BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF CONTROL AND INSTRUMENTATION
ÚSTAV AUTOMATIZACE A MĚŘICÍ TECHNIKY

MONTE CARLO-BASED IDENTIFICATION STRATEGIES FOR STATE-SPACE MODELS
MONTE CARLO IDENTIFIKAČNÍ STRATEGIE PRO STAVOVÉ MODELY

DOCTORAL THESIS (SHORT VERSION)
DIZERTAČNÍ PRÁCE (KRÁTKÁ VERZE)

AUTHOR — **Ing. MILAN PAPEŽ**
AUTOR PRÁCE

ADVISOR — **prof. Ing. PETR PIVOŇKA, CSc.**
VEDOUCÍ PRÁCE

BRNO 2018

ABSTRACT

State-space models are immensely useful in various areas of science and engineering. Their attractiveness results mainly from the fact that they provide a generic tool for describing a wide range of real-world dynamical systems. However, owing to their generality, the associated state and parameter inference objectives are analytically intractable in most practical cases. The present thesis considers two particularly important classes of nonlinear and non-Gaussian state-space models: conditionally conjugate state-space models and jump Markov nonlinear models. A key feature of these models lies in that—despite their intractability—they comprise a tractable substructure. The intractable part requires us to utilize approximate techniques. Monte Carlo computational methods constitute a theoretically and practically well-established tool to address this problem. The advantage of these models is that the tractable part can be exploited to increase the efficiency of Monte Carlo methods by resorting to the Rao-Blackwellization. Specifically, this thesis proposes two Rao-Blackwellized particle filters for identification of either static or time-varying parameters in conditionally conjugate state-space models. Furthermore, this work adopts recent particle Markov chain Monte Carlo methodology to design Rao-Blackwellized particle Gibbs kernels for state smoothing in jump Markov nonlinear models. The kernels are then used to facilitate maximum likelihood parameter inference in the considered models. The resulting experiments demonstrate that the proposed algorithms outperform related techniques in terms of the estimation precision and computational time.

KEYWORDS

Sequential Monte Carlo, particle Markov chain Monte Carlo, nonlinear and non-Gaussian state-space models, conditionally conjugate state-space models, jump Markov nonlinear models, state and parameter inference, identification of static and time-varying parameters

PAPEŽ, Milan. *Monte Carlo-Based Identification Strategies for State-Space Models*. Brno, 2018, 43 p. Doctoral thesis (short version). Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Control and Instrumentation. Advised by prof. Ing. Petr Pivoňka, CSc.

CONTENTS

INTRODUCTION

A state-space model is a generic tool to embody our intuition about time-space dependent and stochastic behaviour of a real-world dynamical system. The necessary step towards drawing conclusions about such a system is to observe data on it. The model and data are then used to carry out various statistical inference objectives, including the estimation of latent states and parameters. However, dynamical systems are mostly nonlinear and non-Gaussian, which makes the associated inference objectives analytically intractable and therefore poses a real challenge on the design of high-fidelity approximation techniques.

Sequential Monte Carlo (SMC) methodology [19] is particularly well suited for this aim. SMC methods provide approximate solutions based on generating a collection of random samples. A range of convergence results [93] for these approaches proves that as the number of samples increases, quantities of interest are approximated with increasingly high precision. This ability comes naturally with the question of high computational complexity. Fortunately, the computational power is still growing—albeit not as rapidly in the sense of the Moore's law as before, but rather in terms of parallel architectures [29]—which makes this question relative, but relevant mainly when the problem is high-dimensional or the computational resources are limited. However, there exist particularly useful and general classes of nonlinear and non-Gaussian state-space models that contain analytically tractable substructures. This feature is commonly utilized in the design of SMC methods in order to improve their computational efficiency through the Rao-Blackwellization [18]. In such cases, an algorithm relying on this principle can have the same estimation precision as an algorithm without this improvement but at a lower computational cost. The requirement of providing highly reliable approximate solutions to various inference objectives in state-space models has recently recorded a significant conceptual shift, namely the particle Markov chain Monte Carlo (MCMC) methodology [2]. Particle MCMC algorithms can be seen as exact approximations of the ideal MCMC procedures. These methods run an SMC method at each iteration in order to produce a single sample of a quantity of interest, making them highly computationally demanding. Therefore, even a slight improvement in the estimation accuracy of these methods can have a profound impact on the computational time.

This thesis is about algorithm design. The aim is to develop computationally more efficient Monte Carlo techniques for two generic classes of nonlinear and non-Gaussian state-space models. The first class is formed by the conditionally conjugate state-space models. Their characteristic feature lies in that they contain an algebraically tractable substructure with respect to the parameters but an intractable substructure with respect to the unobserved states. These models have been applied to a broad range of diverse practical problems, including computer code performance tuning [40], flu epidemics tracking [28], vehicle navigation systems [74], target tracking [70], online recommendation services [97], estimation of the remaining useful life of batteries [63], learning of cellular dynamics in system biology [68], web activity modeling [66], optimization of portfolio returns [45], to mention a few. The second class is given by the jump Markov nonlinear models. Their key aspect is that they are formed by a finite number of nonlinear and non-Gaussian state-space configurations that switch according to a discrete-valued Markov chain. These configurations constitute the intractable part of the model, whereas the discrete chain forms the tractable part. These models have also become substantially popular in various practical applications, such as learning of consumption growth dynamics [46], traffic behavior analysis through video surveillance [7], virus-cell fusion identification [36], molecular bioimaging [83], detection of abrupt changes in financial markets [64], sensor networks [92], simultaneous localization and tracking [54], terrain-based navigation [11], estimation of drivers' behavior [55], etc. The design of precise and fast computational strategies can provide a substantial increase in efficiency in the above applications, potentially decreasing the cost of associated hardware tools.

This document is a short version of the author's doctoral thesis. The content is formed by a number of *separate* chapters that summarize novel methods and solutions. These chapters are substantially shortened versions of the author's published papers. For more details, see the full document.

1 A PROJECTION-BASED PARTICLE FILTER TO ESTIMATE STATIC PARAMETERS IN CONDITIONALLY CONJUGATE STATE-SPACE MODELS

Particle filters constitute today a well-established class of techniques for state filtering in non-linear state-space models. However, online estimation of static parameters under the same framework represents a difficult problem. The solution can be found to some extent within a category of state-space models allowing us to perform parameter estimation in an analytically tractable manner, while still considering non-linearities in data evolution equations. Nevertheless, the well-known particle path degeneracy problem complicates the computation of the statistics that are required to estimate the parameters. The present chapter proposes a simple and efficient method which is experimentally shown to suffer less from this issue.

1.1 Introduction

Context

A state-space model (SSM, [14]) embodies a popular statistical tool for describing dynamical systems in diverse application areas such as signal processing, econometrics, and bioinformatics. This model is especially useful for defining the relation between observed data, latent (unobserved) data, and unknown static parameters. The estimation of the states and parameters based on the observations is the primary task in the aforementioned application areas. A rather general class of state-space models is formed when they contain a *tractable* substructure characterizing the parameters and an *intractable* substructure describing nonlinear, and possibly non-Gaussian, data (observed and unobserved). Such models are herein referred to as the conditionally conjugate SSMs (CCSSMs). Their key feature is that the tractable substructure facilitates recursive updates of statistics related to the posterior distribution of the parameters, but the intractable substructure requires us to use approximate inference to make the parameter estimation feasible. This chapter considers particle filters (PFs, [25]) to perform the approximate inference.

A number of PF-based methods for estimating static parameters in the considered class of models have been developed in the last years [84, 31, 16]. These algorithms utilize the tractable substructure to compute a set of the posterior statistics based on the observations and latent state trajectories simulated by a PF. However, the trajectories are known to suffer from the particle path degeneracy [3], if they are constructed in a single forward pass of a PF. This issue also affects the computation of the posterior statistics, and methods relying on such a principle therefore usually deliver poor performance.

So far, we have only referred to methods that are most relevant to the algorithm proposed in the present chapter. For a thorough overview of PF-based parameter estimation, we refer the reader to a series of recent survey papers [47, 43, 33]. Importantly, there has recently been an increased interest in designing methods based on particle smoothing [59] or particle Markov chain Monte Carlo [2], which are efficient in dealing with the degeneracy issue. However, these procedures are offline, processing repeatedly a fixed batch of data, and since this chapter is concerned with the online estimation, such algorithms are not of a particular interest herein.

Contributions

The main contribution of the present chapter consists in designing an algorithm for estimating parameters in the CCSSMs. The proposed approach shares the similarities with the aforementioned methods in the sense that it also computes the posterior statistics. The design of the method includes two ideas. First, we take advantage of the tractable substructure to integrate out the parameters and thus utilize the Rao-Blackwellization [18]. Second, based on the Kullback-Leibler divergence (KLD, [53]) principle, we formulate an update-project-update cycle to compute the posterior statistics. It is shown that the parameter estimation is then less degenerate.

1.2 Problem Formulation

Let us consider a discrete-time SSM in the form

$$p_\theta(y_t, x_t|x_{t-1}) = g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1}), \tag{1.1}$$

where $x_t \in \mathsf{X} \subseteq \mathbb{R}^{n_x}$ and $y_t \in \mathsf{Y} \subseteq \mathbb{R}^{n_y}$ label the state and observation variables, respectively. The model is characterized by the probability densities $g_\theta(\cdot)$ and $f_\theta(\cdot)$, with $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$ denoting some unknown static parameters. At the initial time step, the state and parameter variables are distributed according to $x_1 \sim p_\theta(x_1)$ and $\theta \sim p_0(\theta)$. We restrict ourselves to SSMs that allow us to express (1.1) by the exponential family (EF, [6]) density

$$\begin{aligned} p_\theta(y_t, x_t|x_{t-1}) = \exp\{&\langle \eta(\theta), s_t(x_{t-1}, x_t, y_t)\rangle \\ &- \zeta(\theta) + \log h(x_{t-1}, x_t, y_t)\}, \end{aligned} \tag{1.2}$$

where η and ζ are respectively the matrix and scalar-valued functions defined on Θ, s_t and h constitute respectively the matrix and scalar-valued functions defined on $\mathsf{X}^2 \times \mathsf{Y}$, and $\langle\cdot,\cdot\rangle$ represents the inner product. The SSM delineated by (1.2) is herein referred to as the CCSSM. The name follows from the fact that (1.2) is analytically tractable with respect to the parameters but intractable with respect to the presumably nonlinear functions s_t and h. More specifically, the model (1.2) facilitates analytical computation of the posterior density of the parameters, if we choose the conjugate prior density according to

$$\begin{aligned} p(\theta|\nu_{t-1}, V_{t-1}) = \exp\{\langle\eta(\theta), V_{t-1}\rangle - \nu_{t-1}\zeta(\theta) \\ - \log\mathcal{I}(\nu_{t-1}, V_{t-1})\}, \end{aligned} \tag{1.3}$$

where V_{t-1} denotes the extended information matrix, ν_{t-1} labels the number of degrees of freedom, and $\mathcal{I}$ defines the normalizing constant. The posterior density $p(\theta|\nu_t, V_t)$ then reproduces the form of (1.3), and its statistics can be updated under the closed-form formulae

$$V_t = V_{t-1} + s_t(x_{t-1}, x_t, y_t), \tag{1.4a}$$

$$\nu_t = \nu_{t-1} + 1. \tag{1.4b}$$

The model (1.2) is also known as the conditionally conjugate latent process model [90, 80]. The generic form (1.2) acknowledges standard probability densities such as Poisson, Gaussian, exponential, etc.

The objective of this chapter is to design an online method for computing the posterior density $p(x_t, \theta|y_{1:t})$ while assuming (1.2), where $y_{1:t} \coloneqq (y_1, \ldots, y_t)$. Nevertheless, the nonlinear functions s_t and h prevent us from computing the posterior analytically. To resolve this problem, we need to resort to approximate techniques. For the ability to deal with almost any nonlinear non-Gaussian SSM, we choose PFs to handle the approximate inference.

1.3 The Proposed Algorithm

The proposed method is summarized in Algorithm 1, where we use the convention $s_1(x_0, x_1, y_1) \coloneqq s_1(x_1, y_1)$ and consider that all i-dependent operations are performed for $i = 1, \ldots, N$. The derivation and more detailed description of Algorithm 1 can be found in the full version of this thesis.

1.4 Experiments and Results

The present section demonstrates the performance of the PBRBPF proposed in Algorithm 1 compared to the Rao-Blackwellized particle filter with linear computational complexity (RBPFN) [84] and the Rao-Blackwellized particle filter with

Algorithm 1 Projection-Based RBPF (PBRBPF)

A. **Initial step:** $(t = 1)$

1. Set $\widehat{\nu}_0$ and $\widehat{V}_0$.
2. Sample $x_1^i \sim q_1(\cdot)$.
3. Compute $w_1^i \propto W_1(x_1^i)$ with

$$W_1(x_1^i) := \frac{p(y_1, x_1^i | \widehat{\nu}_0, \widehat{V}_0)}{q_1(x_1^i)},$$

where

$$p(y_1, x_1^i | \widehat{\nu}_0, \widehat{V}_0) = \int_\Theta p_\theta(y_1, x_1^i) p(\theta | \widehat{\nu}_0, \widehat{V}_0) d\theta.$$

B. **Recursive step:** $(t = 2, \ldots, T)$

1. Sample a_t^i with $\mathbb{P}(a_t^i = j) = w_{t-1}^j$.
2. Sample $x_t^i \sim q_t(\cdot | x_{1:t-1}^{a_t^i})$.
3. Compute $w_t^i \propto W_t(x_{1:t}^i)$ with

$$W_t(x_{1:t}^i) := \frac{p(y_t, x_t^i | \widehat{\nu}_{t-1}, \widehat{V}_{t-1})}{q_t(x_t^i | x_{1:t-1}^{a_t^i})},$$

where

$$p(y_t, x_t^i | \widehat{\nu}_{t-1}, \widehat{V}_{t-1}) = \int_\Theta p_\theta(y_t, x_t^i | x_{t-1}^{a_t^i}) p(\theta | \widehat{\nu}_{t-1}, \widehat{V}_{t-1}) d\theta.$$

C. **Common step:** $(t \geq 1)$

1. Compute $\nu_t^i = \widehat{\nu}_{t-1} + 1$ and $V_t^i = \widehat{V}_{t-1} + s_t(x_{t-1}^{a_t^i}, x_t^i, y_t)$.
2. Compute $\widehat{\nu}_t$ and $\widehat{V}_t$ as the solution of

$$\mathsf{E}[\eta(\theta) | \widehat{\nu}_t, \widehat{V}_t] = \sum_{i=1}^N w_t^i \mathsf{E}[\eta(\theta) | \nu_t^i, V_t^i],$$
$$\mathsf{E}[\zeta(\theta) | \widehat{\nu}_t, \widehat{V}_t] = \sum_{i=1}^N w_t^i \mathsf{E}[\zeta(\theta) | \nu_t^i, V_t^i].$$

---

quadratic computational complexity (RBPFN^2), also known as the Rao-Blackwellized marginal particle filter [60]. We consider the standard benchmark SSM given by

$$x_t = \frac{x_{t-1}}{2} + 25 \frac{x_{t-1}}{1 + x_{t-1}} + 8\cos(1.2t) + w_t,$$
$$y_t = \frac{x_t^2}{20} + v_t,$$

where the variables $w_t \overset{IID}{\sim} \mathcal{N}(\cdot; \mu_w, \sigma_w^2)$ and $v_t \overset{IID}{\sim} \mathcal{N}(\cdot; \mu_v, \sigma_v^2)$ are assumed to be mutually independent. Here, IID stands for independent and identically distributed. The objective is to estimate μ_w, σ_w^2, μ_v, and σ_v^2, whose true values are 1, 2, 1, and 2, respectively. The initial state variable is distributed according to $x_1 \sim \mathcal{N}(\cdot; 0, 1)$. The amount of observations is $T = 2 \cdot 10^4$, and the number of particles is $N = 500$. The simulation is repeated 20 times with different observation sequences.

The time evolution of the parameter estimates over the independent simulation runs is displayed in Fig. 1.1. The results indicate that the proposed PBRBPF al-

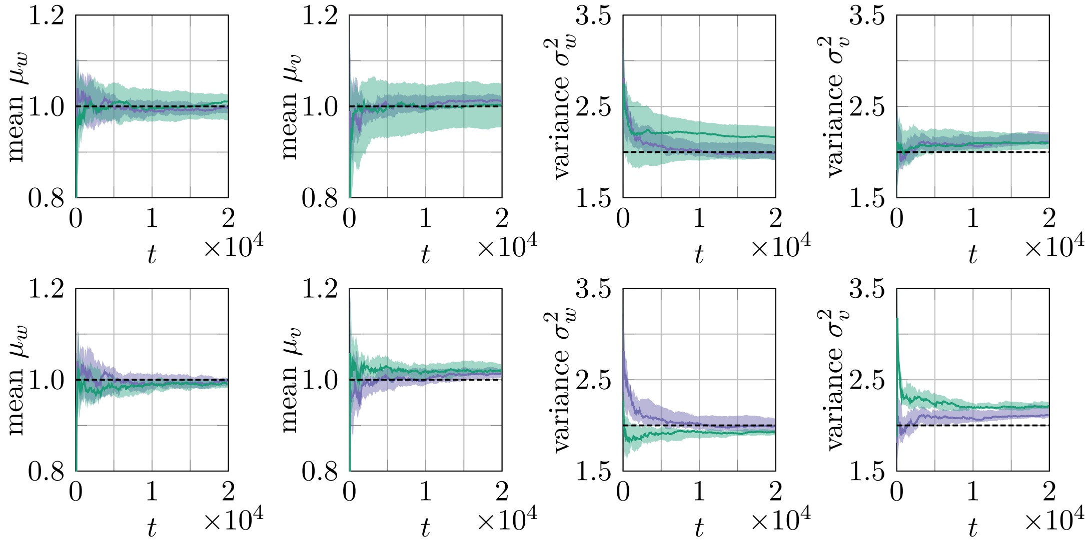

Fig. 1.1: The parameter estimates versus the number of observations. Top: PBRBPF (——) and RBPFN (——) [84]. Bottom: PBRBPF (——) and RBPFN^2 (——) [60]. The results are averaged over 20 independent simulation runs, with the solid line being the median and the shaded area delineating the interquartile range. The true parameter values are indicated with the dashed line (------).

gorithm outperforms the RBPFN method due to its lower bias and variance of the parameter estimates over the multiple simulation runs. From this observation, we can state that the proposed approach is less affected by the particle path degeneracy problem. The average time required to process all the observations with the PBRBPF and RBPFN algorithms was approximately 4.44 and 4.53 seconds, respectively. The proposed PBRBPF approach delivers slightly higher variance than the RBPFN^2 procedure. Nevertheless, the bias provided by the PBRBPF algorithm is lower than the one of the RBPFN^2 technique. Given the fact that the RBPFN^2 approach is computationally highly demanding, we can expect that a small increase in the number of particles of the PBRBPF method can easily compensate for this slightly higher variance.

2 A PARTICLE FILTER TO ESTIMATE TIME-VARYING PARAMETERS IN CONDITIONALLY CONJUGATE STATE-SPACE MODELS

The identification of slowly-varying parameters in dynamical systems constitutes a practically important task in a wide range of applications. The present chapter addresses this problem based on the Bayesian learning and sequential Monte Carlo (SMC) methodology. The proposed approach utilizes an algebraic structure of a specific class of nonlinear and non-Gaussian state-space models in order to enable Rao-Blackwellization of the parameters, thus involving a finite-dimensional sufficient statistic for each particle trajectory into the resulting algorithm. However, relying on basic SMC methods, such techniques are known to suffer from the particle path degeneracy problem. We propose to use alternative stabilized forgetting, which not only allows us to deal with the slowly-varying parameters but also to counteract the degeneracy problem. An experimental study proves the efficiency of the introduced Rao-Blackwellized particle filter compared to some related approaches.

2.1 Introduction

Context

The task of *online* SMC parameter estimation in non-linear state-space models has attracted substantial attention in the last years. Considerable effort has been devoted to maximum likelihood methods, where the aim is to maximize the likelihood $p_\theta(y_{1:t})$ of observed data sequence $y_{1:t} \coloneqq (y_1, \ldots, y_t)$ with respect to some fixed parameterization θ. An algorithmic solution in such cases commonly relies on the computation of expected values of smoothed additive functionals [14], which requires the complete data likelihood $p_\theta(x_{1:t}, y_{1:t})$ to belong to the exponential family [6], where $x_{1:t}$ denotes an unobserved state sequence. The main stream of research in this respect includes the gradient ascent [78] and expectation maximization (EM) methods [13]. However, these SMC-based approaches suffer from the particle path degeneracy problem [3, 44]. Recently, it was recognized in [20] that the forward smoothing algorithm can overcome this issue at the cost of $\mathcal{O}(N^2)$ operations, where N stands for the number of particles. The results of [72] show that the forward smoothing can actually be performed with $\mathcal{O}(N)$ operations by adapting the accept-reject backward sampling of [23].

Bayesian methods interpret unknown parameters as random variables and provide their full description in terms of the posterior density $p(\theta|y_{1:t})$. From this perspective, the earliest SMC approaches apply a particle filter to an augmented state variable $\bar{x}_t = (x_t, \theta)$ while considering a constant model of parameter variations $\theta_t = \theta_{t-1}$. Since the model of constant parameter variations lacks any forgetting properties [14], the diversity of the particle population representing θ decreases with successful resampling steps. The problem is commonly treated by introducing a jittering noise into the model of parameter evolutions [38]. However, a straightforward application of jittering can make the posterior density $p(\theta|y_{1:t})$ unnecessarily diffused. This was addressed in [51] by systematically decreasing the noise variance and later improved by alleviating the artificial variance inflation in [61]. But the simple addition of a jittering noise with a decreasing variance is not always efficient, as it may be difficult to guess a compromise between the number of particles being used and the rate at which the variance should decrease. The advantage of state augmentation techniques is that they can be applied to models without a specific structure. Considering a model with parameters respecting some structure in such a manner that the density $p(\theta|x_{1:t}, y_{1:t})$ is algebraically tractable, the paper [84] proposes to integrate out the parameters and to run a particle filter only for the marginal density $p(x_{1:t}|y_{1:t})$. For each particle trajectory, sampled from this marginal, the density $p(\theta|x_{1:t}, y_{1:t})$ is evaluated in terms of updating the sufficient statistics, which then serves for the parameter estimation. However, this online approach, too, suffers from the particle path degeneracy problem, resulting in a poor approximation of the posterior $p(\theta|x_{1:t}, y_{1:t})$. The related paper [74] imposes exponential forgetting [52] into this algorithm in order to facilitate the estimation of time-varying parameters and counteract the degenerate behavior.

Contributions

This chapter proposes a sequential Monte Carlo-based algorithm which exploits the algebraically tractable substructure of a special class of nonlinear state-space models, here referred to as conditionally conjugate state-space models. A characteristic feature of these models consists in that they facilitate the computation of $p(\theta|x_{1:t}, y_{1:t})$ under a close-form solution. The algorithm is—in its basic structure—similar to the one proposed in [84] but offers an ability to trace time-varying parameters. However, compared to the similar work [74], we accomplish this by utilizing a different forgetting strategy which is known as the alternative stabilized forgetting [50]. We demonstrate that the proposed algorithm outperforms this previous approach in terms of estimation accuracy and computational time.

2.2 Problem Formulation

In this chapter, we are concerned with discrete-time state-space models (SSMs) given by the joint probability density

$$p_{\theta_t}(y_t, x_t|x_{t-1}) = g_{\theta_t}(y_t|x_t) f_{\theta_t}(x_t|x_{t-1}), \tag{2.1}$$

where $x_t \in \mathsf{X} \subseteq \mathbb{R}^{n_x}$ and $y_t \in \mathsf{Y} \subseteq \mathbb{R}^{n_y}$ denote the state and observation variables, respectively. Furthermore, g_{θ_t} and f_{θ_t} constitute observation and state-transition models, with $\theta_t \in \Theta \subseteq \mathbb{R}^{n_\theta}$ being some unknown time-varying parameters. The initial step assumes that the state and parameter variables are distributed as $x_1 \sim p_{\theta_1}(x_1)$ and $\theta_1 \sim p(\theta_1)$. We are particularly interested in a specific class of SSMs which allows us to express (2.1) by the exponential family [6] density

$$\begin{aligned} p_{\theta_t}(y_t, x_t|x_{t-1}) = \exp\{&\langle \eta(\theta_t), s_t(x_{t-1}, x_t, y_t)\rangle \\ &- \zeta(\theta_t) + \log h(x_{t-1}, x_t, y_t)\}, \end{aligned} \tag{2.2}$$

where (η, ζ) and (s_t, h) are functions of appropriate dimensions, defined on Θ and $\mathsf{X}^2 \times \mathsf{Y}$, respectively, and $\langle \cdot, \cdot \rangle$ is the inner product. Due to the fact that (2.2) is analytically intractable with respect to the nonlinear functions (s_t, h) but tractable with respect to the parameter functions (η, ζ), we refer to (2.2) as the conditionally conjugate state-space model (CCSSM), alternatively known as the conditionally conjugate latent process model [90, 80]. The key characteristic of (2.2) consists in that, if we choose the conjugate prior density

$$\begin{aligned} p(\theta_t|\nu_{t|t-1}, V_{t|t-1}) = \exp\{&\langle \eta(\theta_t), V_{t|t-1}\rangle - \nu_{t|t-1}\zeta(\theta_t) \\ &- \log \mathcal{I}(\nu_{t|t-1}, V_{t|t-1})\}; \end{aligned} \tag{2.3}$$

then, we can compute the posterior density, $p(\theta_t|x_{1:t}, y_{1:t}) \coloneqq p(\theta_t|\nu_{t|t}, V_{t|t})$, analytically. In (2.3), $V_{t|t-1}$ is the extended information matrix, $\nu_{t|t-1}$ is the number of degrees of freedom, and $\mathcal{I}$ denotes the normalizing constant. Under this choice, the posterior density $p(\theta_t|\nu_{t|t}, V_{t|t})$ reproduces the form of (2.3), with the statistics being updated according to the closed-form formulae

$$\begin{aligned} V_{t|t} &= V_{t|t-1} + s_t(x_{t-1}, x_t, y_t), \\ \nu_{t|t} &= \nu_{t|t-1} + 1. \end{aligned}$$

Fundamental probability densities, including Poisson, Gaussian, and exponential, fit into the generic form delineated by (2.2).

The objective of this chapter consists in designing an online algorithm for computing the joint posterior density $p(x_t, \theta_t|y_{1:t})$ while assuming the model (2.2). There are, however, two main obstacles in achieving this goal: (i) the nonlinear functions

Algorithm 2 RBPF with Alternative Stabilized Forgetting (RBPFASF)

A. **Initial step:** $(t = 1)$

1. Set $(\widehat{\nu}^i_{1|0}, \widehat{V}^i_{1|0}, \nu_A, V_A)$, and λ.
2. Sample $x^i_1 \sim q_1(\cdot)$.
3. Compute $w^i_1 \propto W_1(x^i_1)$ with

$$W_1(x^i_1) \coloneqq \frac{p(y_1, x^i_1|\widehat{\nu}^i_{1|0}, \widehat{V}^i_{1|0})}{q_t(x^i_1)},$$

where

$$p(y_1, x^i_1|\widehat{\nu}^i_{1|0}, \widehat{V}^i_{1|0}) = \int_\Theta p_{\theta_1}(y_1, x^i_1) p(\theta_1|\widehat{\nu}^i_{1|0}, \widehat{V}^i_{1|0}) d\theta_1.$$

B. **Recursive step:** $(t = 2, \ldots, T)$

1. If $N_{\text{eff}} \leq N_{\text{th}}$, sample a^i_t with $\mathbb{P}(a^i_t = j) = w^j_{t-1}$ and set $\bar{w}^i_{t-1} = 1/N$.
   Else, set $a^i_t = i$ and $\bar{w}^i_{t-1} = w^i_{t-1}$.
2. Sample $x^i_t \sim q_t(\cdot|x^{a^i_t}_{1:t-1})$.
3. Compute $w^i_t \propto W_t(x^i_{1:t})\bar{w}^i_{t-1}$ using

$$W_t(x^i_{1:t}) \coloneqq \frac{p(y_t, x^i_t|\widehat{\nu}^{a^i_t}_{t|t-1}, \widehat{V}^{a^i_t}_{t|t-1})}{q_t(x^i_t|x^{a^i_t}_{1:t-1})},$$

where

$$p(y_t, x^i_t|\widehat{\nu}^{a^i_t}_{t|t-1}, \widehat{V}^{a^i_t}_{t|t-1}) = \int_\Theta p_{\theta_t}(y_t, x^i_t|x^{a^i_t}_{t-1}) p(\theta_t|\widehat{\nu}^{a^i_t}_{t|t-1}, \widehat{V}^{a^i_t}_{t|t-1}) d\theta_t.$$

C. **Common step:** $(t \geq 1)$

1. Compute $\nu^i_{t|t} = \widehat{\nu}^{a^i_t}_{t|t-1} + 1$ and $V^i_t = \widehat{V}^{a^i_t}_{t|t-1} + s_t(x^{a^i_t}_{t-1}, x^i_t, y_t)$.
2. Compute $\widehat{\nu}^i_{t+1|t}$ and $\widehat{V}^i_{t+1|t}$ as the solution of

$$\begin{aligned}
\mathsf{E}[\eta(\theta_{t+1})|\widehat{\nu}^i_{t+1|t}, \widehat{V}^i_{t+1|t}] &= \lambda \mathsf{E}[\eta(\theta_{t+1})|\nu^i_{t|t}, V^i_{t|t}] + (1-\lambda)\mathsf{E}[\eta(\theta_{t+1})|\nu_A, V_A], \\
\mathsf{E}[\zeta(\theta_{t+1})|\widehat{\nu}^i_{t+1|t}, \widehat{V}^i_{t+1|t}] &= \lambda \mathsf{E}[\zeta(\theta_{t+1})|\nu^i_{t|t}, V^i_{t|t}] + (1-\lambda)\mathsf{E}[\zeta(\theta_{t+1})|\nu_A, V_A].
\end{aligned}$$

(s_t, h) prevent us from computing the joint posterior density analytically, and (ii) the parameter time-evolution model $p(\theta_t|\theta_{t-1})$ is unknown. To deal with the first problem, we apply the particle filters, as they constitute a theoretically [93] and practically [26] well-established tool for approximating highly nonlinear probability densities. To resolve the second one, we incorporate—for the first time—the concept of alternative stabilized forgetting [50] into the context of particle filter-based estimation of slowly-varying parameters.

2.3 The Proposed Algorithm

The resulting method is summarized in Algorithm 2, where all i-dependent operations are performed for $i = 1, \ldots, N$. We use the convention that $s_1(x_0, x_1, y_1) \coloneqq s_1(x_1, y_1)$. The derivation and more detailed description of Algorithm 2 can be found in the full version of this thesis.

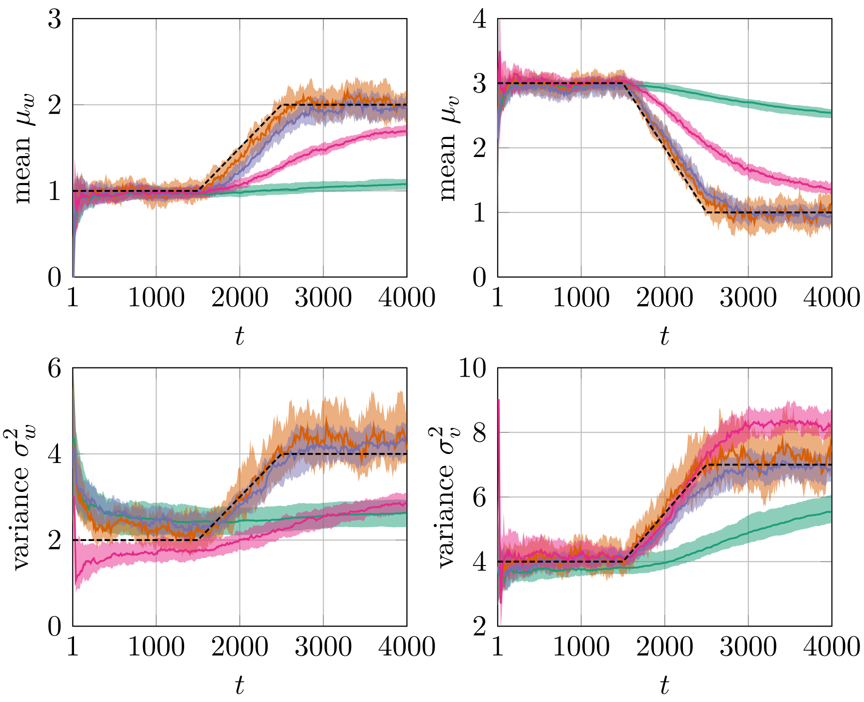

Fig. 2.1: The parameter estimates against the number of observations. The compared methods are PFEM (——) [13], RBPF (——) [84], RBPFEF (——) [74], and RBPFASF (——) Algorithm 2. The number of particles is $N = 512$. The results are averaged over 50 independent simulation runs, with the solid line being the median and the shaded area delineating the interquartile range. The true parameter values are indicated with the dashed line (------).

2.4 Experiments and Results

This section illustrates the behavior of the RBPF with alternative stabilized forgetting (RBPFASF), proposed in Algorithm 2, compared to the particle filter combined with the expectation maximization algorithm (PFEM) [13], the Rao-Blackwellized particle filter for static parameter estimation (RBPF) [84], and the RBPF with exponential forgetting (RBPFEF) [74]. We generate $T = 4000$ observations from the univariate non-stationary growth model

$$\begin{aligned} x_t &= \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + w_t, \\ y_t &= \frac{x_t^2}{20} + v_t, \end{aligned}$$

where $w_t \overset{IID}{\sim} \mathcal{N}(\cdot; \mu_w, \sigma_w^2)$ and $v_t \overset{IID}{\sim} \mathcal{N}(\cdot; \mu_v, \sigma_v^2)$ are mutually independent Gaussian noise variables. The initial value of the state variable is distributed as $x_1 \sim \mathcal{N}(0, 1)$. To be comparative, we follow the pattern of parameter changes outlined in [74]; thus, we have $\mu_{w,1} = 1$, $\Sigma_{w,1} = 2$, $\mu_{v,1} = 3$, $\Sigma_{v,1} = 4$, and $\mu_{w,4000} = 2$, $\Sigma_{w,4000} = 4$,

$\mu_{v,4000} = 1$, $\Sigma_{v,4000} = 7$ for the initial and final steps, respectively. The changes are executed between the times 1500 and 2500, see Fig. 2.1.

The resulting parameter estimates versus the number of observations are depicted in Fig. 2.1. The PFEM algorithm exhibits relatively good performance in terms of learning the static parameters; unfortunately, after the parameters start to change, we can observe that the adaptability of this methods is rather poor. A similar finding holds also for the RBPF, albeit the parameter estimates are obviously more biased. The RBPFEF offers better capability of tracking the changes. This is, however, achieved at the cost of higher variance of the estimated values. In addition, the estimates of σ_v^2 are considerably more biased. The proposed RBPFASF performs more favorably, mostly providing estimates with lower bias and variance compared to the other algorithms.

3 RAO-BLACKWELLIZED PARTICLE GIBBS KERNELS FOR SMOOTHING IN JUMP MARKOV NONLINEAR MODELS

Jump Markov nonlinear models (JMNMs) characterize a dynamical system by a finite number of presumably nonlinear and possibly non-Gaussian state-space configurations that switch according to a discrete-valued hidden Markov process. In this context, the smoothing problem—the task of estimating fixed points or sequences of hidden variables given all available data—is of key relevance to many objectives of statistical inference, including the estimation of static parameters. The present chapter proposes a particle Gibbs with ancestor sampling (PGAS)-based smoother for JMNMs. The design methodology relies on integrating out the discrete process in order to increase the efficiency through Rao-Blackwellization. The experimental evaluation illustrates that the proposed method achieves higher estimation accuracy in less computational time compared to the original PGAS procedure.

3.1 Introduction

Context

Particle Markov chain Monte Carlo (PMCMC) methods [2] have recently emerged as an efficient tool to perform statistical inference in general state-space models (SSMs, [14]). These algorithms apply sequential Monte Carlo (SMC, [25]) to tackle the issue of constructing high-dimensional proposal kernels in MCMC [1]. This makes them particularly well suited for addressing the smoothing problem in jump Markov nonlinear models (JMNMs). The particle Gibbs with ancestor sampling (PGAS) kernel [58], which can be seen as a PMCMC smoother, has proved to be a serious competitor to the prominent SMC-based smoothing strategies such as the backward simulator [37] and generalized SMC two-filter smoother [12]. For a thorough review of existing SMC-based smoothers, see [59] and references therein.

The development in this chapter is motivated by the recent progress in constructing PG kernels specifically tailored for jump Markov linear models (JMLMs) [94, 85]. The methods therein exploit the linear Gaussian substructure of the model to increase their efficiency through Rao-Blackwellization. This is achieved by using the Kalman filter (KF) to design the conditional variants of either the discrete particle filter [32] or Rao-Blackwellized particle filter (RBPF, [24]). A common aspect of these PG methods lies in that the backward information filter (BIF, [67]) is used to further increase the effect of Rao-Blackwellization and to improve the mixing properties [1] via ancestor sampling or backward simulation.

Contributions

The problem with JMNMs is that their nonlinear character prevents us from applying Rao-Blackwellization in the same sense as with JMLMs; nevertheless, there is still a tractable substructure to exploit. The present chapter is concerned with the design of a Rao-Blackwellized PGAS (RBPGAS) kernel that takes advantage of the hierarchical structure formed by the discrete latent process. The method builds on the RBPF proposed in [73], which is similar to that introduced in [24] except it replaces the above-discussed KF with a finite state-space filter; conversely, the particle filter (PF) focuses on the remaining (continuous-valued) part of the latent process. However, the design of a finite state-space BIF turns out to be more intricate in this context as it requires us to introduce a sequence of artificial probability distributions to change the scale of the associated backward recursion.

3.2 Problem Formulation

The generic form of the discrete-time JMNM considered in the present chapter is defined by

$$c_t|c_{t-1} \sim p(c_t|c_{t-1}), \tag{3.1a}$$

$$z_t|c_t, z_{t-1} \sim f(z_t|c_t, z_{t-1}), \tag{3.1b}$$

$$y_t|c_t, z_t \sim g(y_t|c_t, z_t), \tag{3.1c}$$

where the states and measurements are denoted by $z_t \in \mathsf{Z} \subseteq \mathbb{R}^{n_z}$ and $y_t \in \mathsf{Y} \subseteq \mathbb{R}^{n_y}$, respectively. The activity of the current regime of the model is indicated by the discrete *mode* variable $c_t \in \mathsf{C} \coloneqq \{1, \dots, K\}$. We assume to have access only to the measurements y_t, while the state z_t and mode c_t variables are considered hidden. Furthermore, for all $c_t \in \mathsf{C}$, the model is characterized by its state transition and observation probability densities $f(\cdot)$ and $g(\cdot)$, respectively. The switching between the modes is governed by the conditional probability distribution $p(\cdot)$. At the initial time step, the hidden variables are distributed according to $z_1 \sim \mu(z_1|c_1)$ and $c_1 \sim p(c_1)$. For a graphical representation of a JMNM.

Let $x_{1:T} \coloneqq (x_1, \dots, x_T)$ denote a generic sequence of variables defined on some product space X^T, for an integer $T > 0$ denoting the final time point. The aim of this study is to design an efficient PMCMC smoother targeting the joint smoothing density given by

$$p(c_{1:T}, z_{1:T}|y_{1:T}) = \frac{p(c_{1:T}, z_{1:T}, y_{1:T})}{p(y_{1:T})}. \tag{3.2}$$

Algorithm 3 Finite State-Space BIF

Inputs: $z'_{1:T}$ and $\{\xi_t(c_t)\}_{t=1}^T$.

Outputs: $\{\tilde{p}(c_t|y_{t:T}, z'_{t:T})\}_{t=1}^T$.

A. **Initial step:** $(t = T)$

1. Compute $\tilde{p}(c_T|y_T, z'_T) \propto p(y_T|c_T, z'_T)\xi_T(c_T)$.

B. **Recursive step:** $(t = T-1, \dots, 1)$

1. Compute

$$\tilde{p}(c_t|y_{t+1:T}, z_{t:T}) \propto \sum_{c_{t+1}\in \mathsf{C}} \tilde{p}(c_{t+1}|y_{t+1:T}, z_{t+1:T}) \frac{p(z_{t+1}, c_{t+1}|c_t, z_t)\xi_t(c_t)}{\xi_{t+1}(c_{t+1})},$$

2. Compute

$$\tilde{p}(c_t|y_{t:T}, z_{t:T}) \propto p(y_t|c_t, z_t)\tilde{p}(c_t|y_{t+1:T}, z_{t:T}).$$

However, the density (3.2) is intractable even in situations where (3.1b) and (3.1c) are linear and Gaussian. The reason consists in that the marginal likelihood $p(y_{1:T})$ contains summation over K^T values, which is always impossible to compute exactly, except for small data sets. Despite this, we consider (3.1b) and (3.1c) nonlinear and non-Gaussian, making the situation even more difficult as the integral over Z^T in the marginal likelihood $p(y_{1:T})$ cannot be evaluated either.

3.3 The Proposed Algorithm

The proposed RBPGAS kernel is summarized by Algorithms 3 and 4. The derivation and more detailed description of Algorithms 3 and 4 can be found in the full version of this thesis.

3.4 Experiments and Results

This section demonstrates the performance of the proposed RBPGAS kernel (Algorithm 6) in comparison to the PG [2], PGAS [58], RBPG (Algorithm 6 with setting $a_t^N \coloneqq N$ in step B4), and RBPGASnr (Algorithm 6 with the non-rescaled recursion) kernels. Let us consider the nonlinear benchmark model given by

$$\begin{aligned} z_t &= \frac{z_{t-1}}{2} + 25\frac{z_{t-1}}{1+z_{t-1}^2} + 8\cos(1.2t) + v_t, \\ y_t &= \frac{z_t^2}{20} + w_t, \end{aligned}$$

where, for $c_t \in \mathsf{C} \coloneqq \{1, 2\}$, $w_t \sim \mathcal{N}(\mu_{c_t}, \sigma_{c_t}^2)$ denotes a mode-dependent Gaussian noise variable with the mean μ_{c_t} and variance $\sigma_{c_t}^2$. Furthermore, $v_t \sim \mathcal{N}(0, 1)$ is an independent and identically distributed Gaussian noise variable with zero mean and unit variance. The kernel (3.1a) is parameterized by the transition probability

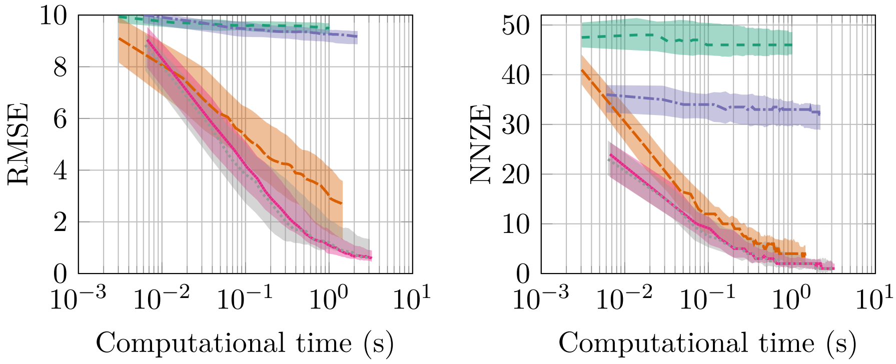

Fig. 3.1: Left: the root-mean-square error (RMSE) between the exact and estimated state trajectories versus the computational time (in seconds). Right: the number of non-zero elements (NNZE) in the error sequence between the exact and estimated mode trajectories versus the computational time. The solid line shows the median, and the shaded area is the interquartile range; both are computed over forty independent runs. The compared methods are PG (- - -), PGAS (- - -), RBPG (-·-·-), RBPGAS (——), and RBPGASnr (·········).

matrix Π according to $p(c_t = j|c_{t-1} = i) \coloneqq \Pi_{ij}$ with $i, j \in \mathsf{C}$. The diagonal entries of this matrix are set as $\Pi_{11} = 0.6$ and $\Pi_{22} = 0.8$. The mode-dependent means and variances are defined by $\mu_1 = 0$, $\mu_2 = 7$ and $\sigma_1^2 = 4$, $\sigma_2^2 = 1$. The state prior density is mode-independent and Gaussian, $\mu(z_1|c_1) \coloneqq \mathcal{N}(z_1; 0, 1)$. Further, the prior distribution of the mode variable $p(c_1)$ is parameterized by the vector λ with the relation $p(c_1 = i) \coloneqq \lambda_i$, where the first entry is $\lambda_1 = 0.5$. Forty independent runs of the considered model were performed, each producing the measurement sequence of the length $T = 100$. The algorithms subjected to comparison were tested with $R = 500$ iterations. To evaluate resulting estimates, the proposed RBPGAS method with $N = 1024$ particles was used to compute 'exact' state and mode trajectories for each of the measurement sequences.

Fig. 3.1 provides a closer look at the situation where the compared algorithms are applied with $N = 2$ particles. We can see that the RMSE of the PGAS and RBPGAS methods is competitive for approximately the first 10^{-1} seconds, with the RBPGAS algorithm starting to be computationally more efficient after this time. This is obvious from the median and interquartile range, which decrease more quickly for the RBPGAS procedure. The right part of Fig. 3.1 reveals that the RBPGAS method achieves lower values of the NNZE in a shorter computational time. For example, we can see that the value 10 is there reached after approximately $2 \cdot 10^{-1}$ seconds with the PGAS method, while the same value is attained after approximately $7 \cdot 10^{-2}$ seconds with the RBPGAS algorithm. It is therefore obvious that the RBPGAS procedure is markedly quicker in approaching the ergodic regime.

Algorithm 4 RBPGAS Kernel for JMNMs (version A)

Inputs: $z'_{1:T} = z_{1:T}[k-1]$.

Outputs: $z_{1:T}[k]$ and $\{z^i_{1:T}, \{p(c_t|z^i_{1:t}, y_{1:t})\}_{t=1}^T, w^i_T\}_{i=1}^N$.

A. **Initial step:** $(t = 1)$

1. Compute the sequence $\{\xi_t(c_t)\}_{t=1}^T$.
2. Use $z'_{1:T}$ and $\{\xi_t(c_t)\}_{t=1}^T$ as the input for Algorithm 3 to produce $\{\tilde{p}(c_t|y_{t:T}, z'_{t:T})\}_{t=1}^T$.
3. Sample $z^i_1 \sim q_1(\cdot)$ for $i = 1, \ldots, N-1$ and set $z^N_1 := z'_1$.
4. Compute $p(c_1|z^i_1, y_1) \propto p(y_1, z^i_1|c_1)p(c_1)$, for $i = 1, \ldots, N$.
5. Compute $w^i_1 \propto W_1(z^i_1)$ according to

$$W_1(z^i_1) = \frac{p(y_1, z^i_1)}{q_1(z^i_1)}, \qquad \text{where} \qquad p(y_1, z^i_1) = \sum_{c_1 \in \mathsf{C}} p(y_1, z^i_1|c_1)p(c_1),$$

   for $i = 1, \ldots, N$.

B. **Recursive step:** $(t = 2, \ldots, T)$

1. Sample a^i_t with $\mathbb{P}(a^i_t = j) = w^j_{t-1}$ for $i = 1, \ldots, N-1$.
2. Sample $z^i_t \sim q_t(\cdot|z^{a^i_t}_{1:t-1})$ for $i = 1, \ldots, N-1$.
3. Compute $w^i_{t-1|T} \propto w^i_{t-1} p(y_{t:T}, z'_{t:T}|z^i_{1:t-1}, y_{1:t-1})$ with

$$p(y_{t:T}, z'_{t:T}|z^i_{1:t-1}, y_{1:t-1}) = \sum_{c_{t-1} \in \mathsf{C}} p(y_{t:T}, z'_{t:T}|z^i_{t-1}, c_{t-1})p(c_{t-1}|z^i_{1:t-1}, y_{1:t-1})$$

   and

$$p(y_{t:T}, z'_{t:T}|z^i_{t-1}, c_{t-1}) \propto \sum_{c_t \in \mathsf{C}} \frac{\tilde{p}(c_t|y_{t:T}, z'_{t:T})}{\xi_t(c_t)} p(z'_t, c_t|c_{t-1}, z^i_{t-1}),$$

   for $i = 1, \ldots, N$.
4. Sample a^N_t with $\mathbb{P}(a^N_t = i) = w^i_{t-1|T}$ and set $z^N_t := z'_t$.
5. Set $z^i_{1:t} := \{z^i_t, z^{a^i_t}_{1:t-1}\}$ for $i = 1, \ldots, N$.
6. Compute $p(c_t|z^i_{1:t}, y_{1:t}) \propto p(y_t, z^i_t|c_t, z^{a^i_t}_{t-1})p(c_t|z^{a^i_t}_{1:t-1}, y_{1:t-1})$ where

$$p(c_t|z^{a^i_t}_{1:t-1}, y_{1:t-1}) = \sum_{c_{t-1} \in \mathsf{C}} p(c_t|c_{t-1})p(c_{t-1}|z^{a^i_t}_{1:t-1}, y_{1:t-1}),$$

   for $i = 1, \ldots, N$.
7. Compute $w^i_t \propto W_t(z^i_{1:t})$ according to

$$W_t(z^i_{1:t}) = \frac{p(y_t, z^i_t|z^{a^i_t}_{1:t-1}, y_{1:t-1})}{q_t(z^i_t|z^{a^i_t}_{1:t-1})},$$

   where

$$p(y_t, z^i_t|z^{a^i_t}_{1:t-1}, y_{1:t-1}) = \sum_{c_t \in \mathsf{C}} p(y_t, z^i_t|c_t, z^{a^i_t}_{t-1})p(c_t|z^{a^i_t}_{1:t-1}, y_{1:t-1}),$$

   for $i = 1, \ldots, N$.

C. **Final step:**

1. Sample k with $\mathbb{P}(k = i) = w^i_T$ and set $z_{1:T}[k] := z^k_{1:T}$.

4 A PARTICLE SAEM ALGORITHM TO IDENTIFY JUMP MARKOV NONLINEAR MODELS

The identification of static parameters in jump Markov nonlinear models (JMNMs) poses a key challenge in explaining nonlinear and abruptly changing behavior of dynamical systems. This chapter introduces a stochastic approximation expectation maximization algorithm to facilitate offline maximum likelihood parameter estimation in JMNMs. The method relies on the construction of a particle Gibbs kernel that takes advantage of the inherent structure of the model to increase the efficiency through Rao-Blackwellization. Numerical examples illustrate that the proposed solution outperforms related approaches.

4.1 Introduction

Context

Jump Markov nonlinear models (JMNMs) can be seen as a particular class of nonlinear and non-Gaussian state-space models (SSMs, [14]) where the observation variable is related to the *latent* state variable that contains a continuous and discrete-valued part. While the continuous part describes the dynamics of a system, the discrete one indicates the switching of different dynamical *modes*.

The expectation maximization (EM) algorithm by [22] has become a standard tool to address the maximum likelihood (ML) parameter estimation in SSMs. The method is favored especially for its inherent feature of splitting the ML problem into two more conveniently tractable steps known as expectation and maximization. In the model class considered here, the expectation step is intractable and requires us to solve the nonlinear smoothing problem [59]. The particle Markov chain Monte Carlo (PMCMC) methods [2], which rely on sequential Monte Carlo (SMC, [25]) to facilitate the construction of high-dimensional proposal *kernels* in (MCMC, [1]), embody an efficient tool to address the issue. The paper [58] recently elaborated on the PMCMC idea and suggested to combine their particle Gibbs with ancestor sampling (PGAS) kernel and the stochastic approximation EM (SAEM) algorithm of [21] to obtain the particle SAEM (PSAEM) procedure. The related paper [85] then extended this design to propose a Rao-Blackwellized PSAEM (RBPSAEM) algorithm for jump Markov linear models by utilizing their linear Gaussian substructure.

A recent EM approach specifically tailored for JMNMs was proposed by [5] who extended the particle smoothing EM (PSEM) framework of [82]. The method proposed herein differs from this approach mainly in using stochastic approximation, Rao-Blackwellization, and PMCMC-based smoothing. Another EM solution was developed by [73] who introduced a Rao-Blackwellized forward smoother, which differs from the present method also in the smoothing methodology but shares similarities with the specific type of Rao-Blackwellization.

Contributions

The contribution of this chapter consists in developing an RBPSAEM method for JMNMs which exploits the substructure related to the discrete state. This is achieved by formulating a conditional version of the RBPF proposed by [73]. To facilitate the ancestor sampling, a finite state-space variant of the backward information filter (BIF, [67]) is proposed, requiring us to change the scale of the associated backward recursion. The experimental evidence indicates that the proposed method offers a higher estimation accuracy compared to competing approaches.

4.2 Problem Formulation

Consider the discrete-time JMNM given by

$$c_t \sim p(c_t|c_{t-1}), \tag{4.1a}$$
$$z_t \sim f(z_t|c_t, z_{t-1}; \theta_{c_t}), \tag{4.1b}$$
$$y_t \sim g(y_t|c_t, z_t; \theta_{c_t}), \tag{4.1c}$$

where the continuous states and observations are denoted by $z_t \in \mathsf{Z} \subseteq \mathbb{R}^{n_z}$ and $y_t \in \mathsf{Y} \subseteq \mathbb{R}^{n_y}$, respectively. The discrete state $c_t \in \mathsf{C} := \{1, \ldots, K\}$ indicates the currently active mode, with K being the total number of the modes. We assume that each mode is described by the probability densities $f(\cdot; \theta_{c_t})$ and $g(\cdot; \theta_{c_t})$, where θ_{c_t} is the associated parameter set. The probability distribution $p(\cdot)$ governs the switching between the modes and is parameterized by the $K \times K$ transition probability matrix Π with the entries

$$\Pi_{mn} := \mathbb{P}(c_t = n|c_{t-1} = m) = p(n|m). \tag{4.2}$$

The set of all unknown parameters, $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$, is defined by $\theta := \{\Pi, \{\theta_n\}_{n=1}^K\}$. At the initial time instance, the latent states are distributed according to $z_1 \sim \mu(z_1|c_1)$ and $c_1 \sim \nu(c_1)$; both μ and ν are assumed to be known.

We search for the parameter estimate maximizing the likelihood of the observed data sequence $y_{1:T} := (y_1, \ldots, y_T)$, with T denoting its length, that is,

$$\widehat{\theta}_{\mathrm{ML}} = \operatorname*{arg\,max}_{\theta \in \Theta} p_\theta(y_{1:T}). \tag{4.3}$$

Algorithm 5 Rao-Blackwellized Stochastic Approximation Expectation Maximization (SAEM) for JMNMs

A. **Initial step:** $(k = 0)$

1. Set $z_{1:T}[0] \in \mathsf{Z}^T$, $\theta[0] \in \Theta$, and $\widehat{\mathcal{Q}}_0(\theta) \coloneqq 0$.

B. **Recursive step:** $(k = 1, \ldots, R)$

1. Sample $z_{1:T}[k] \sim \mathcal{K}_{\theta[k-1]}(\cdot|z_{1:T}[k-1])$.
2. Compute $\widehat{\mathcal{Q}}_k(\theta)$ according to

$$\widehat{\mathcal{Q}}_k(\theta) = (1-\alpha_k)\widehat{\mathcal{Q}}_{k-1}(\theta) + \alpha_k \mathbb{E}_{\theta[k-1]}\big[\log p_\theta(c_{1:T}, z_{1:T}[k], y_{1:T})\big|z_{1:T}[k], y_{1:T}\big].$$

3. Compute $\theta[k] = \arg\max_{\theta\in\Theta} \widehat{\mathcal{Q}}_k(\theta)$.

In the present class of models, the computation of $p_\theta(y_{1:T})$ cannot be conducted exactly, as it contains the summation over K^T possible values, which is infeasible to perform even for a moderate T. Additionally, the integration over Z^T required for evaluating $p_\theta(y_{1:T})$ cannot be performed either, as the model is supposed to contain nonlinearities.

4.3 The Proposed Algorithm

The proposed RBPGAS kernel and the SAEM algorithm are summarized in Algorithms 5 and 6, respectively. The derivation and more detailed description of Algorithms 5 and 6 can be found in the full version of this thesis.

4.4 Experiments and Results

This section illustrates the performance of the proposed RBPSAEM algorithm compared to the PSAEM [57] and PSEM [82] procedures. We perform the experiment on the standard benchmark state-space model

$$\begin{aligned} x_t &= \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t) + v_t, \\ y_t &= \frac{x_t^2}{20} + w_t, \end{aligned}$$

where $v_t \sim \mathcal{N}(0,1)$ is a mode-independent, independent and identically distributed, Gaussian noise variable with zero mean and unit variance, and $w_t \sim \mathcal{N}(\mu_{c_t}, \sigma^2_{c_t})$ is a mode-dependent Gaussian noise variable with the mean μ_{c_t} and variance $\sigma^2_{c_t}$. We consider that the total number of modes is $K = 2$, i.e., $c_t \in \mathsf{C} \coloneqq \{1, 2\}$. The task is to estimate the parameters μ_1, μ_2, σ_1^2, σ_2^2, Π_{11}, and Π_{22}, with their true values given by 0, 8, 5, 1, 0.98, and 0.8, respectively. We repeat the experiment on 20 different observation sequences of the length $T = 1000$.

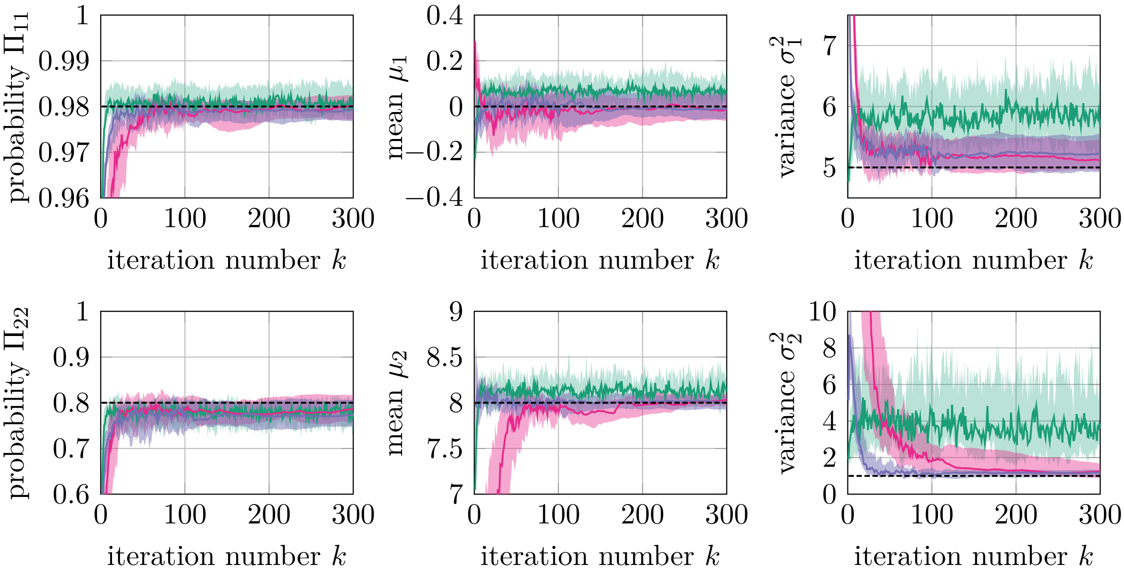

Fig. 4.1: The resulting parameter estimates versus the number of iterations for PSEM (——), PSAEM (——), and RBPSAEM (——). The results are averaged over twenty independent simulation runs, with the solid line being the median and the shaded area delineating the interquartile range. The true parameter values are indicated with the dashed line (------).

Fig. 4.1 shows that the proposed method surpasses PSAEM because of the lower (or very similar) bias and variance of the estimated parameters. Although PSEM is better in estimating the transition probabilities, the remaining estimates converge to incorrect values. The main reason then consists in that PSEM does not rely on the stochastic approximation and thus requires a higher number of particles to perform similarly to the remaining procedures. Moreover, as the probability Π_{11} is close to its upper bound, the method suffers from the degeneracy around the mode changes [27]. Nevertheless, both PSAEM and RBPSAEM seem to be more robust in this respect.

Algorithm 6 RBPGAS Kernel for JMNMs (version B)

Inputs: $z'_{1:T} = z_{1:T}[k-1]$.

Outputs: $z_{1:T}[k]$ and $\{z^i_{1:T}, \{p(c_t|z^i_{1:t}, y_{1:t})\}_{t=1}^T, w^i_T\}_{i=1}^N$.

A. **Initial step:** $(t = 1)$

1. Compute $\alpha_l(c_l|z_{l:T}) \propto p(y_l|c_l, z_l)\alpha_{l+1}(c_l|z_{l:T})$ where

$$\alpha_{l+1}(c_l|z_{l:T}) \propto \sum_{c_{l+1}\in\mathsf{C}} \alpha_{l+1}(c_{l+1}|z_{l+1:T})p(z_{l+1}, c_{l+1}|c_l, z_l),$$

for $l = 1, \dots, T$.

2. Sample $z^i_1 \sim q_1(\cdot)$ for $i = 1, \dots, N-1$ and set $z^N_1 := z'_1$.
3. Compute $p(c_1|z^i_1, y_1) \propto p(y_1, z^i_1|c_1)p(c_1)$, for $i = 1, \dots, N$.
4. Compute $w^i_1 \propto W_1(z^i_1)$ according to

$$W_1(z^i_1) = \frac{p(y_1, z^i_1)}{q_1(z^i_1)}, \qquad \text{where} \qquad p(y_1, z^i_1) = \sum_{c_1\in\mathsf{C}} p(y_1, z^i_1|c_1)p(c_1),$$

for $i = 1, \dots, N$.

B. **Recursive step:** $(t = 2, \dots, T)$

1. Sample a^i_t with $\mathbb{P}(a^i_t = j) = w^j_{t-1}$ for $i = 1, \dots, N-1$.
2. Sample $z^i_t \sim q_t(\cdot|z^{a^i_t}_{1:t-1})$ for $i = 1, \dots, N-1$.
3. Compute $w^i_{t-1|T} \propto w^i_{t-1}p(y_{t:T}, z'_{t:T}|z^i_{1:t-1}, y_{1:t-1})$ with

$$p(y_{t:T}, z'_{t:T}|z^i_{1:t-1}, y_{1:t-1}) = \sum_{c_{t-1}\in\mathsf{C}} p(y_{t:T}, z'_{t:T}|z^i_{t-1}, c_{t-1})p(c_{t-1}|z^i_{1:t-1}, y_{1:t-1}),$$

and

$$p(y_{t:T}, z'_{t:T}|z^i_{t-1}, c_{t-1}) \propto \sum_{c_t\in\mathsf{C}} \alpha_t(c_t|z'_{t:T})p(z'_t, c_t|c_{t-1}, z^i_{t-1}),$$

for $i = 1, \dots, N$.

4. Sample a^N_t with $\mathbb{P}(a^N_t = i) = w^i_{t-1|T}$ and set $z^N_t := z'_t$.
5. Set $z^i_{1:t} := \{z^i_t, z^{a^i_t}_{1:t-1}\}$ for $i = 1, \dots, N$.
6. Compute $p(c_t|z^i_{1:t}, y_{1:t}) \propto p(y_t, z^i_t|c_t, z^{a^i_t}_{t-1})p(c_t|z^{a^i_t}_{1:t-1}, y_{1:t-1})$ where

$$p(c_t|z^{a^i_t}_{1:t-1}, y_{1:t-1}) = \sum_{c_{t-1}\in\mathsf{C}} p(c_t|c_{t-1})p(c_{t-1}|z^{a^i_t}_{1:t-1}, y_{1:t-1}),$$

for $i = 1, \dots, N$.

7. Compute $w^i_t \propto W_t(z^i_{1:t})$ according to

$$W_t(z^i_{1:t}) = \frac{p(y_t, z^i_t|z^{a^i_t}_{1:t-1}, y_{1:t-1})}{q_t(z^i_t|z^{a^i_t}_{1:t-1})},$$

where

$$p(y_t, z^i_t|z^{a^i_t}_{1:t-1}, y_{1:t-1}) = \sum_{c_t\in\mathsf{C}} p(y_t, z^i_t|c_t, z^{a^i_t}_{t-1})p(c_t|z^{a^i_t}_{1:t-1}, y_{1:t-1}),$$

for $i = 1, \dots, N$.

C. **Final step:**

1. Sample k with $\mathbb{P}(k = i) = w^i_T$ and set $z_{1:T}[k] := z^k_{1:T}$.

5 DYNAMIC BAYESIAN KNOWLEDGE TRANSFER BETWEEN A PAIR OF KALMAN FILTERS

Transfer learning is a framework that includes—among other topics—the design of knowledge transfer mechanisms between Bayesian filters. Transfer learning strategies in this context typically rely on a complete stochastic dependence structure being specified between the participating learning procedures (filters). This chapter proposes a method that does not require such a restrictive assumption. The solution in this *incomplete modelling* case is based on the fully probabilistic design of an unknown probability distribution which conditions on knowledge in the form of an externally supplied distribution. We are specifically interested in the situation where the external distribution accumulates knowledge dynamically via Kalman filtering. Simulations illustrate that the proposed algorithm outperforms alternative methods for transferring this dynamic knowledge from the external Kalman filter.

5.1 Introduction

Context

Transfer learning [75] has become a key research direction in statistical machine learning [69]. The basic principle of transfer learning is to utilize the experience of an external learning agent (source task) to improve the learning of a primary agent (target task). Transfer learning has recently witnessed substantial attention in a multitude of theoretically and practically oriented scientific fields, such as reinforcement learning [87], deep learning [8], autonomous driving [42], computer vision [76], sensor networks [89], etc. This chapter focuses on a specific transfer learning context referred to as Bayesian transfer learning and its deployment in statistical signal processing. We are specifically interested in developing a procedure for probabilistic knowledge transfer in sensor networks where each knowledge-bearing node constitutes a Bayesian filter acting on its associated state-space model.

The conventional approach to Bayesian transfer learning involves replacing the prior distribution of standard Bayesian learning with a distribution conditioned on the transferred knowledge [88]. The methods based on this principle differ in the way the knowledge-conditioned prior is elicited [9]. An alternative principle is to define the joint posterior distribution of both source and target quantities of interest given source and target data, and then to compute the posterior distribution of the target quantity by marginalization [48]. Hierarchical Bayesian learning provides another formalization of Bayesian transfer learning [96], where the knowledge is transferred by means of a hyper-prior. However, it seems that a widely accepted consensus on Bayesian transfer learning is missing. This chapter seeks to fill this gap.

Contributions

The common aspect of the above approaches is that they assume existence of an explicit model of all unknown quantities of interest, enabling Bayes' rule to accommodate transfer learning, which we call here the *complete modelling* case. In the present chapter, we are concerned with a scenario where there is not enough knowledge to construct such a model explicitly. We refer to this particular situation as the *incomplete modelling* case. The previous work in this respect [34] involved a static Bayesian knowledge transfer for a pair of Kalman filters, where the external knowledge is transferred in the form of a marginal distribution defined at a single time-step. The present chapter extends this work by designing a mechanism for transferring distributions defined over multiple time-steps, thus achieving dynamic and on-line Bayesian knowledge transfer.

5.2 Problem Formulation

Let us consider a state-space model given by

$$f(x_i|x_{i-1}) \equiv \mathcal{N}_{x_i}(Ax_{i-1}, Q), \tag{5.1a}$$

$$f(z_i|x_i) \equiv \mathcal{N}_{z_i}(Cx_i, R), \tag{5.1b}$$

where $x_i \in \mathsf{X} \subseteq \mathbb{R}^{n_x}$ and $z_i \in \mathsf{Z} \subseteq \mathbb{R}^{n_z}$ are respectively the state and observation variables defined at the discrete-time instants $i = 1, \ldots, n$. The state-space model (5.1) is fully determined by the state transition (5.1a) and observation (5.1b) probability densities, with all their parameters being known. Here, $\mathcal{N}_v(\mu, \Sigma)$ denotes the Gaussian density of a (vector) random variable, v, with the mean, μ, and covariance matrix, Σ; and A and C are matrices of appropriate dimensions. At the initial time step ($i = 1$), the state variable is distributed according to $f(x_1) \equiv \mathcal{N}_{x_1}(\mu_1, \Sigma_1)$. The time-evolution of the state-space model (5.1) is characterized by the joint augmented model

$$f(\mathbf{x}_n, \mathbf{z}_n) = f(\mathbf{z}_n|\mathbf{x}_n) f(\mathbf{x}_n) \equiv \prod_{i=1}^{n} f(z_i|x_i) f(x_i|x_{i-1}), \tag{5.2}$$

where $f(\mathbf{z}_n|\mathbf{x}_n)$ and $f(\mathbf{x}_n)$ define the joint observation model and joint state pre-prior model, respectively. In (5.2), we respect the convention $x_0 \equiv \varnothing$ and use the boldface notation $\mathbf{v}_n \equiv (v_1, \ldots, v_n)$ to denote a sequence of variables $v_i \in \mathsf{V}$, for $i = 1, \ldots, n$.

Moreover, we use the symbols m and f to denote unspecified (variational form) and specified (fixed form) densities, respectively.

We are concerned with the problem of optimally transferring knowledge from an external Bayesian filter (source task) to a primary one (target task). The filters operate on their respective models, processing their local observations, and estimating their local states. The conditional independence structure between the variables in each model is as specified in (5.2). However, an explicit conditioning mechanism describing dependence between quantities of the primary filter, $(\mathbf{x}_n, \mathbf{z}_n)$, and external filter, $(\mathbf{x}_{n,e}, \mathbf{z}_{n,e})$, is assumed missing. The common modelling approach based on a joint density of the external and primary variables is therefore unavailable. This incomplete modelling scenario is addressed here as a problem of optimal design of an unknown probability density, processing the external (distributional) knowledge as a constraint. Specifically, we design a *dynamic* Bayesian knowledge transfer method, where knowledge is transferred in the form of a joint probability density, f_e, describing a sequence of external quantities, $\mathbf{z}_{n,e}$.

5.3 The Proposed Algorithm

The resulting filter with FPD-optimal dynamic transfer is summarized in Algorithm 7. The derivation and more detailed description of Algorithm 7 can be found in the full version of this thesis.

5.4 Experiments and Results

The purpose of this section is to compare the proposed method against alternative approaches. We evaluate the performance of the primary filter when keeping its observation variance R fixed but changing the observation variance of the external filter R_e, which quantifies the confidence of the external knowledge. To assess the resulting state estimates, we use the mean norm squared-error, $\text{MNSE} = \frac{1}{n}\sum_{i=1}^{n} ||x_i - \mu_{i|i}||^2$, with $||\cdot||$ denoting the Euclidean norm. We are concerned with a simple position-velocity state-space model [30] specified by

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad Q = 10^{-5} I_2, \quad R = 10^{-3}.$$

The number of time steps is $n = 50$. The results of the compared algorithms are illustrated in Fig. 5.1.

The MNSE of the NT filter defines a reference level against which the remaining filters are compared. This level is obviously constant as the external observation

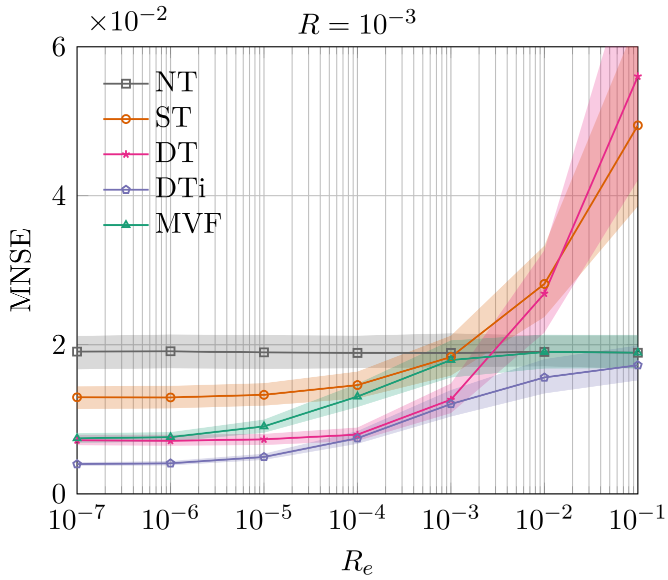

Fig. 5.1: The mean norm squared-error (MNSE) of the primary filter versus the observation variance R_e of the external Kalman filter. The results are averaged over 1000 independent simulation runs, with the solid line being the median and the shaded area delineating the interquartile range. The procedures that are compared are (i) the Kalman filter with *N*o *T*ransfer (*NT*), (ii) *S*tatic Bayesian knowledge *T*ransfer (*ST*) [34], (iii) *D*ynamic Bayesian knowledge *T*ransfer (*DT*) given by Algorithm 7, (iv) an *i*nformally adapted version of DT (*DTi*); and (v) *M*easurement *V*ector *F*usion (*MVF*) [95].

variance does not enter the standard Kalman filter. The error in the remaining filters varies according to the ratio of the primary and external observation variances. We can observe that the proposed DT filter achieves positive knowledge transfer for $R_e < 3 \times 10^{-3}$, which is evidenced by the fact that the error of the DT filter is lower than that of the NT filter in this range. Moreover, the DT filter outperforms the MVF filter in the same interval, and it also outperforms the ST filter for $R_e < 2\times10^{-2}$. The important observation is that the ST and MVF filters meet the performance of the NT filter close to the intersection where $R_e = R$, but the proposed DT filter passes this point with a markedly lower error and meets the NT filter later (i.e. for higher external observation variance). This increased robustness of the DT filter, which now benefits even from external observations that are of a lower quality than the primary ones, is achieved because of its ability to accumulate the external knowledge over multiple time steps via the *dynamic* transfer which is the focus of this chapter. The ST and MVF filters do not have this property, as is evidenced by the fact that their error is, respectively, worse and very similar to the NT filter, above $R_e = R$. However, accumulating external knowledge of increasingly poor quality does lead to a more quickly decreasing performance of the DT filter for $R_e > 2 \times 10^{-2}$.

Algorithm 7 FPD-optimal processing for dynamic transfer between Kalman filters

A. **Backward sweep:**

1. For $i = n$,
   * Compute

$$r_{n|n} = C^\top R^{-1} z_{n|n-1,e},$$
$$S_{n|n} = C^\top R^{-1} C.$$

   * Compute

$$r_{n-1|n} = A^\top (I_{n_x} - L) r_{n|n},$$
$$S_{n-1|n} = A^\top (I_{n_x} - L) S_{n|n} A,$$

   where $L \equiv S_{n|n} Q^{\frac{1}{2}} (Q^{\frac{\top}{2}} S_{n|n} Q^{\frac{1}{2}} + I_{n_x})^{-1} Q^{\frac{\top}{2}}$, I_{n_x} is the $n_x \times n_x$ identity matrix, $Q^{\frac{1}{2}}$ is the Cholesky factor of Q, and $^\top$ denotes matrix transposition.

2. For $i = n-1, \ldots, 2$;
   * Compute

$$r_{i|i} = r_{i|i+1} + C^\top R^{-1} z_{i|i-1,e},$$
$$S_{i|i} = S_{i|i+1} + C^\top R^{-1} C.$$

   * Compute

$$r_{i-1|i} = A^\top (I_{n_x} - L) r_{i|i},$$
$$S_{i-1|i} = A^\top (I_{n_x} - L) S_{i|i} A,$$

   where where $L \equiv S_{i|i} Q^{\frac{1}{2}} (Q^{\frac{\top}{2}} S_{i|i} Q^{\frac{1}{2}} + I_{n_x})^{-1} Q^{\frac{\top}{2}}$.

B. **Forward sweep:**

1. For $i = 1$, set $\mu_{1|0}, \Sigma_{1|0}$ and compute

$$z_{1|0} = C\mu_{1|0},$$
$$R_{1|0} = C\Sigma_{1|0} C^\top + R,$$
$$\mu_{1|1} = \mu_{1|0} + K(z_1 - z_{1|0}),$$
$$\Sigma_{1|1} = \Sigma_{1|0} - K R_{1|0} K^\top,$$

   where $K \equiv \Sigma_{1|0} C^\top R_{1|0}^{-1}$.

2. For $i = 2, \ldots, n$;
   * Compute

$$\mu_{i|i-1} = (I_{n_x} - \Sigma_i^o S_{i|i}) A \mu_{i-1|i-1} + \Sigma_i^o r_{i|i},$$
$$\Sigma_{i|i-1} = (I_{n_x} - \Sigma_i^o S_{i|i}) A \Sigma_{i-1|i-1} A^\top (I_{n_x} - \Sigma_i^o S_{i|i})^\top + \Sigma_i^o,$$

   where $\Sigma_i^o = Q^{\frac{1}{2}} (Q^{\frac{\top}{2}} S_{i|i} Q^{\frac{1}{2}} + I_{n_x})^{-1} Q^{\frac{\top}{2}}$.

   * Compute

$$z_{i|i-1} = C\mu_{i|i-1},$$
$$R_{i|i-1} = C\Sigma_{i|i-1} C^\top + R,$$
$$\mu_{i|i} = \mu_{i|i-1} + K(z_i - z_{i|i-1}),$$
$$\Sigma_{i|i} = \Sigma_{i|i-1} - K R_{i|i-1} K^\top,$$

   where $K \equiv \Sigma_{i|i-1} C^\top R_{i|i-1}^{-1}$.

CONCLUSION

Chapter 1 is concerned with the design of the projection-based Rao-Blackwellized particle filter for estimating static parameters in the conditionally conjugate state-space models. The primary objective was to devise an SMC-based approach which counteracts the particle path degeneracy problem. This was accomplished by formulating the projection-based updates for computing the statistics representing the posterior density of the parameters in order to avoid their resampling and thus make them less affected by the degenerate particle trajectories. The results reveal that the proposed solution indeed decreases the variance of the parameter estimates over multiple simulation runs compared to the plain Rao-Blackwellized particle filter, and it therefore suffers less from the degeneracy problem. Moreover, the proposed approach outperforms a number of alternative techniques for parameter estimation in nonlinear and non-Gaussian state-space models. In the presented experiment, the resulting solution has approximately the same computational complexity as the basic Rao-Blackwellized particle filter but provides an improved estimation precision. Therefore, for the same precision level of both these methods, we obtain a considerable decrease in the computational time in favor of the proposed method. When changing the signal-to-noise ratio in the considered experimental setup, the proposed projection-based Rao-Blackwellized particle filter starts to be more sensitive to the initial setting of the posterior statistics. This increased sensitivity is mainly caused by the adoption of the bootstrap proposal density. Therefore, designing a suitable approximation of the optimal proposal may provide more robustness in this sense.

The proposed algorithm can be applied to, e.g., Bayesian optimization [39], seasonal epidemics detection [56], charge estimation of batteries [62], etc.

The idea of computing the projections seems to provide an interesting opportunity for counteracting the particle path degeneracy problem. Therefore, the primary aim of future work should be focused on different strategies for the evolution of the statistics and investigating dependence of the algorithm on the forgetting properties of the state-space model. A possible generalization of the proposed approach is to use an MCMC procedure [35] at each iteration in order to facilitate application to nonlinear and non-Gaussian state-space models without the tractable substructure with respect to the parameters. An increase in the computational complexity of such an algorithm should be expected. Another possibility is to extend the method to allow for the parameter inference in the conditionally conjugate jump Markov models. Such a method could then be applied to, e.g., traffic flow monitoring [77] and evaluation of the stock return sensitivity to macroeconomic news announcement [41]. Alternatively, to enable tracking of time-varying parameters, it is also tempting to extend the estimation procedure by a suitable forgetting strategy [52, 49].

Chapter 2 investigates the possibility of using alternative stabilized forgetting in the context of SMC-based estimation of slowly-varying parameters in conditionally conjugate state-space models. It is demonstrated that the proposed Rao-Blackwellized particle filter outperforms the one introduced in [74]; more concretely, the estimates of the measurement noise variance are less biased, and the approach also reduces the variance of the estimated parameters. This is achieved in a computationally more efficient way. Specifically, in the present experiment, the proposed method reduces the computational time by an order of magnitude. The algorithm offers a fair degree of adaptability by allowing us to tune the forgetting of the past information by the hyper-parameters of the alternative density. This makes the method slightly more difficult to tune (setting the statistic μ_A of the alternative density to zero always substantially simplifies the initial tuning).

There is a multitude of practical problems for which the proposed technique can be utilized, such as estimating parameters of automotive-grade sensors [10], tire radii estimation [65], etc.

The proposed algorithm—similarly to the one from Chapter 1—can also be extended to incorporate the MCMC steps, thus broadening the range of admissible models to completely nonlinear and non-Gaussian state-space models. However, to simplify the applicability of the proposed method, the main direction of future work will consist in facilitating an autonomous adaptation of the hyper-parameters of the alternative density. A possible approach how to solve this requirement lies in the hierarchical Bayesian modeling [9].

Chapter 3 designs Rao-Blackwellized particle Gibbs kernels for smoothing in jump Markov nonlinear models. The experimental evidence shows that the proposed algorithms are computationally more efficient than the competing approaches. An additional investigation of the proposed (ancestor-sampling-based) procedure revealed that the introduction of the artificial prior is redundant. However, changing the scale of the backward information filtering recursion—provided by the associated design step—is necessary. Practically, this means that we can set the artificial prior to one, while leaving the related derivations intact. The necessary change of scale is then still preserved in the algorithm design. A formally more suitable derivation of this part of the algorithm is provided in Chapter 4. In various experiments, the algorithm *without* the change of scale provided poor estimation precision compared to the one *with* this change. In fact, the former version numerically failed several times during the experiments, whereas the latter one always prevailed.

A possible application scenario for the developed smoothing algorithm consists in offline processing of experimental data in indoor localization [71], target classification [4], fault detection [86], etc. In such cases, the proposed method can serve as a generator of reference trajectories for the development and validation of online algorithms.

Chapter 4 proposes the Rao-Blackwellized particle stochastic approximation expectation algorithm for jump Markov nonlinear models, offering a computationally more efficient alternative to the basic formulation which jointly samples both the latent variables. The efficiency depends on the distance between the individual regimes of the jump Markov nonlinear model. On the one hand, if the regime parameters are substantially different, it is easy to detect the changes in the observations and the algorithm provides best efficiency. On the other hand, if the regime parameters are very similar, it is harder to capture the changes in the observations and the method is less efficient. However, in the latter case, it is no more reasonable to use an algorithm which assumes both continuous state and discrete regime variables, it would suffice to use an algorithm which considers only the continuous state variable. The rationale behind this statement is that the changes in the observations become so small that they will be hidden in the noise, and there is thus no need to consider a jump Markov nonlinear model but rather a plain nonlinear non-Gaussian state-space model. Therefore, the best performance can be expected when the changes are clearly distinguishable from the noise. This behavior is common for all algorithms dealing with switching models.

The method is applicable to parameter identification in diverse application areas such as option pricing in financial markets [17], engine performance diagnosis [91], land vehicle positioning [15], etc.

The proposed Rao-Blackwellized particle stochastic approximation expectation algorithm can be seen as an instance of where the Rao-Blackwellized particle Gibbs kernel from Chapter 3 can be utilized. This building block opens up for the design of various identification strategies in jump Markov nonlinear models, including particle Gibbs with ancestor sampling for Bayesian parameter inference [58].

Chapter 5 devises an FPD-based optimal dynamic Bayesian transfer learning approach and shows its application to probabilistic knowledge transfer between a pair of Kalman filters. The resulting experiments demonstrate that FPD offers a potential for building a versatile framework for Bayesian transfer learning. However, there is still the question of dealing with the aforementioned insensitivity to the second moment transfer. A possible answer to this problem may lie in the recently proposed hierarchical FPD-based Bayesian transfer learning [79], which will be the primary aim of future work.

We have focused thusfar on the basic scenario of one-directional knowledge transfer between two nodes. The natural extension of the proposed approach therefore consists of (i) facilitating the knowledge transfer among a greater number of nodes and (ii) making the transfer bi-directional. Specifically, the former point will require us to introduce an optimal weighting mechanism to assess knowledge in a network of nodes. Another possible extension is to replace the Kalman filters with different forms of Gaussian filters [81]. The application of sequential Monte Carlo methods [25] may also be feasible. Finally, one can change the transferred knowledge and conditional independence assumptions of the adopted model in order to propose other FPD-based transfer learning options, such as transfer of the external joint state predictor.

BIBLIOGRAPHY

[1] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1):5–43, 2003.

[2] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

[3] C. Andrieu, A. Doucet, and V. Tadic. On-line parameter estimation in general state-space models. In *Proceedings of the 44th IEEE Conference on Decision and Control (CDC)*, pages 332–337, 2005.

[4] D. Angelova and L. Mihaylova. Joint target tracking and classification with particle filtering and mixture Kalman filtering using kinematic radar information. *Digital Signal Processing*, 16(2):180–204, 2006.

[5] T. T. Ashley and S. B. Andersson. A sequential Monte Carlo framework for the system identification of jump Markov state space models. In *Proceedings of 2014 American Control Conference (ACC)*, pages 1144–1149, 2014.

[6] O. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory.* Wiley, 1978.

[7] V. Bastani, L. Marcenaro, and C. Regazzoni. A particle filter based sequential trajectory classifier for behavior analysis in video surveillance. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 3690–3694, 2015.

[8] Y. Bengio. Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pages 17–36, 2012.

[9] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory.* Wiley, 1994.

[10] K. Berntorp and S. D. Cairano. Offset and noise estimation of automotive-grade sensors using adaptive particle filtering. In *2018 Annual American Control Conference (ACC)*, pages 4745–4750, 2018.

[11] A. R. Braga, M. G. S. Bruno, E. Özkan, C. Fritsche, and F. Gustafsson. Cooperative terrain based navigation and coverage identification using consensus. In *Proceedings of 18th International Conference on Information Fusion (Fusion)*, pages 1190–1197, 2015.

[12] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state–space models. Technical Report CUED/F-INFENG/TR.498, Cambridge University, 2004.

[13] O. Cappé. Online sequential Monte Carlo EM algorithm. In *15th IEEE Workshop on Statistical Signal Processing*, pages 37–40, 2009.

[14] O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models.* Springer, 2005.

[15] F. Caron, M. Davy, E. Duflos, and P. Vanheeghe. Particle filtering for multisensor data fusion with switching observation models: Application to land vehicle positioning. *IEEE Transactions on Signal Processing*, 55(6):2703–2719, 2007.

[16] C. M. Carvalho, M. S. Johannes, H. F. Lopes, and N. G. Polson. Particle learning and smoothing. *Statistical Science*, 25(1):88–106, 2010.

[17] C. M. Carvalho and H. F. Lopes. Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics & Data Analysis*, 51(9):4526–4542, 2007.

[18] G. Casella and C. P. Robert. Rao-Blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94, 1996.

[19] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[20] P. Del Moral, A. Doucet, and S. Singh. Forward smoothing using sequential Monte Carlo. *arXiv preprint arXiv:1012.5390*, 2010.

[21] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94–128, 1999.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39(1):1–38, 1977.

[23] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *The Annals of Applied Probability*, 21(6):2109–2145, 2011.

[24] A. Doucet, N. J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, 2001.

[25] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. In D. Crisan and B. Rozovsky, editors, *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2009.

[26] A. Doucet, A. Smith, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[27] H. Driessen and Y. Boers. Efficient particle filter for jump Markov nonlinear systems. *IEE Proceedings - Radar, Sonar and Navigation*, 152(5):323–326, 2005.

[28] V. Dukic, H. F. Lopes, and N. G. Polson. Tracking epidemics with Google flu trends data and a state-space SEIR model. *Journal of the American Statistical Association*, 107(500):1410–1426, 2012.

[29] L. Eeckhout. Is Moore's law slowing down? What's next? *IEEE Micro*, 37(4):4–5, 2017.

[30] R. Faragher. Understanding the basis of the Kalman filter via a simple and intuitive derivation. *IEEE Signal Processing Magazine*, 29(5):128–132, 2012.

[31] P. Fearnhead. Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics*, 11(4):848–862, 2002.

[32] P. Fearnhead and P. Clifford. On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(4):887–899, 2003.

[33] P. Fearnhead and H. R. Künsch. Particle filters and data assimilation. *Annual Review of Statistics and Its Application*, 5(1):1–31, 2018.

[34] C. Foley and A. Quinn. Fully probabilistic design for knowledge transfer in a pair of Kalman filters. *IEEE Signal Processing Letters*, 25(4):487–490, 2018.

[35] W. R. Gilks and C. Berzuini. Following a moving target–Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):127–146, 2001.

[36] W. J. Godinez, M. Lampe, P. Koch, R. Eils, B. Muller, and K. Rohr. Identifying virus-cell fusion in two-channel fluorescence microscopy image sequences based on a layered probabilistic approach. *IEEE Transactions on Medical Imaging*, 31(9):1786–1808, 2012.

[37] S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, 99(465):156–168, 2004.

[38] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2):107–113, 1993.

[39] R. B. Gramacy and N. G. Polson. Particle learning of Gaussian process models for sequential design and optimization. *Journal of Computational and Graphical Statistics*, 20(1):102–118, 2011.

[40] R. B. Gramacy, M. Taddy, and S. M. Wild. Variable selection and sensitivity analysis using dynamic trees, with an application to computer code performance tuning. *The Annals of Applied Statistics*, 7(1):51–80, 2013.

[41] T. Hann Law, D. Song, and A. Yaron. Fearing the Fed: How Wall Street reads main street. *SSRN Electronic Journal*, 2017.

[42] D. Isele and A. Cosgun. Transferring autonomous driving knowledge on simulated and real intersections. *arXiv preprint arXiv:1712.01106*, 2017.

[43] P. E. Jacob. Sequential Bayesian inference for implicit hidden Markov models and current limitations. *ESAIM: Proceedings and Surveys*, 51:24–48, 2015.

[44] P. E. Jacob, L. M. Murray, and S. Rubenthaler. Path storage in the particle filter. *Statistics and Computing*, 25(2):487–496, 2015.

[45] M. Johannes, A. Korteweg, and N. Polson. Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69(2):611–644, 2014.

[46] M. Johannes, L. A. Lochstoer, and Y. Mou. Learning about consumption dynamics. *The Journal of Finance*, 71(2):551–600, 2016.

[47] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin. On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351, 2015.

[48] A. Karbalayghareh, X. Qian, and E. R. Dougherty. Optimal Bayesian transfer learning. *arXiv preprint arXiv:1801.00857*, 2018.

[49] M. Kárný. Approximate Bayesian recursive estimation. *Information Sciences*, 285(1):100–111, 2014.

[50] M. Kárný and J. Andrýsek. Use of Kullback-–Leibler divergence for forgetting. *International Journal of Adaptive Control and Signal Processing*, 23(10):961–975, 2009.

[51] G. Kitagawa. A self-organizing state-space model. *Journal of the American Statistical Association*, 93(443):1203–1215, 1998.

[52] R. Kulhavý and M. B. Zarrop. On a general concept of forgetting. *International Journal of Control*, 58(4):905–924, 1993.

[53] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[54] W. Li, Y. Jia, J. Du, and J. Zhang. Distributed multiple-model estimation for simultaneous localization and tracking with NLOS mitigation. *IEEE Transactions on Vehicular Technology*, 62(6):2824–2830, 2013.

[55] K. Lidstrom and T. Larsson. Model-based estimation of driver intentions using particle filtering. In *Proceedings of 11th International IEEE Conference on Intelligent Transportation Systems*, pages 1177–1182, 2008.

[56] J. Lin and M. Ludkovski. Sequential Bayesian inference in hidden Markov stochastic kinetic models with application to detection and response to seasonal epidemics. *Statistics and Computing*, 24(6):1047–1062, 2014.

[57] F. Lindsten. An efficient stochastic approximation EM algorithm using conditional particle filters. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6274–6278, 2013.

[58] F. Lindsten, M. I. Jordan, and T. B. Schön. Particle Gibbs with ancestor sampling. *Journal of Machine Learning Research*, 15(1):2145–2184, 2014.

[59] F. Lindsten and T. B. Schön. Backward simulation methods for Monte Carlo statistical inference. *Foundations and Trends in Machine Learning*, 6(1):1–143, 2013.

[60] F. Lindsten, T. B. Schön, and L. Svensson. A non-degenerate Rao-Blackwellised particle filter for estimating static parameters in dynamical models. *IFAC Proceedings Volumes*, 45(16):1149–1154, 2012.

[61] J. S. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. De Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, chapter 10, pages 197–223. Springer, New York, 2001.

[62] X. Liu, Z. Chen, C. Zhang, and J. Wu. A novel temperature-compensated model for power li-ion batteries with dual-particle-filter state of charge estimation. *Applied Energy*, 123:263–272, 2014.

[63] Z. Liu, G. Sun, S. Bu, J. Han, X. Tang, and M. Pecht. Particle learning framework for estimating the remaining useful life of lithium-ion batteries. *IEEE Transactions on Instrumentation and Measurement*, 66(2):280–293, 2017.

[64] H. F. Lopes and C. M. Carvalho. Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, 137(10):3082 – 3091, 2007.

[65] C. Lundquist, R. Karlsson, E. Ozkan, and F. Gustafsson. Tire radii estimation using a marginalized particle filter. *IEEE Transactions on Intelligent Transportation Systems*, 2(15):663–672, 2014.

[66] C. Mavroforakis, I. Valera, and M. Gomez-Rodriguez. Modeling the dynamics of learning activity on the web. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1421–1430, 2017.

[67] D. Q. Mayne. A solution of the smoothing problem for linear dynamic systems. *Automatica*, 4(2):73–92, 1966.

[68] C. Mukherjee and M. West. Sequential Monte Carlo in model comparison: Example in cellular dynamics in systems biology. In *JSM Proceedings, Section on Bayesian Statistical Science*, pages 1274–1287, 2009.

[69] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[70] C. Nemeth, P. Fearnhead, and L. Mihaylova. Sequential Monte Carlo methods for state and parameter estimation in abruptly changing environments. *IEEE Transactions on Signal Processing*, 62(5):1245–1255, 2014.

[71] M. Nicoli, C. Morelli, and V. Rampa. A jump markov particle filter for localization of moving terminals in multipath indoor scenarios. *IEEE Transactions on Signal Processing*, 56(8):3801–3809, 2008.

[72] J. Olsson, J. Westerborn, et al. Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm. *Bernoulli*, 23(3):1951–1996, 2017.

[73] E. Özkan, F. Lindsten, C. Fritsche, and F. Gustafsson. Recursive maximum likelihood identification of jump Markov nonlinear systems. *IEEE Transactions on Signal Processing*, 63(3):754–765, 2015.

[74] E. Özkan, V. Šmídl, S. Saha, C. Lundquist, and F. Gustafsson. Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters. *Automatica*, 49(6):1566–1575, 2013.

[75] S. J. Pan. Transfer learning. In *Data Classification: Algorithms and Applications*, pages 537–558. Chapman and Hall/CRC, 2015.

[76] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa. Visual domain adaptation: A survey of recent advances. *IEEE Signal Processing Magazine*, 32(3):53–69, 2015.

[77] N. Polson and V. Sokolov. Bayesian particle tracking of traffic flows. *IEEE Transactions on Intelligent Transportation Systems*, 19(2):345–356, 2018.

[78] G. Poyiadjis, A. Doucet, and S. S. Singh. Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98(1):65–80, 2011.

[79] A. Quinn, M. Kárný, and T. V. Guy. Optimal design of priors constrained by external predictors. *International Journal of Approximate Reasoning*, 84:150–158, 2017.

[80] S. Saha, G. Hendeby, and F. Gustafsson. Mixture Kalman filters and beyond. In *Current Trends in Bayesian Methodology with Applications*, pages 537–562, 2015.

[81] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[82] T. B. Schön, A. Wills, and B. Ninness. System identification of nonlinear state-space models. *Automatica*, 47(1):39–49, 2011.

[83] I. Smal, E. Meijering, K. Draegestein, N. Galjart, I. Grigoriev, A. Akhmanova, M. van Royen, A. Houtsmuller, and W. Niessen. Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering. *Medical Image Analysis*, 12(6):764–777, 2008.

[84] G. Storvik. Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, 50(2):281–289, 2002.

[85] A. Svensson, T. B. Schön, and F. Lindsten. Identification of jump Markov linear models using particle filters. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 6504–6509. IEEE, 2014.

[86] S. Tafazoli and X. Sun. Hybrid system state tracking and fault detection using particle filters. *IEEE Transactions on Control Systems Technology*, 14(6):1078–1087, 2006.

[87] M. E. Taylor and P. Stone. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10:1633–1685, 2009.

[88] L. Torrey and J. Shavlik. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, pages 242–264. IGI Global, 2010.

[89] T. L. M. Van Kasteren, G. Englebienne, and B. J. A. Kröse. Transferring knowledge of activity recognition across sensor networks. In *International Conference on Pervasive Computing*, pages 283–300. Springer, 2010.

[90] P. Vidoni. Exponential family state-space models based on a conjugate latent process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):213–221, 1999.

[91] P. Wang and R. X. Gao. Markov nonlinear system estimation for engine performance tracking. *Journal of Engineering for Gas Turbines and Power*, 138(9):091201, 2016.

[92] Y. Wang, V. Gupta, and P. J. Antsaklis. Stochastic passivity of discrete-time markovian jump nonlinear systems. In *2013 American Control Conference*, pages 4879–4884, 2013.

[93] N. Whiteley. Stability properties of some particle filters. *The Annals of Applied Probability*, 23(6):2500–2537, 2013.

[94] N. Whiteley, C. Andrieu, and A. Doucet. Efficient Bayesian inference for switching state-space models using discrete particle Markov chain Monte Carlo methods. *arXiv preprint arXiv:1011.2437*, 2010.

[95] D. Willner, C. B. Chang, and K. P. Dunn. Kalman filter algorithms for a multi-sensor system. In *1976 IEEE Conference on Decision and Control including the 15th Symposium on Adaptive Processes*, volume 15, pages 570–574. IEEE, 1976.

[96] A. Wilson, A. Fern, and P. Tadepalli. Transfer learning in sequential decision problems: A hierarchical Bayesian approach. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pages 217–227, 2012.

[97] C. Zeng, Q. Wang, S. Mokhtari, and T. Li. Online context-aware recommendation with time varying multi-armed bandit. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2025–2034, 2016.

Milan Papež

Work Address Institute of Information Theory and Automation
The Czech Academy of Sciences
Pod Vodárenskou věží 4
182 08 Prague 8
Czech Republic
papez@utia.cas.cz

Research Interests Sequential Monte Carlo, Markov Chain Monte Carlo, Markov Chains, Statistical Signal Processing, Bayesian Decision-Making

Education Brno University of Technology, Czech Republic

Ph.D., Cybernetics, Control and Measurements — 2013 to present
- Thesis Topic: *Monte Carlo-Based Identification Strategies for State-Space Models*
- Supervisor: Prof. Pivoňka P.

M.Eng., Cybernetics, Control and Measurements (with Honors) — 2011 to 2013
- Thesis Topic: *Optimal State Estimation of a Navigation Model System*
- Supervisor: Dr. Dokoupil J.

B.Eng., Automation and Measurement — 2008 to 2011
- Thesis Topic: *System for Temperature Monitoring in Server Rooms*
- Supervisor: Dr. Macho T.

Academic Internships Aalborg University, Denmark — 2017 to 2017
Department of Automation and Control
- Erasmus exchange program (five months)

Research Experience Research Assistant

Institute of Information Theory and Automation — 2018 to present
The Czech Academy of Sciences

Central European Institute of Technology — 2014 to 2017
Brno University of Technology

Awards Student Awards – Brno University of Technology
- dean's prize — 2013
- third place at faculty student competition — 2013

Teaching Experience Teaching Assistant

Aalborg University, Denmark — 2017 to 2017
Department of Automation and Control
- Matrix Computations and Convex Optimization

Brno University of Technology, Czech Republic — 2013 to 2017
Department of Control and Instrumentation
- Computer Science in Automation
- Computer Control
- Optimization of Controllers

Hardware & Software Skills	Matlab(+OOP), Simulink, C, C++, LaTeX, Embedded system design
Language Skills	Czech - Native or bilingual proficiency English - Full professional working proficiency German - Elementary proficiency

Published Papers

1. **Papež, M.**, Quinn A. "Dynamic Bayesian knowledge transfer between a pair of Kalman filters" In *28th IEEE International Workshop on Machine Learning for Signal Processing*, pages 1-6, 2018.
2. **Papež, M.** "A particle stochastic approximation EM algorithm to identify jump Markov nonlinear models" In *18th IFAC International Symposium on System Identification*, pages 676-681, 2018.
3. **Papež, M.** "Rao-Blackwellized particle Gibbs kernels for smoothing in jump Markov nonlinear models" In *16th IEEE European Control Conference*, pages 2466-2471, 2018.
4. **Papež, M.** "A projection-based Rao-Blackwellized particle filter to estimate parameters in conditionally conjugate state-space models" In *20th IEEE Workshop on Statistical Signal Processing*, pages 268-272, 2018.
5. **Papež, M.** "A Rao-Blackwellized particle filter to estimate the time-varying noise parameters in non-linear state-space models using alternative stabilized forgetting" In *16th IEEE International Symposium on Signal Processing and Information Technology*, pages 229-234, 2016.
6. **Papež, M.** "Sequential Monte Carlo estimation of transition probabilities in mixture filtering problems" In *19th IEEE International Conference on Information Fusion*, pages 1063-1070, 2016.
7. **Papež, M.** "Approximate Bayesian inference methods for mixture filtering with known model of switching" In *17th IEEE International Carpathian Control Conference*, pages 545-551, 2016.
8. **Papež, M.** "On Bayesian decision-making and approximation of probability densities" In *38th IEEE International Conference on Telecommunications and Signal Processing*, pages 499-503, 2015.
9. Dokoupil, J., **Papež, M.**, Václavek, P. "Bayesian comparison of Kalman filters with known covariance matrices" In *13th AIP International Conference on Numerical Analysis and Applied Mathematics ICNAAM 2014*, 2015.
10. Dokoupil, J., **Papež, M.**, Václavek, P. "Comparison of Kalman filters formulated as the statistics of the Normal-inverse-Wishart distribution" In *54th IEEE Conference on Decision and Control*, pages 5008-5013, 2015.
11. **Papež, M.**, Pivoňka, P. "Numerical aspects of inertial navigation" In *12th IFAC Conference on Programmable Devices and Embedded Systems*, pages 262-267, 2013.