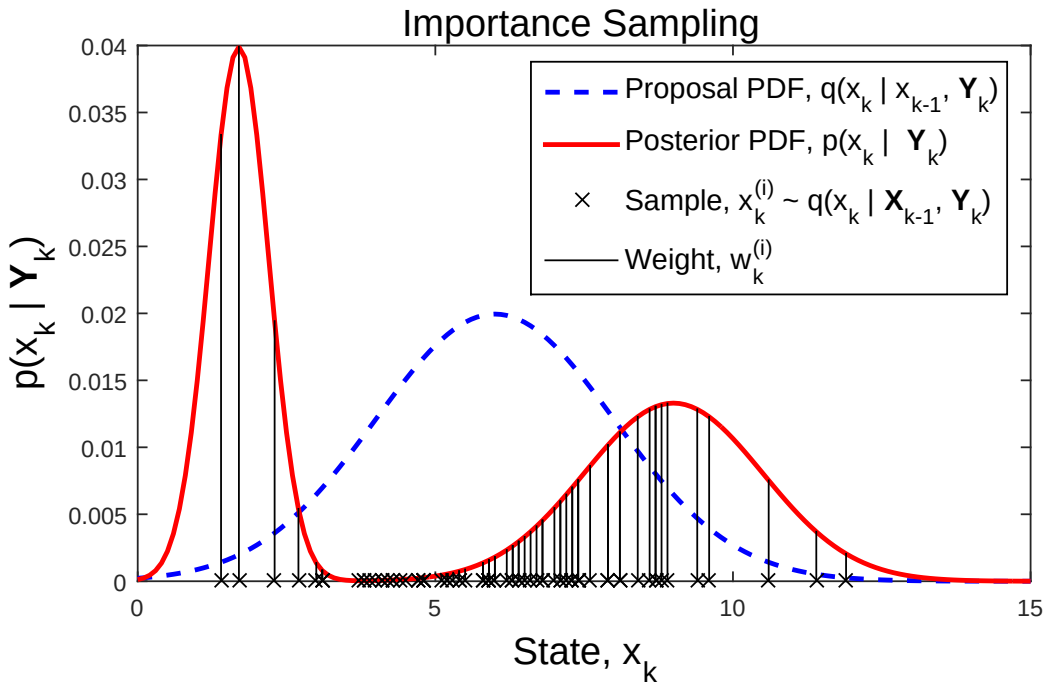


Figure 3. An approximation of the PDF $p(x_k|Y_k)$ using samples $x^{(i)}$ generated from $x^{(i)} \sim q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}, \boldsymbol{Y}_k)$. These samples are weighted and the result is an estimate of the posterior PDF $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$.

the true density $p(\boldsymbol{X}_k|\boldsymbol{Y}_k)$ [20]. The last step is formulating importance sampling into a sequential algorithm that can approximate the posterior density $p(\boldsymbol{X}_k|\boldsymbol{Y}_k)$ at every time step using the previous estimate. We will focus on the marginal distribution, $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$, because its mean estimate is the optimal estimate of the state $\boldsymbol{x}_k$.

2.4.4 The Particle Filter. In Sequential Importance Sampling, we developed a method that allows us to estimate the pdf $p(\boldsymbol{X}_k|\boldsymbol{Y}_k)$ indirectly using another "easier" to sample importance density $q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}, \boldsymbol{Y}_k)$. First, N particles are sampled from $q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}^{(i)}, \boldsymbol{Y}_k)$. Then, we weight these samples using Eq. (2.27) and normalize the weights using Eq. (2.23). This leads to an estimate of the desired pdf $p(\boldsymbol{X}_k|\boldsymbol{Y}_k)$ as described in Eq. (2.21).

In most cases, only the marginal posterior distribution, $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$, is needed at time step k. In fact, the optimal state estimate is given by the conditional mean estimate of the marginal posterior:

$$E[\boldsymbol{x}_n|\boldsymbol{Y}^n] \approx \hat{\boldsymbol{x}}_k = \int \boldsymbol{x}_k \, \hat{p}(\boldsymbol{x}_k|\boldsymbol{Y}^k) d\boldsymbol{x}_x = \sum_{i=1}^{N} w_n^{(i)} \boldsymbol{x}_k^{(i)}. \tag{2.28}$$

In these cases, the importance density can be written as $q(\boldsymbol{x}_k|\boldsymbol{X}_{k-1}^{(i)}, \boldsymbol{Y}_k) = q(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{Y}_k)$ [3]. The un-normalized weights are now given by

$$\bar{w}_k^{(i)} = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)}) p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{q(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{Y}_k)} w_{k-1}^{(i)} \tag{2.29}$$

and the marginal density is calculated as

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) \approx \sum_{i=1}^{N} w^{(i)} \delta(\boldsymbol{x} - \boldsymbol{x}^{(i)}), \tag{2.30}$$

Importance sampling, however, suffers from a *degeneracy* problem, where after few iterations, only one particle will have a significant weight and the others will have almost zero weight. When degeneracy happens, the filter "degenerates" to using a few of the N particles to represent the pdf. This problem occurs because the variance of the weights can only increase over time [21], which makes it impossible to avoid this degeneracy phenomenon. Not only does this phenomenon inherent to Sequential Importance Sampling reduce the estimation accuracy because it suggests that over time only a few samples will be representing an entire pdf, but it also implies that we will be performing an expensive operation in weighting N particles and only a handful of these particles will actually be

used!

At this point, the Particle Filter (PF) has still not been formally introduced. Now that the degeneracy phenomenon has been discussed, however, the familiar form of the Particle Filter can be discussed, which is based on Sequential Importance *Resampling*. There are two ways to reduce the accumulation of the variance in the weights. The first is to pick a proposal distribution that is close to the actual joint posterior distribution [21]. There have been many different types of proposals discussed in literature, including a theoretically optimal proposal that minimizes the variance of the weights. However, this optimal distribution is not easy to sample from. The second way to reduce the variance of the weights is through a method called *resampling*. This method is the main contribution to the Sequential Importance Resampling framework.

There are many variants of the Particle filter in use today, whose performance depends on the types of state-space models and assumptions. The generic Particle filter follows the following structure. The PF follows the same procedure of sampling N particles from the proposal distribution $q(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{Y}_k)$ and then weighting those samples using Eqs. (2.29) and (2.23). The change that Sequential Importance Resampling makes is that it adds one additional step to the procedure: the resampling step. There are many different methods for resampling that try to create a trade-off between sampling variance reduction and computational complexity. However, the basic idea of all resampling techniques is to eliminate the particles that have small weights (low probability areas) and keep the particles that have large weights (high probability areas). This procedure results in an entirely new set of particles ${x_k^{(i)*}}_{i=1}^{N}$, which have been sampled (with replacement) N times from an the approximate discrete representation of the posterior distribution. The result is a new set of

for $k = 1, 2, ...$ **do**
  for $i = 1, 2, ..., N$ **do**
    Draw $\boldsymbol{x}_k^{(i)} \sim q(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{Y}_k)$.
    Evaluate the un-normalized weight $\hat{w}_k^{(i)}$ using Eq. (2.29).
  end
  Normalize the weights using Eq. (2.23).
  Resample the particles if appropriate.
  Estimate the state using the mean estimate: $\hat{\boldsymbol{x}}_k^{(i)} = \sum_{i=1}^{N} w_k^{(i)} \boldsymbol{x}_k^{(i)}$.
end

Figure 4. The generic particle filter algorithm.

particles with each particle having the same weight $w_k^{(i)} = \frac{1}{N}$ [22]. Figure 4 outlines the procedure of the generic Particle Filter.

Theoretically, the proposal distribution q can be any distribution which support includes the support of the posterior density p. However, in practice, the number of particles is finite, and thus, the choice of the proposal distribution is very important in the performance of the particle filter. A popular simplification of the algorithm involves setting the proposal distribution to be the prior distribution, i.e., $q(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{Y}_k) = p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1})$. This simplification is mainly used because of the availability of the prior distribution (usually a Gaussian when the system noise is Gaussian), and it also leads to a simplification of the weight calculation as $\bar{w}_k^{(i)} = p(\boldsymbol{y}_k | \boldsymbol{x}_k^{(i)}) w_{k-1}^{(i)}$. That is, the weights of the particles are given by their likelihood functions. This choice of the proposal distribution, however, is usually a poor one because the prior density does not take into account the most recent observation, and hence introduces a large amount of variance into the weights [3].

Finally, we would like to mention an important problem with Particle Filtering, known

as the *curse of dimensionality*, which is still an active area of research. As with most numerical algorithms, it has been found that the particle filter's performance starts to decay in higher dimensional state-spaces. Specifically, as the dimensionality of the system continues to increase, the particles needed to accurately sample the space increases super-exponentially [11]. This limitation can restrict the particle filter to not being practical in high-dimensional applications. In this thesis, we will tackle, among other problems, the dimensionally issue of the PF and propose a solution for mixed state-spaces that incorporate both linear and non-linear structures.

Chapter 3

The Constrained Particle Filter

3.1 Introduction

The state of many dynamical systems is often required to satisfy certain constraints arising from basic physical laws, mathematical properties or geometric considerations, e.g., maximum power or transmission capacity, energy conservation laws and bounded parameters. In fact, constrained systems are already omnipresent in many real-world applications including camera tracking [23], fault diagnosis [24], chemical processes [10], vision-based systems [25], target tracking [5, 6], biomedical systems [26], robotics [7] and navigation [27].

Particle Filters (PF) are a broad class of Monte Carlo algorithms, which provide approximate solutions to analytically intractable inference problems, which can include nonlinear and non-Gaussian modeling scenarios. PFs can solve these problems by using *particles*, which sample the state space of the system. These particles are then weighted to estimate the state posterior density. The estimation converges, in the mean-square error, to the true posterior density of the state. PFs have become a viable alternative to more traditional techniques, such as the Extended Kalman Filter (EKF) due to the PF's ability to calculate posterior densities without using functional approximation such as local linearization techniques or assume Gaussian noise.

However, the very numerical nature of the particle filters, which constitutes their strength for multidimensional numerical integration, becomes their major weakness in handling constraints on the state. The main difficulty of the constrained PF problem stems from the fact that every particle in the particle approximation of the state posterior density is a local representation of the density, and thus cannot characterize global properties of the density,

such as constraints on the conditional mean or any other functional expectation. The current trend in constrained particle filtering simply enforces the constraints on all particles of the PF. This approach, however, constrains the posterior density of the state rather than its mean, which leads to more stringent conditions and possibly a completely different condition than the original constraints (see Figure 5). We refer to the approach of constraining all particles as the *Pointwise Density Truncation* (PDT) method.

In this thesis, we introduce a new approach called *Mean Density Truncation* (MDT) that imposes the state constraints on the conditional mean estimate without further restraining the posterior distribution of the state. This chapter is organized in the following way: the unconstrained PF framework is reviewed in Section 3.2, the PDT and MDT approaches are advanced in Section 3.3, and simulation results that compare the PDT and MDT approaches are presented in Section 3.4.

3.2 The Unconstrained Particle Filter

We consider a discrete-time state-space model defined by the following state and measurement equations:

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_k, \\ \boldsymbol{y}_k &= \boldsymbol{h}_k(\boldsymbol{x}_k) + \boldsymbol{v}_k, \end{aligned} \tag{3.1}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{n_x}$ and $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$ represent the system state and the system output, respectively; $\boldsymbol{f}_k$ and $\boldsymbol{h}_k$ are known and possibly non-linear function mappings; $\boldsymbol{w}_k$ is a realization from the zero-mean process noise with known pdf, $\boldsymbol{w}_k \sim p(\boldsymbol{w})$; $\boldsymbol{v}_k$ is a realization of the

the zero-mean measurement noise with known pdf, $\boldsymbol{v}_k \sim p(\boldsymbol{v})$; and $k = 0, 1, ...$ is the current time-step that is being evaluated.

Let $\boldsymbol{Y}_k = [\boldsymbol{y}_1, ..., \boldsymbol{y}_k]$ denote the history of observations up to time k. In the Bayesian context, inference of $\boldsymbol{x}_k$ given a realization of the observations $\boldsymbol{Y}_k$ relies upon the posterior density $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$. Using the Bayesian rule, we can obtain the following two-step Bayesian recursion formula:

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})d\boldsymbol{x}_{k-1} \tag{3.2}$$

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})}{\int p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})d\boldsymbol{x}_k} \tag{3.3}$$

Equations (3.2)-(3.3) are a conceptual solution because the integrals defined are, in general, intractable. For the linear Gaussian model, it is easy to check that $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ is a Gaussian distribution whose mean and covariance can be computed using the Kalman filter. However, for most nonlinear non-Gaussian models, it is not possible to compute these distributions in closed-form.

The PF approximates the posterior pdf using an ensemble of *particles* $\{\boldsymbol{x}_k^{(i)}\}_{i=1}^N$ and their associated weights $\{w_k^{(i)}\}$:

$$\hat{p}(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \sum_{i=1}^{N} w_k^{(i)} \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}) \tag{3.4}$$

where $\delta(.)$ is the Dirac delta function and N is the number of particles. Ideally, the particles are required to be sampled from the true posterior, $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$, which is not available. Therefore, another distribution, referred to as the *importance distribution* or the

proposal distribution, $q(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_k)$, is used. The particles at time k are sampled from $\boldsymbol{x}_k^{(i)} \sim q(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_k)$. The importance weight of each particle $\boldsymbol{x}_k^{(i)}$ is computed as

$$\bar{w}_k^{(i)} = w_{k-1}^{(i)} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)}) p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{q(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_k)}, \tag{3.5}$$

where $\bar{\boldsymbol{w}}_k$ are the un-normalized weights [28]. The normalized weights in (3.4) are given by $w_k^{(i)} = \bar{w}_k^{(i)} / \sum_{j=1}^{N} w_k^{(j)}$.

3.3 The Constrained Particle Filter

We focus on the discrete state-space model in (3.1) augmented with the following general constraint

$$\boldsymbol{a}_k \leq \boldsymbol{\phi}_k(\boldsymbol{x}_k) \leq \boldsymbol{b}_k, \tag{3.6}$$

where ϕ_n is the constraint function at time n and the inequality holds for all elements. It is important to emphasize the fact that the constraint needs to only be satisfied by the state estimate given by the conditional mean, i.e., we must have

$$\boldsymbol{a}_k \leq \phi_k(\hat{\boldsymbol{x}}_k) = \phi_k(E[\boldsymbol{x}_k|\boldsymbol{Y}_k]) \approx \phi_k(\sum_{j=1}^{N} w_k^{(i)} \boldsymbol{x}_k^{(i)}) \leq \boldsymbol{b}_k$$

This mean constraint is not a local condition, meaning there are many ways to globally constrain the mean. Projection of the unconstrained density onto the constraint set is only one possible option. The widely used approach in constrained sequential Monte Carlo is the *acceptance/rejection* approach, which enforces the constraints by simply rejecting the particles violating them [29]. The acceptance/rejection procedure does not make any

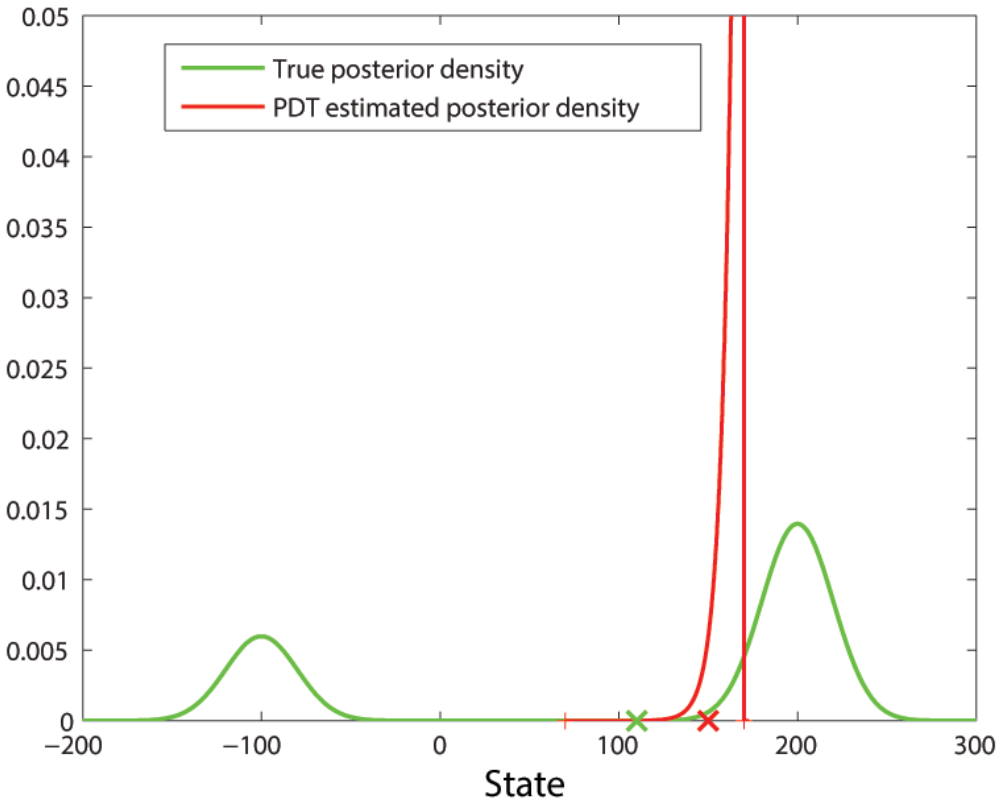


Figure 5. Illustration of the PDT approach for an interval-constrained system, $x_k \in [70, 170]$ for all k. The true posterior density (green curve) is multimodel with mean 110 (green x-mark). If all particles are constrained to be within the interval $[70, 170]$, then the estimated posterior density (red curve) will be a truncated exponential density that is dramatically different from the true posterior distribution.

assumption on the distributions and therefore maintains the generic property of the particle filter. However, the number of samples will be reduced and hence the estimation accuracy may decrease, especially with a poor choice of the proposal density. An extreme example is when most (or all) of the particle violate the constraint and the algorithm fails [30].

3.3.1 Pointwise Density Truncation (PDT). The current practice in the literature constrains the mean of the posterior distribution by imposing the constraints on all particles

of the PF [30–37]. However, this is not true. Imposing the constraint on all particles results, in general, in a stronger constraint and possibly a completely different or even irrelevant condition. To see this, let us consider the scalar case with $\mathcal{C}_k = [a, b]$ for all k: the state estimate is constrained in the interval $[a, b]$ or $a \leq x_k \leq b$. Constraining every particle to be within the interval $[a, b]$ is equivalent to constraining the support of the posterior distribution to this interval, which is a much stronger condition than constraining the mean of the distribution, or any point estimate, to be inside the interval. We refer to this approach as *pointwise density truncation* or *particle density truncation* (PDT). Since the particle filter estimates the posterior density of the state, imposing stronger constraints may, and in general will, result in an erroneous estimation of the density, as illustrated in Figure 5.

3.3.2 Mean Density Truncation (MDT). In the constrained state-space model, the constraints must be satisfied by the estimate of the conditional mean. Unlike the pointwise density truncation approach, which enforces the constraints on all particles, we propose the mean density truncation (MDT) approach, which constrains only one particle in order to confine the estimated mean to the desired constraints. In the MDT approach, $(N-1)$ unconstrained particles are drawn from the proposal distribution. Then, the N^{th} particle is constrained in order to impose the conditions on the sample mean. A constraint of the form $a_k \leq \phi_k(\boldsymbol{x}_k) \leq b_k$ can be equivalently expressed as

$$a_k \leq \phi_k\left(\sum_{j=1}^{N} \omega_k^{(j)} \boldsymbol{x}_k^{(j)}\right) \leq b_k. \tag{3.7}$$

For simplicity, we will assume that the weights are given by the likelihood, i.e., the proposal density is the prior distribution function; the essence of the MDT method remains the same in the general case, where the proposal density is different from the prior distribution.

Separating the summation of the $(N-1)$ unconstrained particles from the N^{th} particle, and taking into account the normalization of the weights, the constraint becomes

$$a_k \leq \phi_k\left(\frac{\sum_{j=1}^{N-1} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(j)})\boldsymbol{x}_k^{(j)} + p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(N)})\boldsymbol{x}_k^{(N)}}{\sum_{j=1}^{N} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(j)})}\right) \leq b_k$$

Then, conditions on the N^{th} particle can be derived depending on the explicit expression of the constraint function ϕ_k. For instance, if we consider the interval-type constraint, i.e., ϕ_k is the identity function, then the above inequality becomes equivalent to the two inequalities,

$$\sum_{j=1}^{N-1} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(j)})(a_k - \boldsymbol{x}_k^{(j)}) \quad \leq p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(N)})(\boldsymbol{x}_k^{(N)} - a_k), \tag{3.8}$$

$$\sum_{j=1}^{N-1} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(j)})(b_k - \boldsymbol{x}_k^{(j)}) \quad \geq p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(N)})(\boldsymbol{x}_k^{(N)} - b_k). \tag{3.9}$$

Letting $q_1(\boldsymbol{x}_k^{(N)}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(N)})(\boldsymbol{x}_k^{(N)} - a_k)$ and $q_2(\boldsymbol{x}_k^{(N)}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(N)})(\boldsymbol{x}_k^{(N)} - b_k)$, we obtain the two inequalities

$$\begin{aligned} q_1(\boldsymbol{x}_k^{(N)}) &\geq C_1(\{\boldsymbol{x}_k^{(j)}\}_{j=1}^{N-1}), \\ q_2(\boldsymbol{x}_k^{(N)}) &\leq C_2(\{\boldsymbol{x}_k^{(j)}\}_{j=1}^{N-1}), \end{aligned} \tag{3.10}$$

which have to be satisfied for the N^{th} particle only. C_1 and C_2 are two constants, which depend only on the already sampled $(N-1)$ unconstrained particles and their weights. Depending on the likelihood function, Eq (3.10) can be solved analytically or numerically. The solution to (3.10) may not be unique. Many "N^{th} particles" can satisfy (3.10), all of them enforcing the original constraint on the sample mean estimate. These different solutions may lead to different constrained estimates. We found, in our preliminary results, that

the solution with the highest weight (here, likelihood) leads to the most accurate estimator among all other solutions having lower weights.

If the proposal distribution is chosen poorly, the $(N-1)$ unconstrained particles will lie in a low probability region of the posterior density of the state. In this case, it may not be possible to find an N^{th} particle that satisfies (3.10), thus imposing the constraint on the sample mean. Intuitively, if the initial particle sampling is poor, then one additional particle may not be able to force the mean to satisfy the desired constraints. We advance two solutions to ensure the existence of an N^{th} particle that will enforce the constraint on the sample mean: m^{th}-order MDT and inductive MDT (IMDT).

In the case where one particle may not be sufficient to constrain the mean, it seems reasonable to consider constraining more than one particle, e.g., two, three or up to $m \leq N$ particles. These m constrained particles will ensure that the sample mean satisfies the desired constraint. The MDT method is thus termed 1^{st}-order MDT, and its extension to m constrained particles is called m^{th}-order MDT. In the m^{th}-order MDT, $(N-m)$ unconstrained particles are sampled from the proposal distribution, and the remaining m particles are constrained in order to satisfy the condition on the sample mean. It is important to notice that when $m = N$, the N^{th}-order MDT is very different from the PDT method: In the PDT approach, the original constraint is imposed on all particles. On the other hand, the N^{th}-order MDT constrains the particles, as in Eq. (3.7), in order to impose the desired condition on the sample mean.

3.4 Simulation Results

We consider the following nonlinear dynamic system

$$x_{k+1} = \frac{x_k}{2} + 25\frac{x_k}{1+x_k^2} + 8\cos(1.2k) + w_k \tag{3.11}$$
$$y_k = \frac{x_k^3}{25} + x_k + v_k; \quad -5 \leq x_k \leq 5,$$

where w_k and v_k are zero-mean Gaussian white noise. This example is severely nonlinear, both in the system and the measurement equations. It was shown in [38] that the EKF fails in estimating the true state value of this unconstrained system.

Figure 7(a) shows the true and estimated trajectories using 1^{st}-order MDT and PDT. The results are shown for 1000 Monte Carlo simulations. It is seen that, on average, the 1^{st}-order MDT leads to more accurate estimation of the dynamic state, where both the mean-square error and the variance are smaller. Figures 7 (b),(c) show the posterior density of the constrained state as it evolves over time, for 1^{st}-order MDT and PDT, respectively. First, observe that the PDT approach (Figure 7(c)) produces posterior distributions with a bounded support at all time points, whereas the MDT approach results in proper unbounded support densities. Moreover, the PDT approach results in multiple spurious peaks within the densities. These large peaks are located mainly at the boundary of the constraining interval. These spurious peaks are due to the fact that sampled particles that do not satisfy the constraint are projected onto the boundary, thus creating a significant positive mass at the boundary of the constraint set and a small density mass elsewhere. In other words, in PDT, the density outside of the interval $[-5, 5]$ is projected onto the boundary points. On the other hand, the 1^{st}-order MDT method does not suffer from the 'boundary spurious peaks' problem and estimates smooth (multinomial) densities over time, which results in more accurate estimation of the conditional mean.

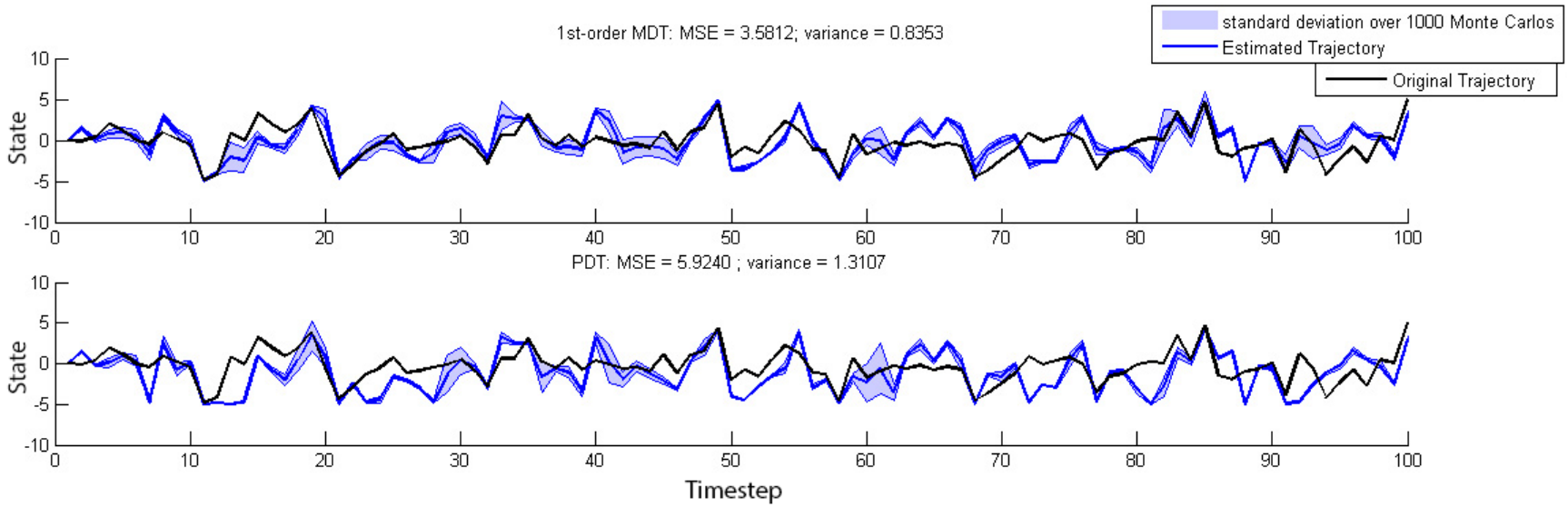
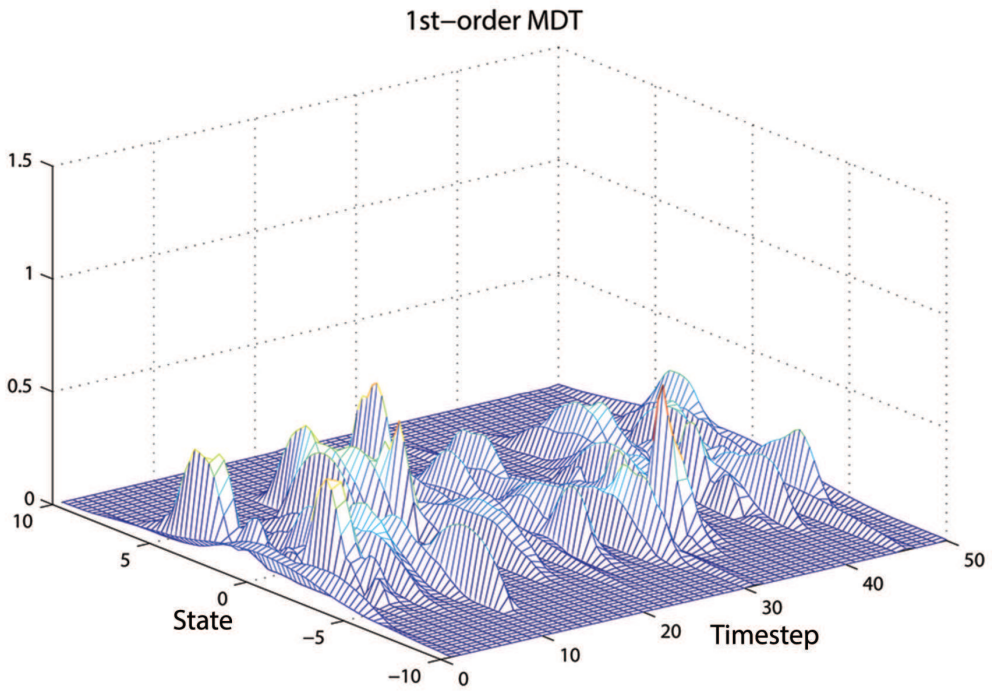
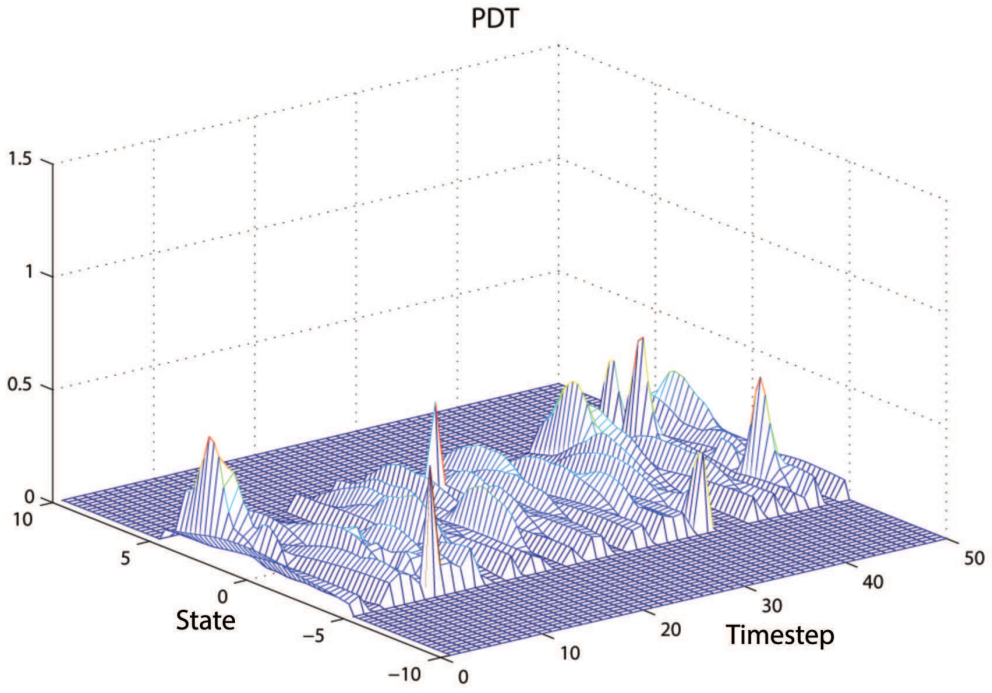


Figure 6. Constrained state-estimation of the nonlinear dynamic system in (3.11). State estimation for 1000 Monte Carlo simulations. The shading represents a two standard deviations band. Top row: 1^{st}-order MDT (MSE=3.58, $\sigma = 0.83$); bottom row: PDT (MSE=5.36, $\sigma = 2.03$).



(a) 1^{st}-order MDT



(b) PDT

Figure 7. Constrained state-estimation of the nonlinear dynamic system in (3.11). State posterior densities evolving over time for (a) 1^{st}-order MDT and (b) PDT.

Chapter 4

The Marginalized Particle Filter
Application to EEG Dynamic Source Localization

4.1 Introduction

Electroencephalography is a widely used technology in neurology because it is non-invasive, portable, low cost, and has a high temporal resolution. There are many technological and clinical applications of EEG, including brain computer interface (BCI) technologies, which have the potential to interface the brain with computers directly using the EEG measurements as inputs, and the diagnosis of specific electroclinical syndromes, such as epilepsy. The main sources of EEG potentials, which are measured at the scalp, emanate from the simultaneous current flows of many neighboring neurons in the same direction. The total electric current in an activated region of the brain can often be modeled by a mathematical current dipole with an adequate dipole moment. Also, many of these current dipoles representing microscopic current flows with the same orientation can be replaced by an equivalent current dipole [12].

Currently, there are two major research areas in modeling brain neural generators [13]. The first modeling technique utilizes imaging models, which explains the data using a dense set of current dipoles distributed at fixed locations. The second technique uses a parametric approach and models the dense set of current dipoles in each region using one equivalent current dipole. Although the imaging-based techniques can create a detailed map of the brain's neuronal activity, the parametric approach provides a direct mapping of the EEG measurements to a small number of parameters. By using the parametric approach, the equivalent current models can provide more intuitive interpretations that explain the

electrical activity in the brain and can be fostered in emerging technologies, such as BCI systems [14]. However, an important challenge of the parametric approach is the estimation of the locations of the equivalent dipole sources using EEGs recorded from the scalp. Locating the dipole sources in the brain is very useful in clinical and research applications. For instance, accurate localization of the epileptic focus can be used to plan presurgical investigations. Similarly, obtaining information about the regions of the brain that are associated with various sensory modalities may shed light on the functioning of the brain. Ideally, we would like to localize the equivalent dipole source positions, with no *a priori* knowledge of the active regions in the brain and without performing exhaustive search of the entire head volume. There are many popular methods, such as the Multiple Signal Classification (MuSiC) algorithm and its modifications, the solution to inverse problems [39], the construction of spatial filters (beamformers) by data-independent [40] or data-driven methods [41] and blind source separation techniques [42], which can estimate the equivalent dipole moment amplitude and orientation [43]. However, these methods assume that the correct spatial positions of the dipoles were found or given beforehand. There is also a very strict assumption in these models that the correct spatial positions of the dipoles do not move. By lifting the constraint on the spatial position of the current dipoles, a more accurate representation of the equivalent current dipole model can be used.

We formulate the brain source localization problem as a (nonlinear) state-space model, where the positions and moments of the neural generators constitute the unknown or hidden state and the EEG measurements are the observations of the system. In a Bayesian context, inference of the hidden state given a realization of the observations relies upon the posterior density function (pdf) [1]. For systems with linear dynamics and Gaussian noise,

the posterior distribution is Gaussian whose mean and covariance can be computed using the Kalman filter. For systems with non-linear dynamics, several approximations can be used: the linearized Extended Kalman Filter (EKF) [1] and the Unscented Kalman filter (UKF), which uses a deterministic set of samples, called *sigma points*, to propagate the posterior mean and covariance [2]. Approximation methods, however, have an underlying assumption of linearity or Gaussianity, an unrealistic assumption in the source localization problem. A Monte Carlo method, called the *Particle Filter* (PF) has emerged, which uses the concept of Sequential Importance Sampling (SIS) to estimate the posterior pdf using a finite number of weighted samples [3]. In particular, the PF does not make any assumptions about the pdfs or the linearity of the system model.

The power of the PF, however, comes at a computational cost. In particular, the number of particles needed for the estimation increases super-exponentially with the dimension of the state [11]. This problem is commonly known as the "curse of dimensionality", and makes it unreasonable to use the Particle filter for tracking problems in high dimensional spaces. In the context of EEG source localization, the dimension of the state space is six times the number of dipoles, causing the tracking of even two dipoles (12-dimensional problem) to be inaccurate unless a very large number of particles are used. To deal with the high-dimensionality issue, we propose to *marginalize* out the states in the system that are linear with respect to the measurements [44]. This allows the linear states in the state-space model to be estimated optimally using the Kalman Filter, whereas the non-linear states are estimated using the PF. By decreasing the dimensionality of the state, less particles can be used, allowing a decrease in computation time. Simulation results show that even a two dipole model cannot be localized using the traditional PF, but can be tracked accurately

using the marginalized PF.

This chapter is organized as follows: Section 4.2 formulates the EEG source localization problem as a state-space model. Section 4.3 reviews the theory of the Particle Filter. This section also outlines the shortcomings of the PF in terms of computational complexity and curse of dimensionality. Section 4.4 presents the Marginalized Particle Filter, and its algorithm. Section 4.5 discusses the simulation results and compares between the traditional and marginalized PFs.

4.2 EEG Source Localization Model

Given M equivalent active dipoles in the brain, the measured multichannel EEG signal $\boldsymbol{y}_k$ from n_y sensors at time k can be modeled as follows:

$$\boldsymbol{y}_k = \sum_{m=1}^{M} \boldsymbol{L}_m(\boldsymbol{d}_k(m))\boldsymbol{s}_k(m) + \boldsymbol{v}_k, \tag{4.1}$$

where M is the total number of dipoles, $\boldsymbol{d}_k(m)$ is a 3×1 spatial position vector in the brain of dipole m at discrete time k. Each dipole $m = 1, \cdots, M$, is defined as $\boldsymbol{d}_k(m) = [x_k(m), y_k(m), z_k(m)]^t$. $\boldsymbol{L}_m(\boldsymbol{d}_k(m))$ is the $n_z \times 3$-dimensional lead-field matrix for the m^{th} dipole. $\boldsymbol{s}_k(m)$ is a 3×1-dimensional moment of the m^{th} dipole at time k. $\boldsymbol{v}_k$ is a zero-mean white Gaussian noise with covariance $\boldsymbol{R}_k$. Most notably, the components of the leadfield matrix $\boldsymbol{L}_m$ are non-linear functions of the dipole locations, electrodes' positions and head position [45]. The EEG measurement equation described in (4.1) can be written concisely as

$$\boldsymbol{y}_k = \boldsymbol{L}_k(\boldsymbol{d}_k)\boldsymbol{s}_k + \boldsymbol{v}_k, \tag{4.2}$$

where $\boldsymbol{d}_k = [\boldsymbol{d}_k(1), ..., \boldsymbol{d}_k(M)]^T$ is a $3M \times 1$-dimensional vector representing the 3D location coordinates of the M dipoles at time k. $\boldsymbol{L}(\boldsymbol{d}_k) = [\boldsymbol{L}_1(\boldsymbol{d}_k(1)), \cdots, \boldsymbol{L}_M(\boldsymbol{d}_k(M))]$ is a $n_x \times 3M$ leadfield matrix of the M dipoles at time k, and $\boldsymbol{s}_k = [\boldsymbol{s}_k(1), \cdots, \boldsymbol{s}_k(M)]$ is a $3M \times 1$-dimensional vector of brain source locations of the M dipoles. The hidden state (to be estimated) is given by the dipole positions and moments: $\boldsymbol{x}_k = [\boldsymbol{d}_k^t, \boldsymbol{s}_k^t]^t$.

It is important to note that the measurements $\boldsymbol{y}_k$ are linear with respect to the dipole moments $\boldsymbol{s}_k$ and non-linear with respect to the dipole spatial positions $\boldsymbol{d}_k$. This model allows us to consider marginalization of the linear states to be estimated by the Kalman filter, thus reducing the dimensionality of the state estimated by the PF. We further assume a random walk model for the state transition dynamics. A random walk model does not assume any *a priori* knowledge about the source locations and moments. The EEG source localization state-space model is then given by

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \boldsymbol{w}_k, \\ \boldsymbol{y}_k = \boldsymbol{L}(\boldsymbol{d}_k)\boldsymbol{s}_k + \boldsymbol{v}_k, \end{cases} \tag{4.3}$$

where $\boldsymbol{w}_k$ is the state noise at time k, assumed to be zero-mean, white Gaussian process. The goal is to use the model in (4.3) to estimate, at every time instant, the dipole locations $\boldsymbol{d}_k$ and moments $\boldsymbol{s}_k$ given the EEG measurements $\boldsymbol{y}_k$.

4.3 The Particle Filter

Consider the following discrete-time state-space model defined by possibly nonlinear state and measurement equations

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_k, \\ \boldsymbol{y}_k &= \boldsymbol{h}_k(\boldsymbol{x}_k) + \boldsymbol{v}_k, \end{aligned} \tag{4.4}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{n_x}$ and $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$ represent, respectively, the hidden state and the measurement vectors. The functions f_k and h_k are known, possibly non-linear, mappings; and $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are realizations of the zero-mean process and measurement noise with known pdfs. We wish to estimate the state of the system $\boldsymbol{x}_k$ at every time step k, given the history of measurements $\boldsymbol{Y}_k = \{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_k\}$.

In the Bayesian framework, the optimal state estimate is given by the mean of the posterior density $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$. Using Bayes rule, the posterior distribution, at time k, can be computed sequentially from the following two-step prediction-update formula [38]:

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1}) d\boldsymbol{x}_{k-1} \tag{4.5}$$

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})}{\int p(\boldsymbol{y}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) d\boldsymbol{x}_k}. \tag{4.6}$$

Unfortunately, except for the linear case, these equations are only a conceptual solution, due to the intractability of the integrals defined. The PF is a Monte Carlo method that represents the posterior pdf, at time k, using a set of N *particles* $\{\boldsymbol{x}_k^{(i)}\}_{i=1}^N$ and their associated weights

$\{w_k^{(i)}\}_{i=1}^N$:

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}), \tag{4.7}$$

where $\delta(.)$ is the Dirac delta function and N is the number of particles [3]. The conditional mean estimate at time k is then given by

$$\hat{\boldsymbol{x}}_k = E[\boldsymbol{x}_k|\boldsymbol{Y}_k] \approx \sum_{i=1}^{N} w_k^{(i)} \boldsymbol{x}_k^{(i)}. \tag{4.8}$$

Ideally, the particles should be sampled from the posterior distribution $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ itself; however it is not available to us directly. Therefore, another known pdf, called the *importance distribution* or *proposal distribution*, $q(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_k)$, is used to sample the particles: $\boldsymbol{x}_k^{(i)} \sim q(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_k)$. To make up for the difference between the importance distribution and the posterior density, the weight of each particle $\boldsymbol{x}_k^{(i)}$ is computed as

$$\bar{w}_k^{(i)} = w_{k-1}^{(i)} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)}) p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{q(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_k)}, \tag{4.9}$$

where $\{\bar{w}_k^{(i)}\}_{i=1}^N$ are the un-normalized weights [28]. The normalized weights in Eq. (4.7) are given by

$$w_k^{(i)} = \frac{\bar{w}_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}}.$$

The most popular choice for the importance distribution is the prior, i.e., $q(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_k) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$, leading to the importance weights, in Eq. (4.9), to simplify

to $\bar{w}_k^{(i)} = w_{k-1}^{(i)} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)})$ [3]. PFs using the prior as the importance distribution are usually referred to as *generic PFs*. It can be shown that, as long as the support of the importance distribution contains the support of the posterior distribution, the PF estimates converge to the optimal Bayesian estimates in the mean square error sense when the number of particles tends to infinity [1]. The price to be paid for the flexibility and numerical power of the PF is computational. This computational cost is especially prohibitive in higher dimensional state spaces, where the number of particles needed increases super-exponentially with the dimension of the state [11].

4.4 The Marginalized Particle Filter

The main idea of the marginalized PF (MPF) is to partition the state vector as $\boldsymbol{x}_k = [(\boldsymbol{x}_k^l)^t, (\boldsymbol{x}_k^n)^t]^t$, where $\boldsymbol{x}_k^l$ denotes the state variable partition with conditionally linear dynamics and $\boldsymbol{x}_k^n$ denotes the state variable partition with non-linear dynamics. Using Bayes' Theorem, we can then marginalize out the linear state variables and estimate them using the Kalman Filter. This technique is reminiscent of Rao-Blackwellization because it is related to the Rao-Blackwell formula [22]. Let us consider the following marginalization model:

$$\begin{cases} \boldsymbol{x}_k^n &= \boldsymbol{f}_k(\boldsymbol{x}_{k-1}^n) + \boldsymbol{w}_k^n, \\ \boldsymbol{x}_k^l &= \boldsymbol{A}_k(\boldsymbol{x}_{k-1}^n)\boldsymbol{x}_{k-1}^l + \boldsymbol{w}_k^l, \\ \boldsymbol{y}_k &= \boldsymbol{h}_k(\boldsymbol{x}_k^n) + \boldsymbol{C}_k(\boldsymbol{x}_k^n)\boldsymbol{x}_k^l + \boldsymbol{v}_k, \end{cases} \tag{4.10}$$

where f_k and h_k are non-linear functions, $\boldsymbol{A}_k(\boldsymbol{x}_{k-1}^n)$ is conditionally linear on $\boldsymbol{x}_{k-1}^l$, and $\boldsymbol{C}_k(\boldsymbol{x}_k^n)$ is conditionally linear on $\boldsymbol{x}_k^l$. The system and measurement noise are assumed to

be white Gaussian processes distributed according to

$$\boldsymbol{v}_k \sim N(0, \boldsymbol{R}_k),$$

$$\boldsymbol{w}_k = \begin{bmatrix} \boldsymbol{w}_k^l \\ \\ \boldsymbol{w}_k^n \end{bmatrix} \sim N(0, \boldsymbol{Q}_k), \quad \boldsymbol{Q}_k = \begin{bmatrix} \boldsymbol{Q}_k^l & \mathbf{0} \\ \\ \mathbf{0} & \boldsymbol{Q}_k^n \end{bmatrix}.$$

The posterior pdf of the state $\boldsymbol{x}_k$ can then be found as $p(\boldsymbol{x}_k^l, \boldsymbol{X}_k^n | \boldsymbol{Y}_k)$, where $\boldsymbol{X}_k^n = \{\boldsymbol{x}_0^n, \boldsymbol{x}_1^n, ..., \boldsymbol{x}_k^n\}$ and the marginal of which is $p(\boldsymbol{x}_k^l, \boldsymbol{x}_k^n | \boldsymbol{Y}_k)$. By marginalizing out the conditionally linear states $\boldsymbol{x}_k^l$ using Bayes' theorem, we have

$$p(\boldsymbol{x}_k^l, \boldsymbol{X}_k^n | \boldsymbol{Y}_k) = p(\boldsymbol{x}_k^l | \boldsymbol{X}_k^n, \boldsymbol{Y}_k) p(\boldsymbol{X}_k^n | \boldsymbol{Y}_k). \tag{4.11}$$

The distribution $p(\boldsymbol{x}_k^l | \boldsymbol{X}_k^n, \boldsymbol{Y}_k)$ is analytically tractable because it is conditioned on the non-linear states $\boldsymbol{X}_k^n$ and, therefore, can be found optimally using the Kalman Filter [44]. The distribution $p(\boldsymbol{X}_k^n | \boldsymbol{Y}_k)$ depends only on the nonlinear states and can be estimated using the PF. Since the non-linear state vector $\boldsymbol{x}_k^n$ is of smaller dimension than the original state $\boldsymbol{x}_k$, the dimensionality of the PF, and thus its computational complexity, has been reduced.

Since the pdf $p(\boldsymbol{x}_k^l | \boldsymbol{X}_k^n, \boldsymbol{Y}_k)$ is conditioned on the non-linear states $\boldsymbol{X}_k^n$, we have $\boldsymbol{A}_k(\boldsymbol{x}_{k-1}^n)$ and $\boldsymbol{C}_k(\boldsymbol{x}_k^n)$ as fixed constant matrices, thus allowing for a statistically optimal estimate of the linear state. The Kalman Filter algorithm to compute the optimal estimate of $\boldsymbol{x}_k^l$ given the non-linear states is given by

Given the initial conditions $\boldsymbol{x}_{0|0}^l$ and $\boldsymbol{P}_{0|0}$, compute the prediction equations

$$\begin{aligned} \boldsymbol{x}_{k|k-1}^l &= \boldsymbol{A}_k(\boldsymbol{x}_{k-1}^n)\boldsymbol{x}_{k-1|k-1}^l \\ \boldsymbol{P}_{k|k-1} &= \boldsymbol{A}_k(\boldsymbol{x}_{k-1}^n)\boldsymbol{P}_{k-1|k-1}\boldsymbol{A}_k(\boldsymbol{x}_{k-1}^n)^T + \boldsymbol{Q}_k^l \end{aligned} \tag{4.12}$$

and the update equations

$$\begin{aligned} \boldsymbol{S}_k &= \boldsymbol{C}_k(\boldsymbol{x}_k^n)\boldsymbol{P}_{k|k-1}\boldsymbol{C}_k(\boldsymbol{x}_k^n)^T + \boldsymbol{R}_k \\ \boldsymbol{K}_k &= \boldsymbol{P}_{k|k-1}\boldsymbol{C}_k(\boldsymbol{x}_k^n)^T\boldsymbol{S}_k^{-1} \\ \boldsymbol{x}_{k|k}^l &= \boldsymbol{x}_{k|k-1}^l + \boldsymbol{K}_k\left(\boldsymbol{y}_k - \boldsymbol{h}_k(\boldsymbol{x}_k^n) - \boldsymbol{C}_k(\boldsymbol{x}_k^n)\boldsymbol{x}_{k|k-1}^l\right) \\ \boldsymbol{P}_{k|k} &= \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k\boldsymbol{C}_k(\boldsymbol{x}_k^n)\boldsymbol{P}_{k|k-1} \end{aligned} \tag{4.13}$$

The optimal estimate of $\boldsymbol{x}_k^l$ is then given by $\boldsymbol{x}_{k|k}^l$.

The second conditional probability in Eq. (4.11), $p(\boldsymbol{X}_k^n|\boldsymbol{Y}_k)$, can be expressed as

$$p(\boldsymbol{X}_k^n|\boldsymbol{Y}_k) \propto p(\boldsymbol{y}_k|\boldsymbol{x}_k^n)p(\boldsymbol{x}_k^n|\boldsymbol{x}_{k-1}^n)p(\boldsymbol{X}_{k-1}^n|\boldsymbol{Y}_{k-1}) \tag{4.14}$$

We use the distribution $p(\boldsymbol{x}_k^n|\boldsymbol{x}_{k-1}^n)$ as the importance density. We have $p(\boldsymbol{x}_k^n|\boldsymbol{x}_{k-1}^n) = \mathcal{N}(\boldsymbol{f}_k(\boldsymbol{x}_{k-1}^n), \boldsymbol{Q}_k^n)$, where $\mathcal{N}(\boldsymbol{x}, \boldsymbol{C})$ denotes the normal distribution with mean $\boldsymbol{x}$ and covariance matrix $\boldsymbol{C}$. The weights of the particles are calculated as

$$\bar{w}_k^{(i)} = w_{k-1}^{(i)} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{n,(i)}), \tag{4.15}$$

where

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k^n) = \mathcal{N}\left(\boldsymbol{h}_k(\boldsymbol{x}_k^n) + \boldsymbol{C}_k(\boldsymbol{x}_k^n)\boldsymbol{x}_{k|k-1}^l, \boldsymbol{S}_k\right).$$

for $i = 1, 2, ..., N$ **do**
  Initialize particles $\boldsymbol{x}_0^{n,(i)} \sim p_{\boldsymbol{x}_0^n}(\boldsymbol{x}_0^n)$ and set $\{\boldsymbol{x}_{0|-1}^{l,(i)}, \boldsymbol{P}_{0|-1}^{(i)}\} = \{\bar{\boldsymbol{x}}_0^l, \bar{\boldsymbol{P}}_0^l\}$
end
for $k = 1, 2, ...$ **do**
  for $i = 1, 2, ..., N$ **do**
    Evaluate the weights $\tilde{w}_k^{(i)}$ using Eq. (4.15).
  end
  Normalize the weights $\boldsymbol{w}_k \leftarrow \frac{\boldsymbol{w}_k}{\sum_{i=1}^N \tilde{w}_k^{(i)}}$.
  for $i = 1, 2, ..., N$ **do**
    Update $\hat{\boldsymbol{x}}_{k|k}^l$ using Eq. (4.13).
    Update $\boldsymbol{P}_{k|k}$ using Eq. (4.13).
    Calculate the mean estimate $\hat{\boldsymbol{x}}_k$ using Eq. (4.16).
    Sample $\hat{\boldsymbol{x}}_{k+1}^{n,(i)} \sim p(\boldsymbol{x}_{k+1}^n | \boldsymbol{X}_k^n, \boldsymbol{Y}_k)$.
    Update $\hat{\boldsymbol{x}}_{k+1|k}^l$ using Equation (4.12).
    Update $\boldsymbol{P}_{k+1|k}$ using Equation (4.12).
  end
end

Figure 8. The marginalized particle filter algorithm.

The optimal state estimate at time k is then given by

$$\hat{\boldsymbol{x}}_k = \sum_{i=1}^{N} w_k^{(i)} \left[\boldsymbol{x}_{k|k}^{l,(i)}, \boldsymbol{x}_k^{n,(i)} \right]^t. \tag{4.16}$$

The Marginalized PF algorithm is summarized in Figure 8.

A resampling step may be introduced after normalizing the weights to avoid degeneracy of the PF [1].

4.5 Results and Discussion

4.5.1 Simulation Results on Synthetic Data. In our experiments, we considered one and two moving dipoles generating the observed EEG measurements. We compared the performance of the "traditional" Particle Filter algorithm, where both the linear and nonlinear components of the state are estimated using the PF, with the proposed Marginalized PF. Tables 1 and 2 show the starting and ending positions of the dipoles in the brain and moment amplitudes and frequencies, respectively. The moments are assumed to be sinusoidal waveforms with varying amplitudes and frequencies. Dipole 1 is used in the one dipole test and Dipoles 2a and 2b are used in the two dipoles test. We performed 100 Monte Carlo simulations and computed the Mean Squared Error (MSE) of the true state $\boldsymbol{x}_k$ versus the estimated state using the following formula MSE$= \frac{1}{100}\sum_{i=1}^{100}[(\hat{\boldsymbol{x}}_{i,k}-\boldsymbol{x}_k)^2+(\hat{\boldsymbol{y}}_{i,k}-\boldsymbol{y}_k)^2+(\hat{\boldsymbol{z}}_{i,k}-\boldsymbol{z}_k)^2]$, where $\hat{\boldsymbol{x}}_{i,k}$ is the estimated state at time k at the i^{th} Monte Carlo run.

For the one dipole test, Dipole 1 was tracked using the Marginalized PF and the classical PF with $N = 500$ particles and run over 100 Monte Carlo simulations. The tracking results are shown in Figure 9. The MSE curve of the 100 Monte Carlo runs is shown in Figure 10. We also note that the convergence times of the Marginalized PF are considerably faster than the PF due to the smaller state dimension. In this case, the classical PF converges to the true state and only performs marginally worse once locked on.

The real performance enhancement of the Marginalized PF occurs when tracking two (or more) dipoles. Dipole 2a and Dipole 2b were tracked using the Marginalized PF and the classic PF with 500 particles. The simulation was repeated for 100 Monte Carlo runs. The mean state tracking results for the Marginalized PF and PF are shown in Figures 11

Table 1

Moving dipoles starting and ending positions (in meters) in the brain

	Start Position (m)			End Position (m)		
	x	y	z	x	y	z
Dipole 1	-0.01	0.07	0.04	-0.04	-0.07	0.01
Dipole 2a	-0.01	0.07	0.01	-0.04	-0.05	0.04
Dipole 2b	0.04	-0.05	0.01	0.01	0.07	0.04

and 12, respectively. The average MSE is shown in Figure 13. The advantage of using the Marginalized PF is more obvious in this case. Not only does the PF produce a larger MSE, but in 12D it fails to track the state. The Marginalized PF is able to track the state because it is only using the PF to estimate the non-linear part and uses the Kalman filter to estimate the linear part. For the EEG localization model presented In this thesis, half of the states are linear, allowing the Marginalized PF to calculate the same result using a reduced 6D state space model.

4.5.2 Application to Real EEG Data. In this section, we apply the proposed Marginalized PF algorithm to real EEG data recorded from twelve female subjects (20-28 years old). The experimental setup was designed by Santos *et al*. [46] for their study on subject attention and perception using Visually Evoked Potentials (VEP). VEPs are transient components in the EEG generated in response to visual stimulus. The subjects were exposed to a sequence of images of different facial expressions (neutral, fearful and disgusted) superimposed on houses as shown in Figure 15. The images were divided into two experimental

Table 2

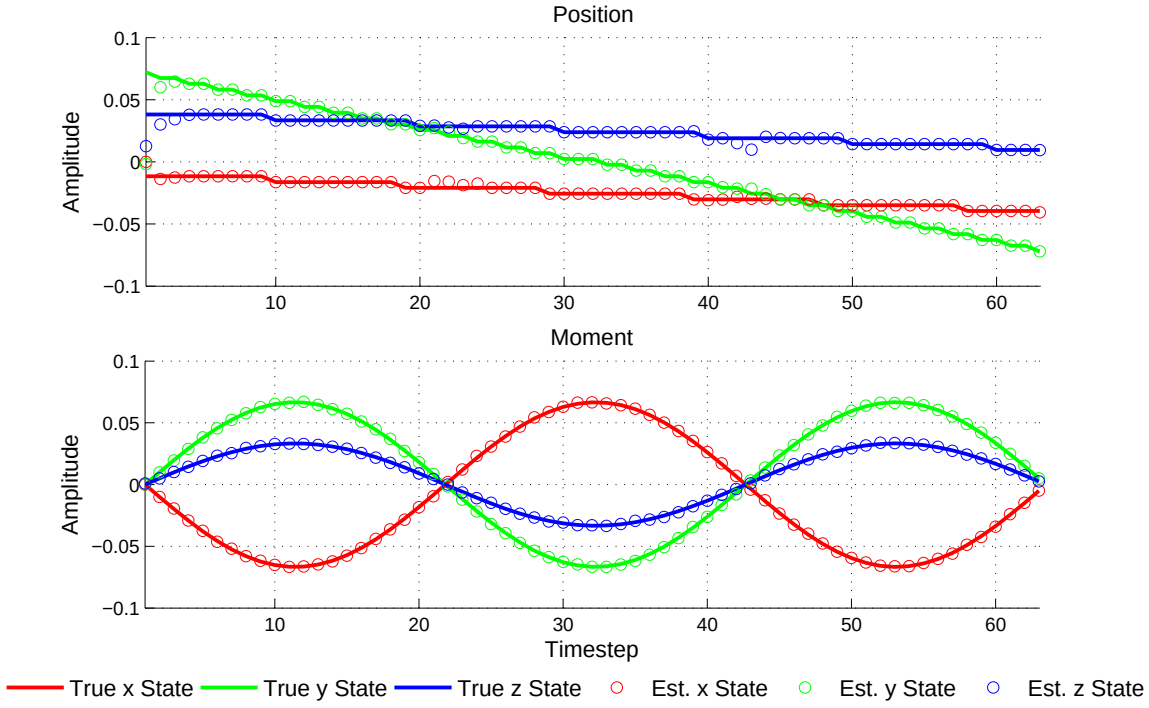
Moving dipoles moment amplitude and frequency

	Sinusoid Moment			
	Amp. x	Amp. y	Amp. z	Freq.
Dipole 1	-0.6	0.6	0.4	10 Hz.
Dipole 2a	-0.6	0.6	0.4	10 Hz.
Dipole 2b	0.6	-0.6	0.4	15 Hz.

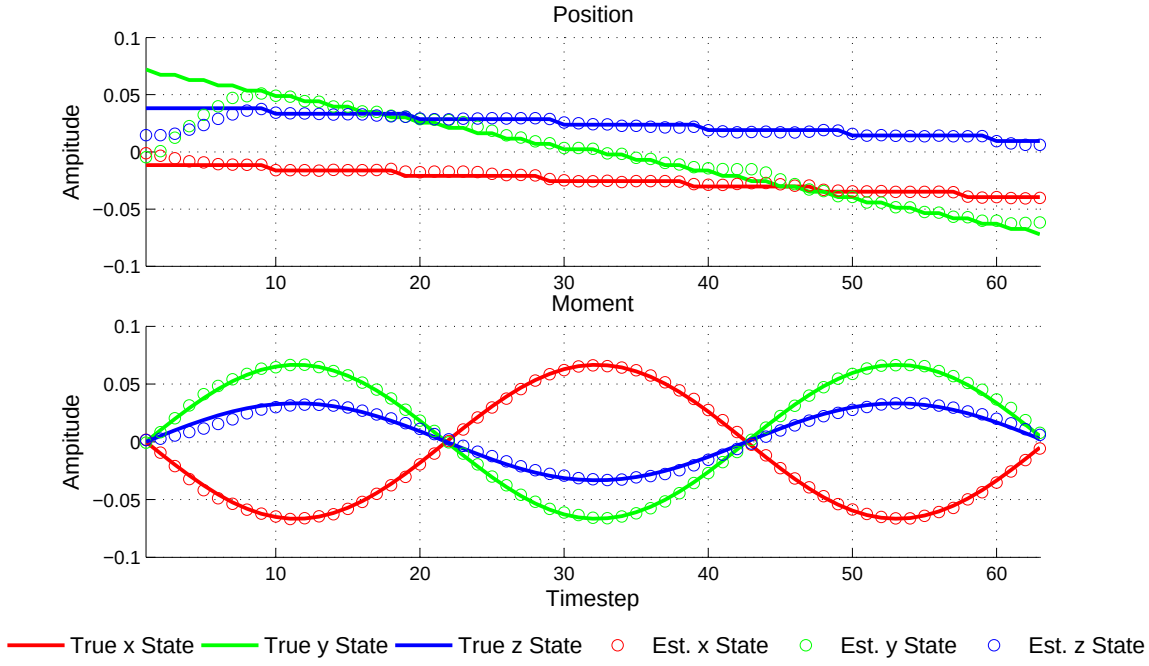
blocks. In the first, the participants were required to attend to the houses (ignoring the faces) and in the other they were required to attend to the faces (ignoring the houses). The participants task was to determine, on each trial, if the current house or face (depending on the experimental block) is the same as the one presented on the previous trial. Each trial lasts 1600 ms (400 samples with sampling rate 250 Hz) comprising a pre-stimulus interval of 148 ms (37 samples) and post-stimulus onset interval of 1452 ms. Only trials with correct responses were included in the data set. EEG signals were recorded from 16 channels (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2; F7, F8, Fz, Cz, Pz, Oz) and two Electrooculogram (EOG) channels (horizontal and vertical EOG) located according to the 10/20 International system (see Figure 14). The raw brain signals were first eye-movement corrected, baseline compensated and segmented into trials using NeuroScan software. Trials with excessive EOG artifacts were eliminated. In order to reduce the measurement noise and to eliminate the electrical supply frequency of 50 Hz, Principal Component Analysis (PCA) and a But-

terworth band-pass filter between 1- 40 Hz were applied. Since the primary brain task in this experiment is perception of visual stimulus, the neural activity is supposed to happen in the visual cortex. Therefore, the Marginalized PF algorithm is expected to estimate the strongest dipoles that may have originated the registered VEPs in the occipital brain zone which corresponds to the visual cortex.

We considered the estimation of two sources for each patient. We used 1000 particles in the Marginalized PF for the real data. The results of of the Marginalized particle filter estimate for two patients over three trials are shown in Figure 16 and Figure 17. It is very interesting to observe that the dipole coordinates are located in the zone of the primary visual cortex as shown in Figure 18. In this sense, the proposed approach seems to be coherent in tracking the brain sources over time. Another noteworthy observation is the fact that the 3D locations of the dipoles does not vary significantly over time or between trials. This is due to the fact that the EEG experimental setup was designed to study attention and perception. We postulate that in order to observe significant or abrupt changes in brain source locations, we need to design an experiment, where two or more areas of the brain (e.g., visual and motor) are invoked. We also observed that there is no significant variability between the subjects in the locations of the brain dipoles. However, there was a notable variability in the moments between the subjects.



(a) Marginalized Particle Filter



(b) Classical Particle Filter

Figure 9. Tracking of a one dipole: Trajectory (top) and moment (bottom). Top figure: marginalized PF and bottom figure: classical PF. The solid lines represent the true state and the circles represent the estimated state for each dimension. The x-dimension is in red, the y-dimension in green, and the z-dimension in blue.

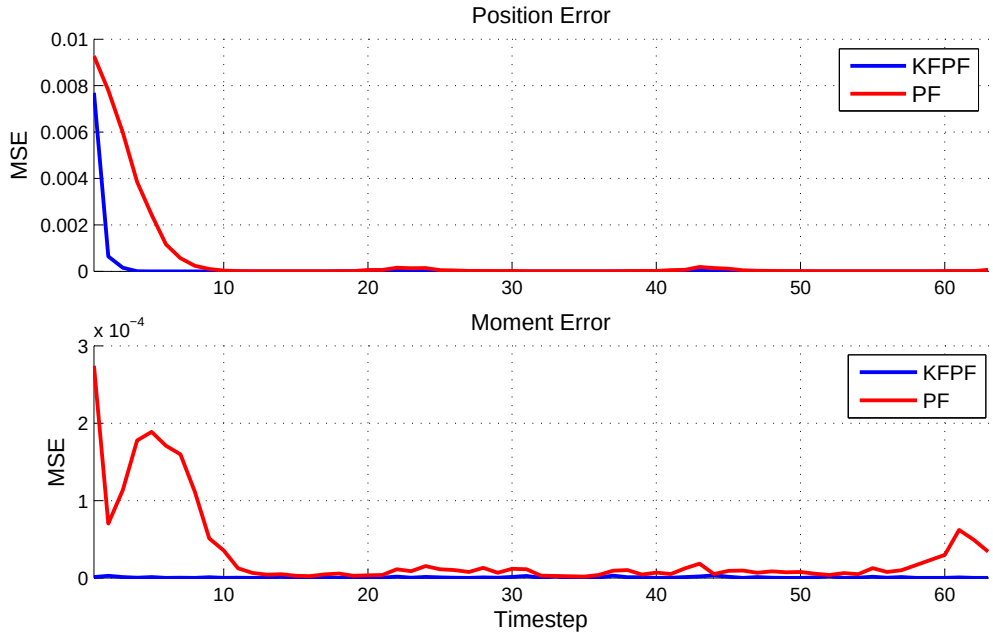
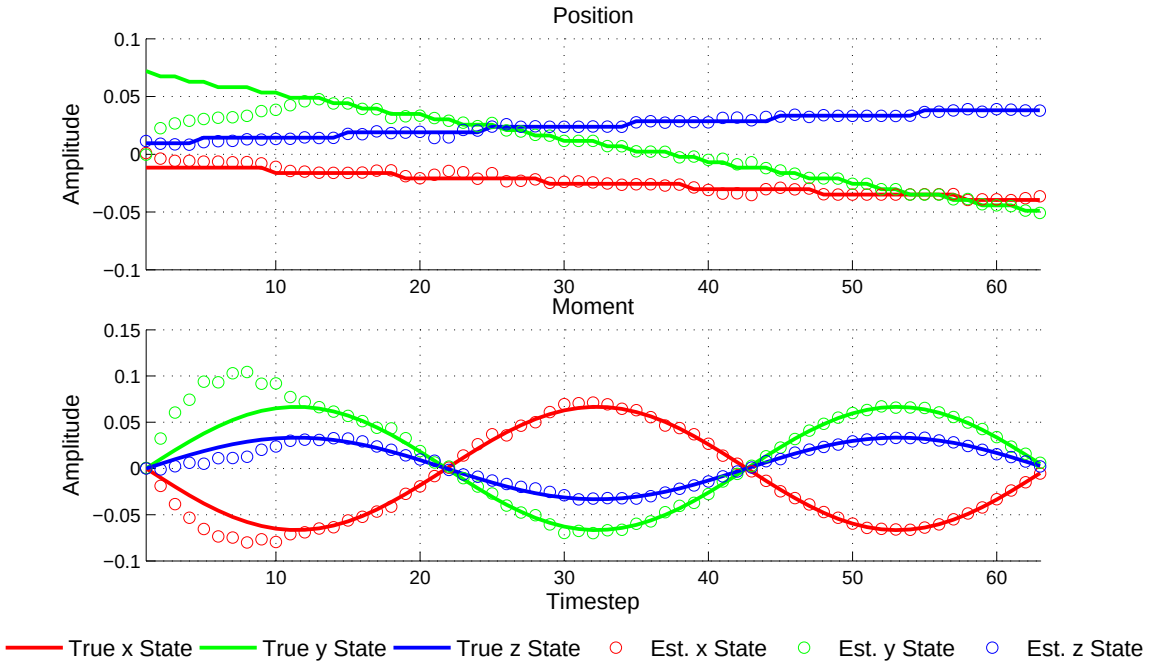
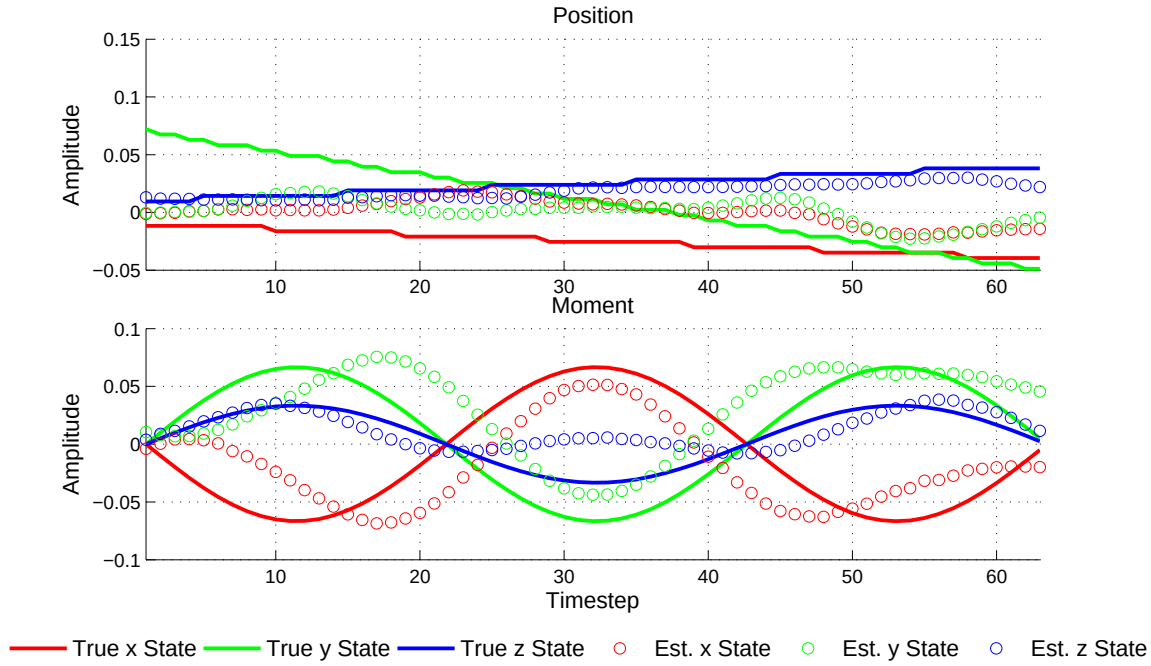


Figure 10. Mean Squared Error of tracking one dipole: Location (top) and moment (bottom). The Marginalized Particle filter is in blue and the classical Particle Filter is in red.

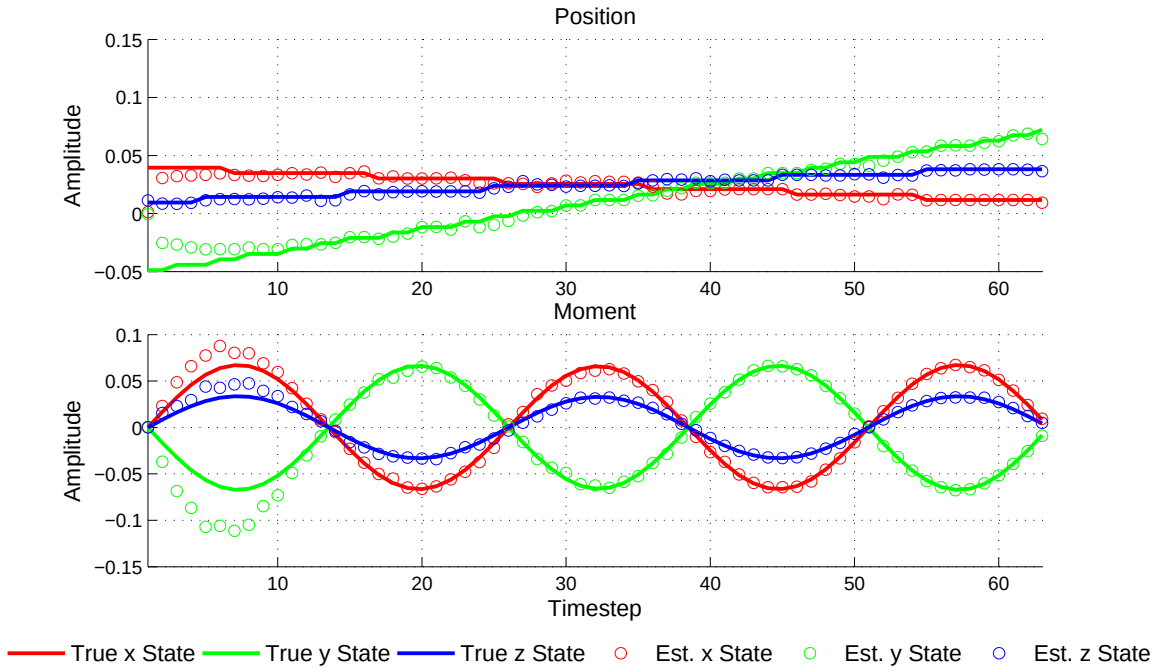


(a) Marginalized Particle Filter

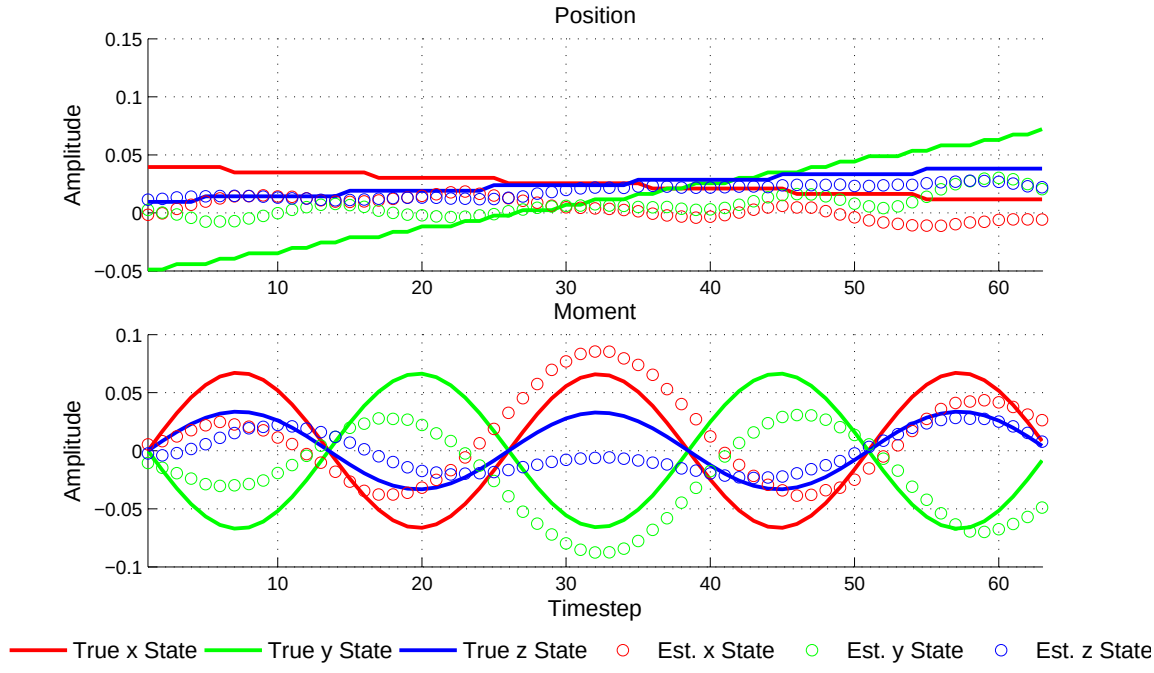


(b) Classical Particle Filter

Figure 11. Tracking of two dipoles: first dipole trajectory (top) and moments (bottom). Top figure: marginalized PF and bottom figure: classical PF. The solid lines represent the true state and the circles represent the estimated state for each dimension. The x-dimension is in red, the y-dimension in green, and the z-dimension in blue.

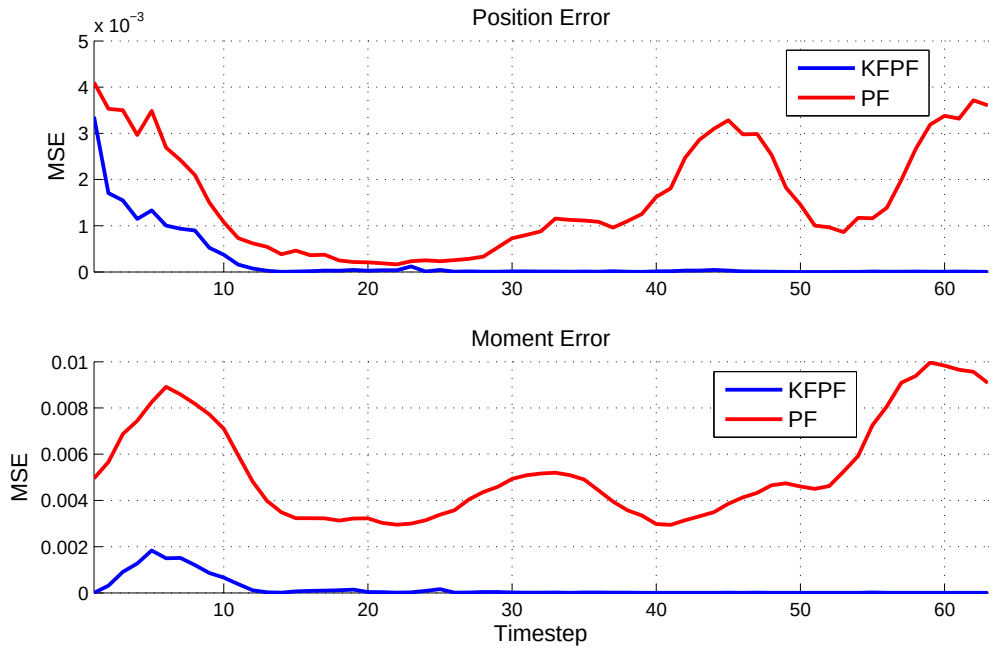


(a) Marginalized Particle Filter

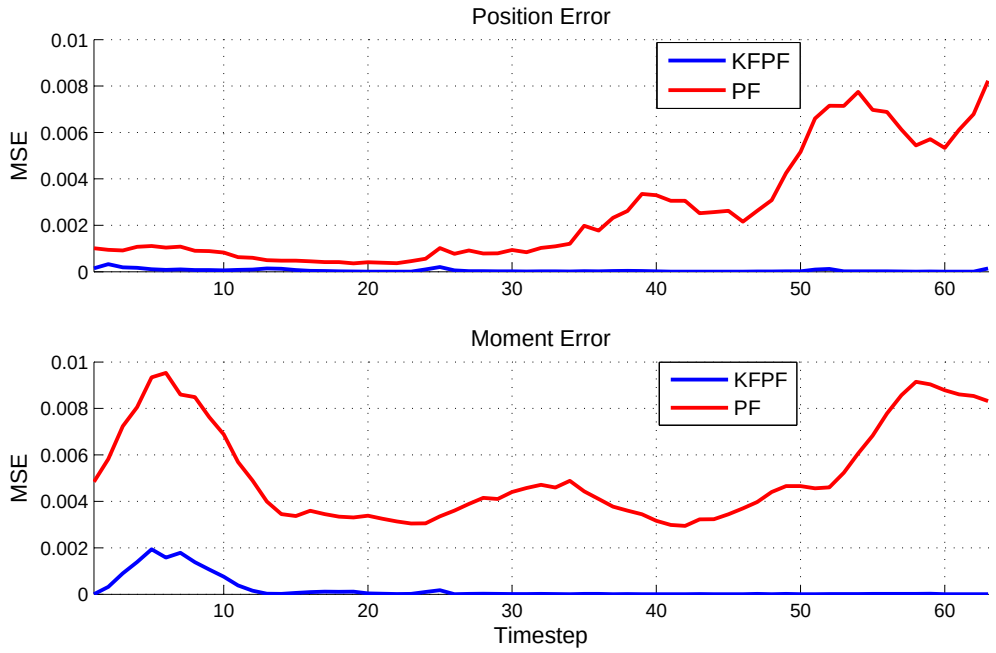


(b) Classical Particle Filter

Figure 12. Tracking of two dipoles: second dipole trajectory (top) and moments (bottom). Top figure: marginalized PF and bottom figure: classical PF. The solid lines represent the true state and the circles represent the estimated state for each dimension. The x-dimension is in red, the y-dimension in green, and the z-dimension in blue.

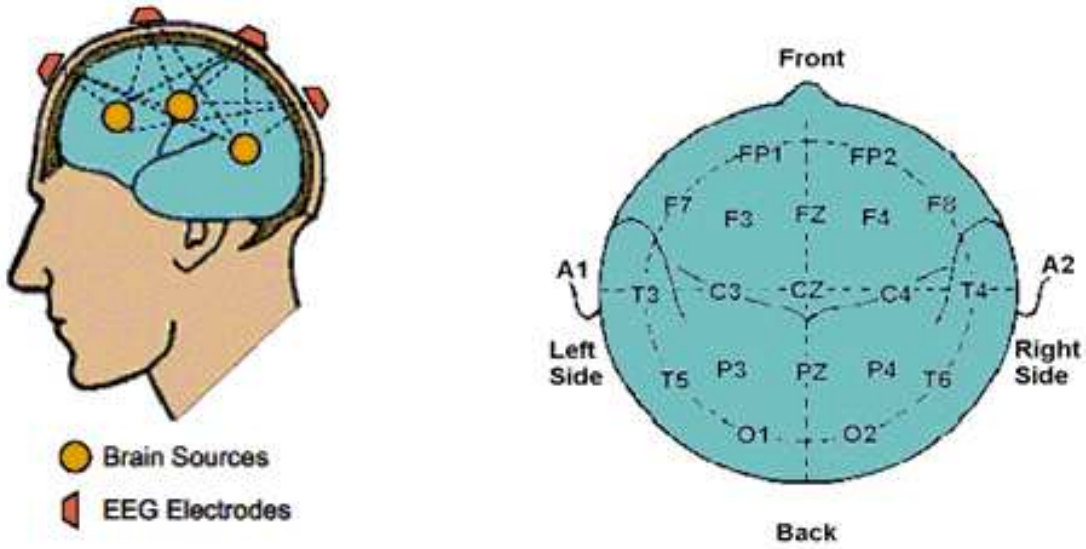


(a) First dipole



(b) Second dipole

Figure 13. Mean Squared Error of Tracking the position (top) and moment (bottom) of the first dipole (top figure) and second dipole (bottom figure). The Marginalized PF tracking is in blue and and the classical PF tracking is in red.



(a) (b)

Figure 14. The head model: (a) Depiction of a realistic EEG experiment; (b) Spatial scalp location of the EEG electrodes.

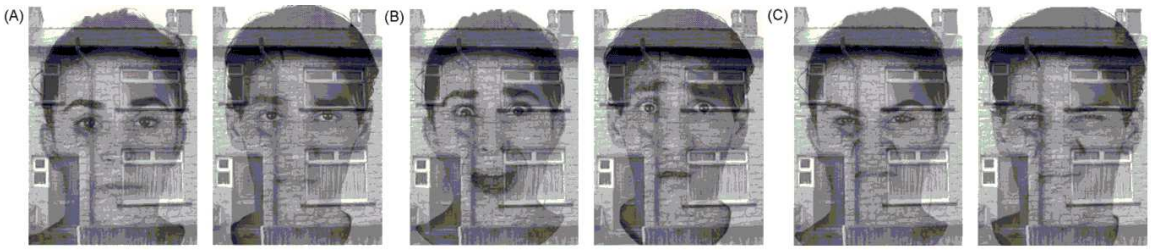
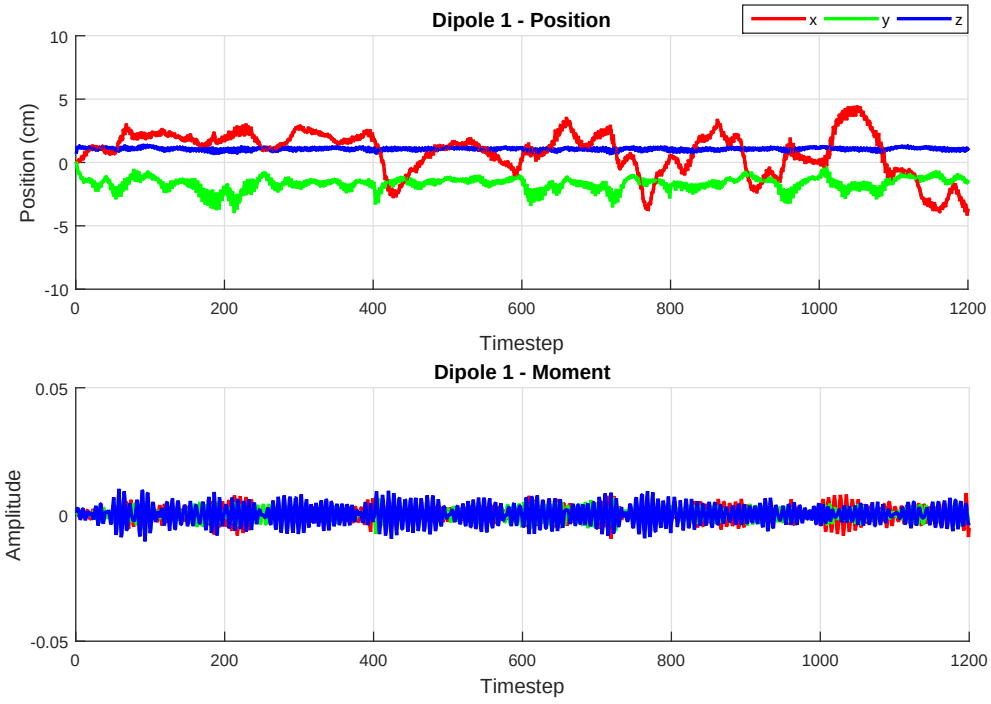
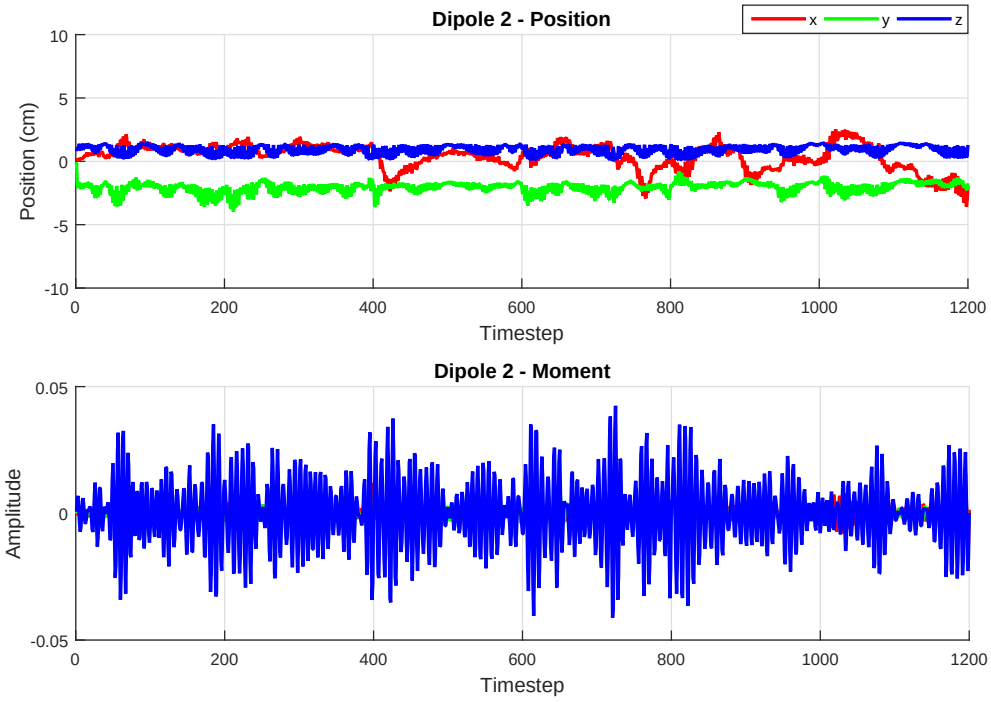


Figure 15. A set of images shown to the subjects in the experimental setup.

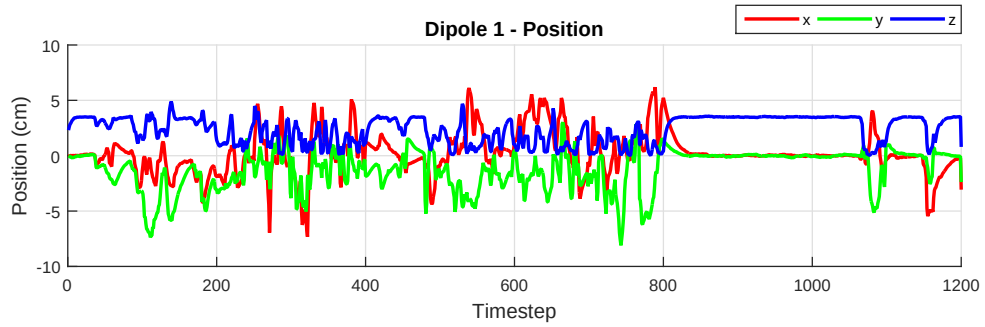


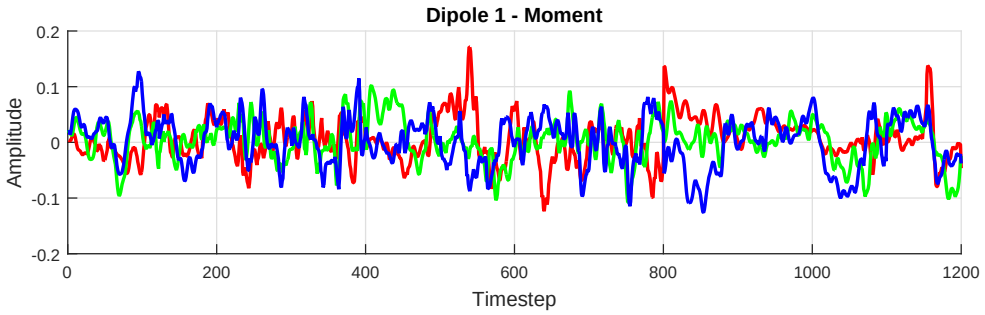
(a) Subject 1 - Dipole 1



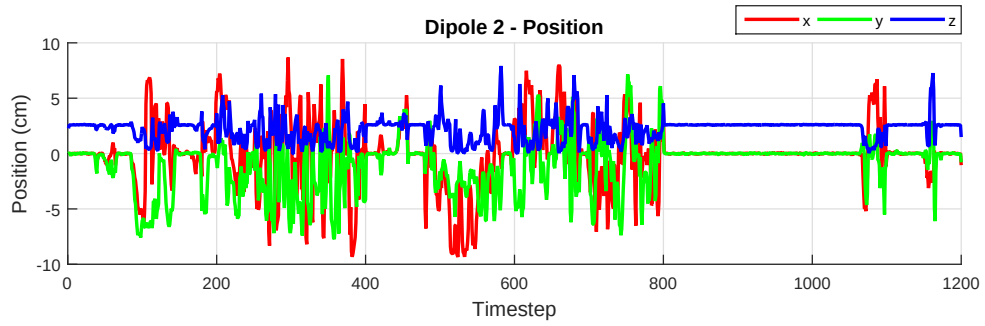
(b) Subject 1 - Dipole 2

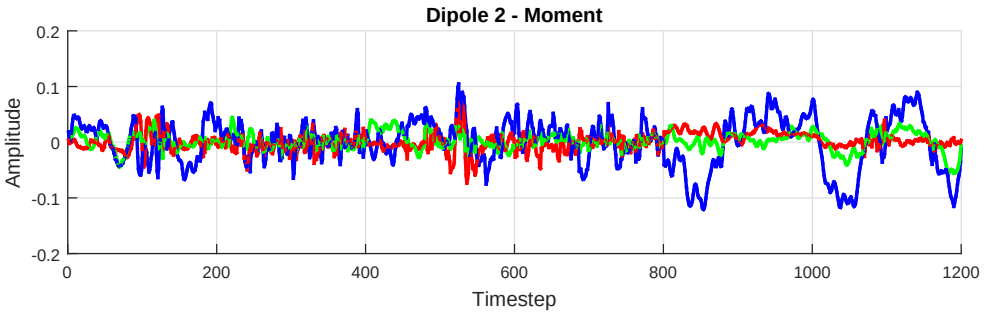
Figure 16. Dipoles and moments tracking over 3 trials for subject one. The first row shows tracking of the locations of the dipoles. The second row shows the dipole moments over time. The average location of dipole 1 is $(\bar{x}_1^1 = 0.83, \bar{y}_1^1 = -1.67, \bar{z}_1^1 = 1)$ and dipole 2 is $(\bar{x}_2^1 = 0.37, \bar{y}_2^1 = -2, \bar{z}_2^1 = 0.9)$.





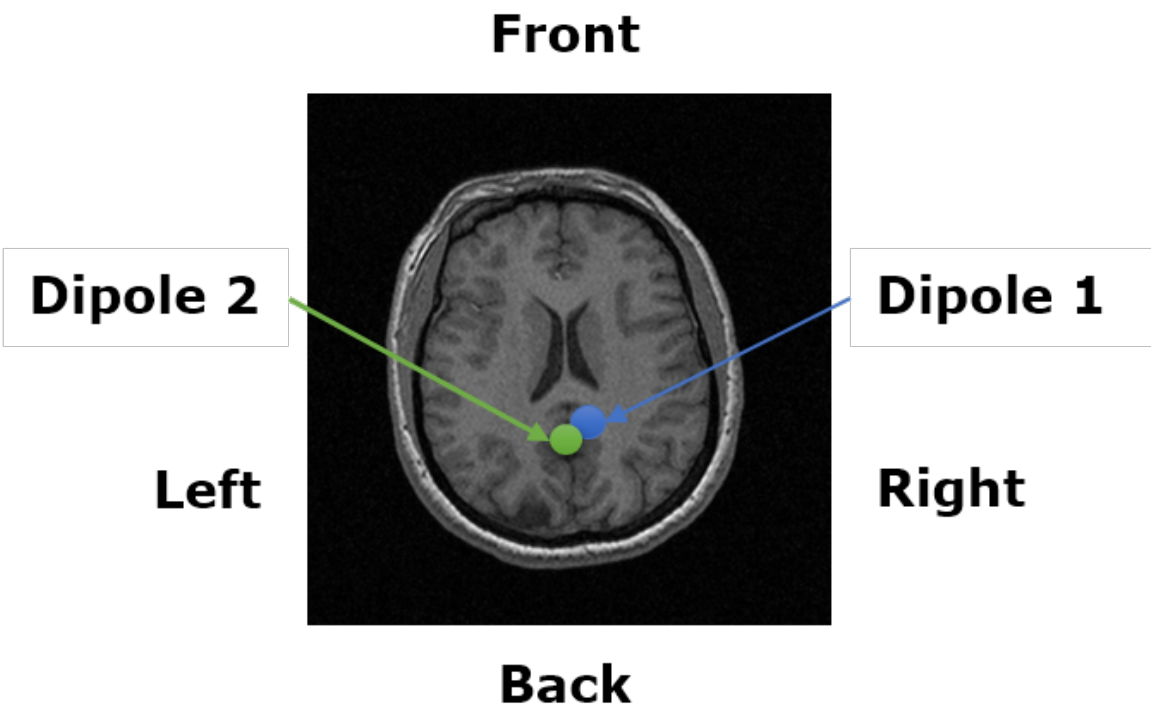
(a) Subject 2 - Dipole 1



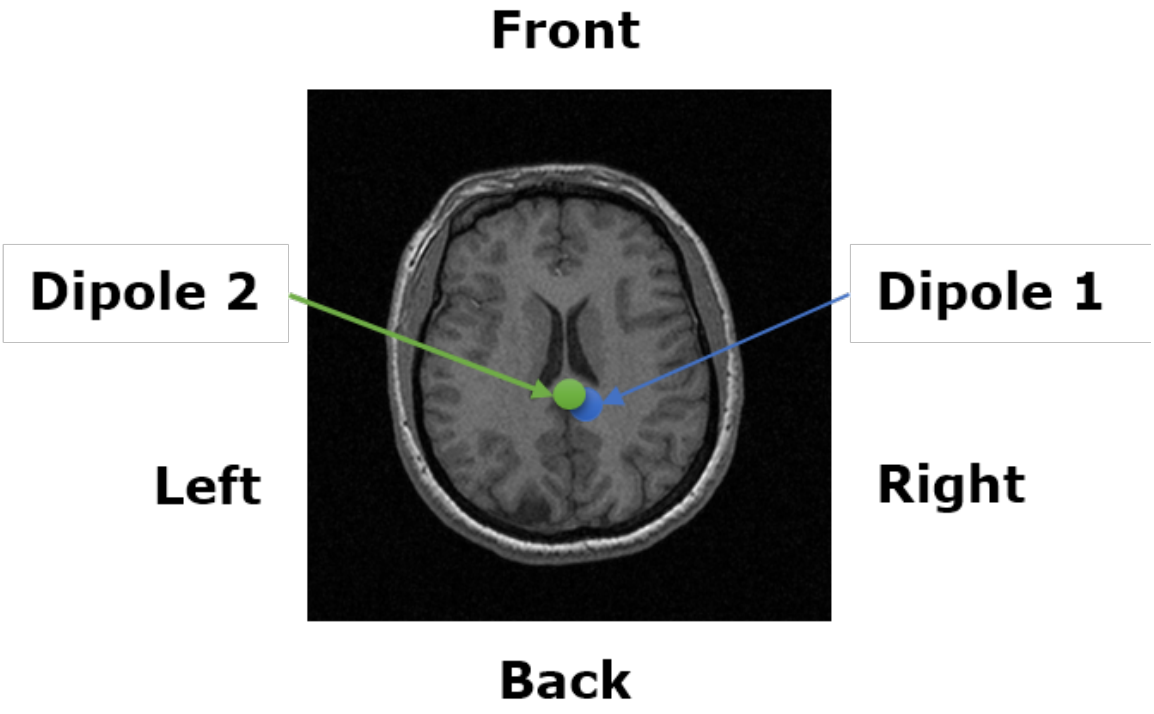


(b) Subject 2 - Dipole 2

Figure 17. Dipoles and moments tracking over 3 trials for subject two. The first row shows tracking of the locations of the dipoles. The second row shows the dipole moments over time. The average location of dipole 1 is $(\bar{x}_1^2 = 0.2, \bar{y}_1^2 = -1.25, \bar{z}_1^2 = 2.3)$ and dipole 2 is $(\bar{x}_2^2 = -0.07, \bar{y}_2^2 = -1, \bar{z}_2^2 = 2.2)$.



(a) Subject 1



(b) Subject 2

Figure 18. Axial view of primary visual cortex zone. The arrows point at the estimated source locations.

Chapter 5

Conclusion and Future Perspectives

The Particle Filter framework is a powerful tool for non-linear and non-Gaussian state-space modeling. In this thesis, we presented novel solutions to two research areas related to the Particle Filter: constrained dynamical systems and high-dimensionality in mixed linear and non-linear state-spaces.

The first problem of constrained state-spaces is crucial in real-world dynamical systems that are subject to constraints arising from physical principles and process restrictions. Therefore, constraints must be taken into account in order to obtain physically meaningful estimation results. In this thesis, we argued that constraining all particles is equivalent to constraining the posterior distribution of the state. This may lead either to a stronger condition or to a different (unrelated) condition; both of which result in incorrect estimation of the posterior distribution of the state. We, subsequently, advanced a new approach, Mean Density Truncation (MDT), which imposes the desired constraints on the conditional mean estimate without further restricting the posterior density of the state; and hence preserving the convergence properties of the particle filter towards the optimal posterior density of the state. Future research directions include efficient algorithmic implementation of the MDT approach and its variants.

We motivated the second problem, dealing with high-dimensional state-spaces in particle filtering, by considering for the first time moving dipoles in the brain. This dynamical framework of brain sources would contribute to a physiologically more plausible brain technologies such as source-based BCI. Non-linear tracking algorithms, notably the Particle Filter (PF), are emerging as promising solutions in the localization of equivalent current

dipole models from EEG measurements. However, the numerical nature of particle filters, which constitutes their strength for multidimensional numerical integration, becomes their major weakness in high-dimensional state-space models. It has been shown that the number of particles needs to grow super-exponentially with the dimension of the state in order to maintain tracking accuracy. This problem is known as the "curse of dimensionality" of the PF. In the EEG source localization problem, every dipole that is being tracked is a 6-dimensional vector; making the dimension of the state $6M$, with M being the number of dipoles. In this thesis, we proposed to handle the curse of dimensionality problem in the PF by taking advantage of the linear substructures in the EEG state space model. In this state-space model, the measurements are linear with respect to the moments of each dipole and non-linear with respect to the dipole position. The moments were "Marginalized" out and computed optimally using the Kalman filter. The remaining non-linear positions were then estimated numerically using the classical Particle Filter. We showed that the Marginalized Particle Filter was able to successfully track two dipoles with no a priori knowledge of their positions or moments using only 500 particles. The classic PF failed in tracking this same system, due to the high-dimensionality of the problem and the small number of particles used. The Marginalized PF is based on a reformulation of the problem so that the linear part of the state vector is estimated by an optimal estimator (e.g., a Kalman filter) and the nonlinear part of the state vector is estimated by a nonlinear estimator (a particle filter). The Marginalized PF works over a reduced state space which leads to a reduction of the computational complexity. In our future research, we will also investigate Markov Chain Monte Carlo (MCMC) techniques, which have the potential to alleviate the curse of dimensionality of the PF. On the experimental side, we will design an EEG experiment

that involves more than one functional cortex of the brain (e.g., visual followed by motor stimuli) that will allow us to topographically track the dynamics of the brain sources as they move from one cortex of the brain to the other. Also, it would be interesting to see if there is any correlation between the sex, age and health of the subjects and the dynamics of their dipoles in the brain. This subject information will be saved in future experiments.

In summary, the main contributions of this thesis are:

- A new method of constrained Particle Filtering, called *Mean Density Truncation*, that computes the estimate of the state that satisfies the constraint without imposing stronger conditions on the posterior distribution of the state.

- A new approach to EEG source localization using a non-linear state-space model and using the Particle Filter to track the moving dipoles and their corresponding moments in the 3D volume of the brain.

- Using the Marginalized Particle Filter, which takes advantage of the linear substructures in the EEG state-space model, in order to reduce the dimension of the state estimated by the particle filter.

- Applying the MPF algorithm to real EEG data and showing that the obtained results correspond to the regions of the brain that are expected to be active in the experiment.

References

[1] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *Handbook of Nonlinear Filtering*, vol. 12, 2009, pp. 656–704.

[2] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems." in *Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls*, 1997.

[3] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[4] E. A. Wan and R. van der Merwe, "The unscented Kalman filter for nonlinear estimation," 2000.

[5] L. S. Wang, Y. T. Chiang, and F. R. Chang, "Filtering method for nonlinear systems with constraints," *IEE Proceedings in Control Theory and Applications*, vol. 149, no. 6, pp. 525 – 531, November 2002.

[6] L. Xu, X. R. Li, Z. Duan, and J. Lan, "Modeling and state estimation for dynamic systems with linear equality constraints," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2927 – 2939, June 2013.

[7] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Modeling and Control*. John Wiley and Sons, Inc., 2005.

[8] N. Gupta, "Constrained Kalman filtering and predicting behaviour in agent-based financial market models," Ph.D., Oxford University, 2007.

[9] H. F. Lopez and R. S. Tsay, "Particle filters and Bayesian inference in financial econometrics," *Journal of Forecasting*, vol. 30, pp. 168–209, 2010.

[10] P. Vachhani, R. Rengaswamy, V. Gangwal, and S. Narasimhan, "Recursive estimation in constrained nonlinear dynamical systems," *AIChE Journal*, vol. 51, no. 3, pp. 946–959, March 2005.

[11] T. Bengtsson, P. Bickel, and B. Li, *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*. Institute of Mathematical Statistics, 2008, vol. 2, pp. 316–334.

[12] S. Sanei and J. Chambers, *EEG Signal Processing*. John Wiley and Sons, 2007.

[13] A. Galka, O. Yamashita, T. Ozaki, R. Biscay, and P. Valds-Sosa, "A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering," *Neuroimage*, vol. 23, no. 2, pp. 435–453, 2004.

[14] B. Kamousi, Z. Liu, and B. He, "An EEG inverse solution based brain computer interface," *The International Journal of Bioelectromagnetism*, vol. 7, no. 2, pp. 292–294, 2005.

[15] S. J. Kiebel, J. Daunizeau, C. Phillips, and K. J. Friston, "Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG," *NeuroImage*, vol. 39, no. 2, pp. 728–741, January 2008.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc, 1993, vol. 1, ch. The Bayesian Philosophy, pp. 309–340.

[17] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME - Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[18] D. Simon, *Optimal state estimation:* K*alman,* H∞*, and nonlinear approaches*. Wiley-Interscience, 2006.

[19] G. A. Einicke, *Smoothing, Filtering and Prediction - Estimating The Past, Present and Future*. InTech, February 2012.

[20] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer, 2001, ch. An Introduction to Sequential Monte Carlo Methods, pp. 3–14.

[21] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, July 2000.

[22] S. G. A. Doucet and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," in *Statistics and Computing*, vol. 10, no. 3, 2000, pp. 197–208.

[23] S. J. Julier and J. J. LaViola, "On Kalman filtering with nonlinear equality constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2774 – 2784, June 2007.

[24] D. Simon and D. L. Simon, "Kalman filtering with inequality constraints for turbofan engine health estimation," *IEE Proceedings in Control Theory and Applications*, vol. 153, no. 3, pp. 371 – 378, May 2006.

[25] J. Porrill, "Optimal combination and constraints for geometrical sensor data," *International Journal of Robotics Research - Special Issue on Sensor Data Fusion*, vol. 7, no. 6, pp. 66 – 77, December 1988.

[26] G. Rasool, K. Iqbal, N. Bouaynaya, and G. White, "Neural drive estimation using the hypothesis of muscle synergies and the state-constrained Kalman filter," in *International IEEE EMBS Neural Engineering Conference*, November 2013.

[27] D. Simon and T. L. Chia, "Kalman filtering with state equality constraints," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 1, pp. 128 – 136, January 2002.

[28] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, September 1998.

[29] L. Lang, W.-S. Chen, B. R. Bakshi, P. K. Goel, and S. Ungarala, "Bayesian estimation via sequential Monte Carlo sampling Constrained dynamic systems," *Automatica*, vol. 43, no. 9, pp. 1615–1622, September 2007.

[30] X. Shao, B. Huang, and J. M. Lee, "Constrained Bayesian state estimation: A comparative study and a new particle filter based approach," *Journal of Process Control*, vol. 20, no. 2, pp. 143–157, 2010.

[31] J. Prakash, S. C. Patwardhan, and S. L. Shah, "Constrained state estimation using particle filters," *oral presentation in IFAC-2008*, 2008.

[32] O. Straka, J. Dunik, and M. Simandl, "Truncated unscented particle filter," in *American Control Conference*, 2011, pp. 1825–1830.

[33] J. H. Kotecha and P. Djuric, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, 2003.

[34] M.-A. Beyer and G. Reinig, "Constrained particle filtering using gaussian sum approximations," in *Proceedings of the UKACC International Conference on Control*, 2008.

[35] S. Ungarala, "A direct sampling particle filter from approximate conditional density function supported on constrained state space," *Computers & Chemical Engineering*, vol. 35, no. 6, pp. 1110–1118, June 2011.

[36] O. Straka, J. Dunik, and M. Simandl, "Truncation nonlinear filters for state estimation with nonlinear inequality constraints," *Automatica*, vol. 48, no. 2, pp. 273–286, February 2012.

[37] F. Papi, M. Podt, Y. Boers, and G. Battistello, "On constraints exploitation for particle filtering based target tracking," in *International Conference on Information Fusion*, July 2012, pp. 455 – 462.

[38] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-gaussian Bayesian state estimation," *IEEE Proceedings on Radar and Signal Processing*, vol. 140, no. 2, pp. 107 – 113, April 1993.

[39] C. Michel, M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. G. de Peralta, "EEG source imaging," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2195–222, October 2004.

[40] J. Gross and A. Ioannides, "Linear transformations of data space in MEG," *Physics in Medicine and Biology*, vol. 44, no. 8, pp. 2081–2097, August 1999.

[41] B. V. Veen, W. V. Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filter," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 867–880, September 1997.

[42] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applicaions*. John Wiley & Sons, July 2002.

[43] S. Baillet, J. Mosher, and R. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, 2001.

[44] T. Schon, F. Gustafsson, and P. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, July 2005.

[45] C. Chen, A. Pogosyan, L. Zrinzo, S. Tisch, and et al., "Intra-operative recordings of local field potentials can help localize the subthalamic nucleus in parkinsons disease surgery," *Exp. Neurology*, pp. 214–221, 2006.

[46] I. Santos, J. Iglesias, E. I. Olivares, and A. Young, "Differential effects of object-based attention on evoked potentials to fearful and disgusted faces," *Neuropsychologia*, vol. 46, no. 5, p. 14681479, 2008.