Article:
Carmi, A.Y., Mihaylova, L. and Septier, F. (2015) Subgradient-Based Markov Chain Monte Carlo Particle Methods for Discrete-Time Nonlinear Filtering. Signal Processing. ISSN 0165-1684

https://doi.org/10.1016/j.sigpro.2015.10.015

Subgradient-Based Markov Chain Monte Carlo Particle Methods for Discrete-Time Nonlinear Filtering

Avishy Y. Carmi*, Lyudmila Mihaylova and François Septier§**

*Department of Mechanical Engineering, Ben-Gurion University of the Negev, Israel

**Automatic Control and Systems Engineering, University of Sheffield, United Kingdom
Corresponding author

§Institut Mines Tlcom / Télécom Lille, CRIStAL UMR CNRS 9189, France
Email: avishycarmi@gmail.com, l.s.mihaylova@sheffield.ac.uk, francois.septier@telecom-lille.fr

Abstract

This work shows how a carefully designed instrumental distribution can improve the performance of a Markov chain Monte Carlo (MCMC) filter for systems with a high state dimension. We propose a special subgradient-based kernel from which candidate moves are drawn. This facilitates the implementation of the filtering algorithm in high dimensional settings using a remarkably small number of particles. We demonstrate our approach in solving a nonlinear non-Gaussian high-dimensional problem in comparison with a recently developed block particle filter and over a dynamic compressed sensing (l_1 constrained) algorithm. The results show high estimation accuracy.

1. Introduction

Over the past decade we have witnessed an immense rise in the complexity of inference tasks, primarily owing to recent advancements in computation power and ever increasing performance expectations. Filtering algorithms nowadays confront the curse of dimensionality in many real world high dimensional applications which include, among others, multiple object and crowd tracking [1], learning algorithms, and reasoning in multi-agent systems. In many typical problems in these domains and alike, even the most conservative restrictions on the underlying spatio-temporal complexities would not alleviate the dimensionality issue.

A thoughtful study of the class of sequential Monte Carlo methods, otherwise known as particle filters (PFs), has shown that it is the well-known importance sampling technique which renders most state-of-the-art PF schemes inadequate for reasoning in complex high dimensional settings [2]. Recently, various approaches have been proposed for overcoming this downfall. These account for MCMC methods [3, 4], population PFs and log-homotopy particle flow [5]. A number of local sequential Monte Carlo methods, called also block

particle filters have been proposed by representing complex probability density functions (pdfs) with a product of independent pdfs [6, 7, 8].

In this work we derive an efficient version of a genuine and simple MCMC particle filtering algorithm of which several variants appeared in the literature. In particular, our proposed methodology endows the plain MCMC filtering algorithm with an informative instrumental density from which new moves are generated. As part of this, the local subgradient of the likelihood is exploited for potentially steering the produced chain to highly probable regions of the exploration space. This approach possesses an advantage over the prevalent particle refinement technique which utilizes a Metropolis-within-Gibbs stage for making conditional draws - a stage which has been conjectured to deteriorate the chain mixing time [3]. We demonstrate the potential of our approach in complex settings involving nonlinear state dynamics.

The rest of the paper is organised in the following way. Section 2 presents the problem formulation. Section 3 describes a few existing MCMC filtering techniques for high dimensional systems. Section 4 describes a new subgradient-based sampling approach for filtering in potentially high dimensional state spaces. A few illustrative examples of the proposed approach are provided in Section 5. Finally, Section 6 summarises the results.

2. General Particle Filtering Framework

Consider the problem of estimating the state of a dynamic system:

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

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{r}_k), \tag{1b}$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the unknown system state vector, $\mathbf{z}_k \in \mathbb{R}^{n_z}$ is the observed measurement vector and $f(.)$ and $h(.)$ are the process and measurement functions, respectively. The above equations are driven by $\mathbf{v}_k$ and $\mathbf{r}_k$ which stand for the process and observation noises.

According to Bayes rule, the state filtering pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ of the state vector $\mathbf{x}_k$ given the measurement history $\mathbf{z}_{1:k} = \{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$ may be written as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}, \tag{2}$$

where $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ is the normalising constant. The state predictive distribution is given by

the Chapman-Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int_{\mathbb{R}^{n_x}} p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) d\mathbf{x}_k. \tag{3}$$

The evaluation of the right hand side of (2) involves integration which can be avoided in the particle filtering approach by approximating the filtering pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ with a set of particles $\mathbf{x}_{0:k}^{(i)}$, $i = 1, \ldots, N$ and their corresponding weights $w_k^{(i)}$ [9]. Then the posterior density can be written as follows

$$p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}) = \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^{(i)}), \tag{4}$$

where $\delta(.)$ is the Dirac delta function, and the weights are normalised such that $\sum_i w_k^{(i)} = 1$. Each pair $\{\mathbf{x}_{0:k}^{(i)}, w_k^{(i)}\}$ characterises the belief that the system is in state $\mathbf{x}_{0:k}^{(i)}$. An estimate of the variable of interest is obtained by the weighted sum of particles. Two major stages can be distinguished: *prediction* and *update*. During prediction, each particle is modified according to the state model, including the addition of random noise in order to simulate the effect of the noise on the state. In the update stage, each particle's weight is re-evaluated based on the new data. A *resampling* procedure introduces variety in the particles by eliminating those with small weights and replicating the particles with larger weights such that the approximation in (4) still holds. This paper proposes a resample-move scheme which is presented in details as Algorithms 1 and 2.

3. High Dimensional Particle Schemes

The importance sampling approach, which essentially forms the core of every PF algorithm, becomes prohibitively inefficient in high dimensions [2]. Over the past decade this caveat has motivated the derivation of far more sophisticated particle schemes, most of which rely on MCMC techniques [3, 4, 10]. The main purpose of this paper is to corroborate and extend a single promising direction in this regard. We demonstrate the strong potential of a class of genuine MCMC-based particle algorithms.

3.1. Sequential MCMC Filtering

The following sequential filtering scheme is closely related to the inference algorithms presented in [10, 11] (see also [4]). Suppose that at time $k-1$ there are N samples $\{\mathbf{x}_{k-1}^{(i)}\}_{i=1}^{N}$ drawn approximately from the filtering density $p(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})$ (i.e., the previous time

target distribution). A new set of samples $\{\mathbf{x}_k^{(i)}\}_{i=1}^N$ representing $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$ can be then simulated using a tailored Metropolis Hastings (MH) scheme.

The MH algorithm generates samples from an aperiodic and irreducible Markov chain with a predetermined (possibly unnormalised) stationary distribution. This is a constructive method which specifies the Markov transition kernel by means of acceptance probabilities based on the preceding time outcome. Setting the stationary density as the joint filtering pdf $p(\mathbf{x}_k, \mathbf{x}_{k-1} \mid \mathbf{z}_{1:k})$ (the marginal of which is the desired filtering pdf $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$), a new set of samples from this distribution can be obtained after the MH burn-in period. This procedure is described next.

First, we simulate a sample $\mathbf{x}'_k$ from the joint pdf $p(\mathbf{x}_k, \mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})$ by drawing

$$\mathbf{x}'_k \sim p(\mathbf{x}_k \mid \mathbf{x}'_{k-1}), \tag{5}$$

where $\mathbf{x}'_{k-1}$ is uniformly drawn from the empirical approximation of $p(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})$ given by

$$\hat{p}(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1}) = N^{-1} \sum_{i=1}^{N} \delta(\mathbf{x}_{k-1}^{(i)} - \mathbf{x}_{k-1}). \tag{6}$$

This sample $\mathbf{x}'_k$ is accepted or rejected using the following Metropolis rule.

Let $(\mathbf{x}_k^{(i)}, \mathbf{x}_{k-1}^{(i)})$ be a sample from the realised chain of which the stationary distribution is the joint filtering pdf. The MH algorithm accepts the new candidate pair $(\mathbf{x}'_k, \mathbf{x}'_{k-1})$ as the next realisation from the chain with probability

$$\alpha = \min\left\{1, p(\mathbf{z}_k \mid \mathbf{x}'_k)/p(\mathbf{z}_k \mid \mathbf{x}_k^{(i)})\right\}, \tag{7}$$

that is,

$$(\mathbf{x}_k^{(i+1)}, \mathbf{x}_{k-1}^{(i+1)}) = \begin{cases} (\mathbf{x}'_k, \mathbf{x}'_{k-1}), & \text{if } u \leq \alpha \\ (\mathbf{x}_k^{(i)}, \mathbf{x}_{k-1}^{(i)}), & \text{otherwise} \end{cases} \tag{8}$$

with the uniform random variable $u \sim U[0, 1]$. The above sampling scheme may be inefficient in exploring the sample space as the underlying proposal density of a well behaved system (i.e., of which the process noise is of low intensity) introduces relatively small moves. This drawback can be alleviated by a secondary Metropolis-within-Gibbs refinement stage [10, 11].

4. Subgradient-Based Efficient Sampling

The efficiency of MCMC samplers, which mainly refers to the mixing properties of the produced chain, is prominently affected by the proposal density from which the candidate moves are drawn. Good proposals facilitate the exploration of the sample space and in particular of high probability regions irrespectively of the initial conditions. The mixing problem is partially alleviated in both [10] and [11] where Metropolis within Gibbs refinement stages are incorporated into the basic MCMC scheme. This approach has proved itself viable for various multi object tracking applications.

In this work we investigate a rather different type of proposal in which the (sub)gradient information of the likelihood is taken into account. As it would be demonstrated in the ensuing this unique proposal facilitates the application of MCMC filtering in high dimensional state spaces (as far as particle filtering is concerned) using a remarkably small number of particles. The idea consists of constructing a proposal out of set of improved samples using the joint propagated pdf $p(\mathbf{x}_k, \mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})$. Thus, the obtained samples from (5) are pushed towards high probability regions based on the (sub)gradient of the likelihood

$$\bar{\mathbf{x}}_k^{(i)} = \mathbf{x}_k'^{(i)} - \lambda^{(i)} \frac{\log p(\mathbf{z}_k \mid \mathbf{x}_k'^{(i)})}{\| t^{(i)} \|_2^2} t^{(i)}, \quad i = 1, \ldots, N, \tag{9}$$

where $\| \cdot \|_2$ denotes the Euclidean norm, the relaxation parameter $\lambda^{(i)} \sim p_\lambda$ is sampled for every i from some prescribed *steering distribution* p_λ (e.g., uniform), and $t^{(i)} := \partial \log p(\mathbf{z}_k \mid \mathbf{x}_k)/\partial \mathbf{x}_k$ is the associated subgradient with respect to $\mathbf{x}_k$, computed at $\mathbf{x}'^{(i)}_k$. This technique is essentially related to the acclaimed iterative convex optimisation method known as subgradient projection [12]. Having the set of improved and propagated particles, $\{\bar{\mathbf{x}}_k^{(i)}, \mathbf{x}_{k-1}'^{(i)}\}_{i=1}^N$, a regularised proposal is constructed in the following manner

$$q(\bar{\mathbf{y}}_k) \propto \sum_{i=1}^{N} \mathcal{N}(\bar{\mathbf{y}}_k \mid \bar{\mathbf{y}}_k^{(i)}, \sigma^2), \;\; \bar{\mathbf{y}}_k = [\bar{\mathbf{x}}_k^T, \mathbf{x}_{k-1}'^T]^T, \; \bar{\mathbf{y}}_k^{(i)} = [(\bar{\mathbf{x}}_k^{(i)})^T, (\mathbf{x}_{k-1}'^{(i)})^T]^T, \tag{10}$$

where $\mathcal{N}(\cdot)$ and σ denote the normal distribution and a roughening intensity parameter, respectively. A MH procedure is then carried out in a fashion similar to (8). This time, however, the acceptance probability of a new candidate pair $(\bar{\mathbf{x}}_k, \mathbf{x}'_{k-1}) \sim q(\bar{\mathbf{y}}_k)$, is given by

$$\alpha = \min \left\{ 1, \frac{p(\mathbf{z}_k \mid \bar{\mathbf{x}}_k) \hat{p}(\bar{\mathbf{x}}_k, \mathbf{x}'_{k-1} \mid \mathbf{z}_{1:k-1}) q(\mathbf{y}_k^{(i)})}{p(\mathbf{z}_k \mid \mathbf{x}_k^{(i)}) \hat{p}(\mathbf{x}_k^{(i)}, \mathbf{x}_{k-1}^{(i)} \mid \mathbf{z}_{1:k-1}) q(\bar{\mathbf{y}}_k)} \right\}, \tag{11}$$

where $\bar{\mathbf{y}}_k) = \left[\mathbf{x}_k^T, \mathbf{x}_{k-1}^T\right]^T$ is a vector containing the pair before the mix and $\hat{p}(\bar{\mathbf{x}}_k, \mathbf{x}'_{k-1} \mid$

$\mathbf{z}_{1:k-1}$) is the density of the move.

4.1. Setting Proposal and Steering Distributions

In the above scheme the proposal distribution $q(\bar{\mathbf{y}}_k)$ is obtained prior to the application of the MH stage. This computationally excessive, albeit necessary, step is used in conjunction with the MH for producing a reversible chain that will ultimately converge to the prescribed stationary distribution. An efficient alternative for computing $q(\bar{\mathbf{y}}_k)$ is by replacing (10) with a single Gaussian of which the statistical moments correspond to the sample mean and covariance of the population $\{\bar{\mathbf{y}}_k^{(i)}\}_{i=1}^N$. A pseudo-code of this variant of the MCMC particle filter is provided in Algorithm 1.

An approach that is likely to have an improved MH acceptance rate relies on using two distinct proposals $q_l(\bar{\mathbf{y}}_k)$, $l = 1, 2$ each of which involves a different steering distribution $p_\lambda^l(\lambda)$. The new moves are then sampled from each of these two proposals in an alternating fashion. This approach, which is provided here without a theoretical justification, is summarised in Algorithm 2.

Algorithm 1 MCMC Particle Filtering Algorithm

1: Given previous time samples $\mathbf{x}_{k-1}^{\prime(i)}$, $i = 1, \ldots, N$ perform the following steps.
2: Draw $\mathbf{x}_k^{\prime\,(i)} \sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{\prime(i)})$, $i = 1, \ldots, N$.
3: Use (9) for producing $\bar{\mathbf{x}}_k^{(i)}$, $i = 1, \ldots, N$. The set $\{\mathbf{x}_k^{\prime\,(i)}, \mathbf{x}_{k-1}^{\prime(i)}\}_{i=1}^N$ simulates $\hat{p}(\mathbf{x}_k, \mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})$, whereas $\{\bar{\mathbf{y}}_k^{(i)}\}_{i=1}^N$, $\bar{\mathbf{y}}_k^{(i)} = [(\bar{\mathbf{x}}_k^{(i)})^T, (\mathbf{x}_{k-1}^{\prime(i)})^T]^T$ simulates $q(\bar{\mathbf{y}}_k) = \mathcal{N}(\bar{\mathbf{y}}_k \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$ where

$$\boldsymbol{\mu}_k = N^{-1} \sum_{i=1}^{N} \bar{\mathbf{y}}_k^{(i)}, \quad \boldsymbol{\Sigma}_k = N^{-1} \sum_{i=1}^{N} \left[\bar{\mathbf{y}}_k^{(i)} - \boldsymbol{\mu}_k\right] \left[\bar{\mathbf{y}}_k^{(i)} - \boldsymbol{\mu}_k\right]^T$$

4: **for** i=1, ..., $N + N_{Burn-in}$ **do**
5: Draw $(\bar{\mathbf{x}}_k, \mathbf{x}_{k-1}) \sim q(\bar{\mathbf{y}}_k)$.
6: Accept the new move as a sample in the chain $\mathbf{x}_k^{(i)} = \bar{\mathbf{x}}_k$ with probability α given in (11).
7: **end for**
8: Retain only N samples $\mathbf{x}_k^{(i)}$ subsequent to the end of the burn-in period.

Algorithm 2 Alternate Steering MCMC

1: Simulate $q_1(\bar{\mathbf{y}}_k)$ and $q_2(\bar{\mathbf{y}}_k)$ using two distinct steering distributions p_λ^1 and p_λ^2, respectively.
2: **for** i=1, ..., $N + N_{Burn-in}$ **do**
3: Draw $(\bar{\mathbf{x}}_k, \mathbf{x}_{k-1}) \sim q(\bar{\mathbf{y}}_k)$ where $q(\bar{\mathbf{y}}_k) = q_1(\bar{\mathbf{y}}_k)$ if $(i \bmod 2) = 1$, and $q(\bar{\mathbf{y}}_k) = q_2(\bar{\mathbf{y}}_k)$, otherwise.
4: Accept the new move as a sample in the chain $\mathbf{x}_k^{(i)} = \bar{\mathbf{x}}_k$ with probability α given in (11).
5: **end for**
6: Retain only N samples $\mathbf{x}_k^{(i)}$ subsequent to the end of the burn-in period.

5. Illustrative Examples

In the following examples we compare the performance of a few nonlinear filtering algorithms applied to systems with $n_x = 100$ states. The filters refer to an Extended Kalman Filter (EKF), the compressed sensing Kalman filter (CSKF) of [13], and the proposed MCMC particle filtering algorithm. The MCMC schemes use no more than 3000 particles and 1000 burn-in samples. The steering distributions p_λ^1 and p_λ^2 are set as $U[-6, 6]$ and $U[0, 6]$, respectively. The non alternating MCMC scheme uses $\lambda \sim p_\lambda^1$.

In the first example we consider a system model, which is an extension of the wide spread example from [9]

$$\mathbf{x}_k^j = \mathbf{x}_{k-1}^j + \frac{25 \sum_{i=1}^{n_x} \mathbf{x}_{k-1}^i}{1 + \left(\sum_{i=1}^{n_x} \mathbf{x}_{k-1}^i\right)^2} + \cos(1.2k) + \mathbf{v}_{k-1}^j \tag{12a}$$

$$\mathbf{z}_k^j = \frac{\left(\mathbf{x}_k^j\right)^2}{20} + \mathbf{r}_k^j, \;\; j = 1, \ldots, n_x \tag{12b}$$

where the superscript j denotes the jth element in the vector. The noises $\mathbf{v}_k^j$ and $\mathbf{r}_k^j$ are assumed to be Gaussian with unit covariance matrices. All other related parameters are set as in [9]. The residual resampling algorithm [14] is applied here. This is a two step procedure making use of sampling-importance-resampling (SIR) scheme.

The performance of the EKF and of both MCMC variants, namely, the alternate steering MCMC (alternating) and the non-alternating, is shown in Figures 1 and 2. The normalised RMSE, defined as $E\left[\| \hat{\mathbf{x}}_k - \mathbf{x}_k \|_2 \; / \; \| \mathbf{x}_k \|_2\right]^{1/2}$, is approximated based on 50 Monte Carlo runs and is shown in Fig. 1. Figure 2 presents the mean acceptance rate of both MCMC variants. Figure 2 demonstrates the superiority of the alternating MCMC filtering approach in terms of estimation accuracy and sampling efficiency (i.e., improved acceptance rate). A comparison with the block PF developed in [8] is also presented. The block PF splits the high dimensional state vector into low-dimensional state vectors and calculate likelihoods over these low dimensional state sub-vectors. The block PF is implemented with the same number $N = 3000$ of particles as in the proposed MCMC filters and respectively with blocks with state sub-vectors, with sizes, respectively, 25 and 100. The results are shown on Figure 1. Figure 1 demonstrates the outperformance of the sub-gradient MCMC algorithms compared with the block PF of Rebeschini and the EKF. The block PF [8] calculates independent likelihoods over blocks of smaller state sub-spaces. However since the subgradient MCMC filter uses the latest measurements and moves the particles towards more likely regions, the subgradient MCMC algorithm outperforms the block PF.

In our last example we demonstrate the performance of the non-alternating MCMC approach in solving a dynamic compressed sensing problem [13]. The system model is similar to the one considered in [13] with $n_x = 100$ states out of which only 10 are either non-vanishing or non-compressible (i.e., the state process is sparse/compressible). The signal itself becomes corrupted over time and its corresponding complexity in the sense of sparseness rises (see illustration in Fig. 3). At each time step the observations are generated from $\mathbf{z}_k = H\mathbf{x}_k + \mathbf{r}_k$. The system and measurement noise covariances are unit matrices and the sensing matrix is in the form: $\mathbf{H} = [\text{sign}(z(1), \ldots, \text{sign}(z(n_z))]$ where $\text{sign}(z(i))$ denotes the sign function of the ith element of $\mathbf{z}_k$ (i.e. $\text{sign}(z_k(i)) = 1$ if $z_k(i) > 0$) and $\text{sign}(z_k(i)) = -1$ otherwise. The likelihood function of the MCMC algorithm is given as

$$p(\mathbf{z}_k \mid \mathbf{x}_k) \propto \exp\left\{-0.5 \parallel \mathbf{z}_k - H\mathbf{x}_k \parallel_2^2 - \beta \parallel \mathbf{x}_k \parallel_1\right\} \tag{13}$$

with $\beta = 100$.

The performance shown in Figures 3 and 4 demonstrates the viability of the MCMC approach in solving the compressed sensing problem. The estimation errors of both the CSKF and the MCMC nearly coincide from a certain compressibility level (the increased complexity is manifested by the positive slope of the estimation error lines). The estimation accuracy of a conventional Kalman filter is also depicted showing its uselessness for such a problem.

6. Conclusions

This paper presents a Markov chain Monte Carlo approach for high dimensional nonlinear filtering. The new algorithm utilises an improved proposal distribution that essentially incorporates the latest measurement and subgradient information of the underlying likelihood function. This proposal is then used for generating candidate moves in high probability regions of the sample space. The alternating MCMC scheme with the steering distribution accelerates the generation of new samples and has a much higher success rate than the MCMC filter with nonalternating sampling rate. The subgradient MCMC algorithm is shown to outperform the block PF [8] which subdivides the whole state vector into blocks with smaller dimensions and calculates individual likelihoods for each block. The numerical study demonstrates the potential of the new filtering scheme for high dimensional nonlinear state estimation.

Acknowledgements. We acknowledge the support from the UK Engineering and Physical Sciences Research Council (EPSRC) via the Bayesian Tracking and Reasoning over Time (BTaRoT) grant EP/K021516/1 and EC Seventh Framework Programme [FP7 2013-2017] TRAcking in compleX sensor systems (TRAX) Grant agreement no.: 607400.

References

[1] L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. K. Pang, S. Godsill, Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking, Digital Signal Processing 25 (1) (2014) 1 – 16.

[2] F. Daum, J. Huang, Curse of Dimensionality and Particle Filters, in: IEEE Aerospace Conf., 2003, pp. 1979–1993.

[3] C. Andrieu, A. Doucet, R. Holenstein, Particle Markov chain Monte Carlo methods, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74 (3) (2010) 269–342.

[4] C. Berzuini, G. Nicola, W. R. Gilks, C. Larizza, Dynamic Conditional Independence Models and Markov Chain Monte Carlo Methods, J. Amer. Statist. Assoc. 92 (440) (1997) 1403–1412.

[5] F. Daum, J. Huang, Particle flow for nonlinear filters with log-homotopy, in: SPIE conference on Signal and Data Processing of Small Targets, 2008.

[6] P. Djuric, M. F. Bugallo, Particle filtering for high-dimensional systems, in: Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2013, pp. 352–355.

[7] L. Mihaylova, A. Hegyi, A. Gning, R. Boel, Parallelized particle and Gaussian sum particle filters for large scale freeway traffic systems, IEEE Transactions on Intelligent Transportation Systems 13 (1) (2012) 36–48.

[8] P. Rebeschini, R. van Handel, Can local particle filters beat the curse of dimensionality?, Annals of Applied Probability, 2015, 25 (5) (2015) 2809–2866.

[9] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. on Signal Proc. 50 (2) (2002) 174–188.

[10] S. K. Pang, J. Li, S. J. Godsill, Detection and tracking of coordinated groups, IEEE Transactions on Aerospace and Electronic Systems 47 (1) (2011) 472 – 502.

[11] A. Carmi, F. Septier, S. J. Godsill, The Gaussian mixture MCMC particle algorithm for dynamic cluster tracking, Automatica 48 (2012) 2454–2467.

[12] N. Z. Shor, Minimization Methods for Non-differentiable Functions, Springer-Verlag, 1985.

[13] A. Carmi, P. Gurfil, D. Kanevsky, Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms, IEEE Transactions on Signal Processing 58 (4) (2010) 2405–2409.

[14] J. Liu, R. Chen, Sequential Monte Carlo methods for dynamic systems, Journal of the American Statist. Association 93 (443) (1998) 1032–1044.

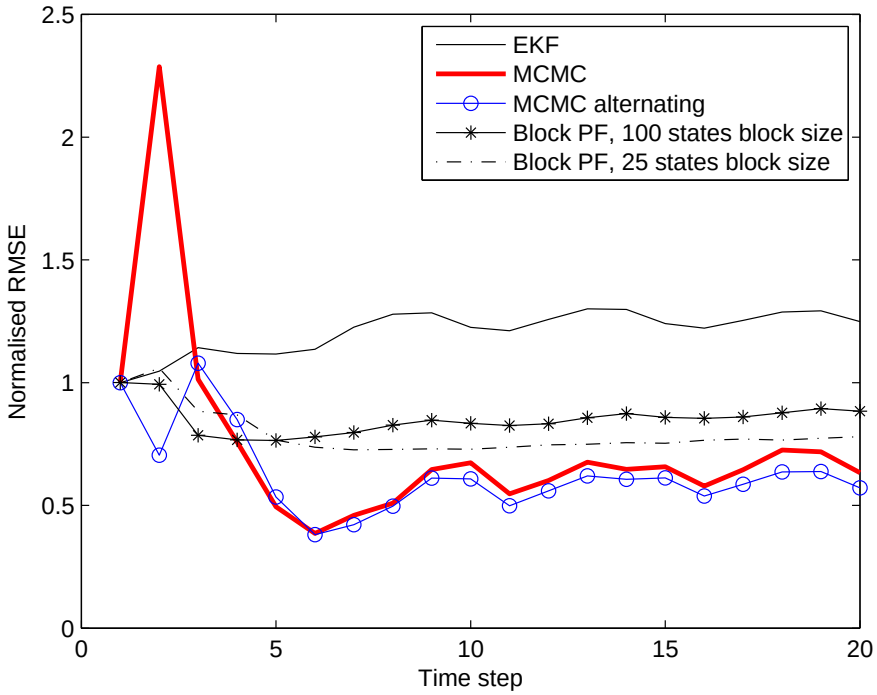

Figure 1. The RMSE based on 50 Monte Carlo runs, 3000 particles of the subgradient-based MCMC algorithm with Metropolis Hastings step compared with the EKF and block PF. Nonlinear system with 100 states.

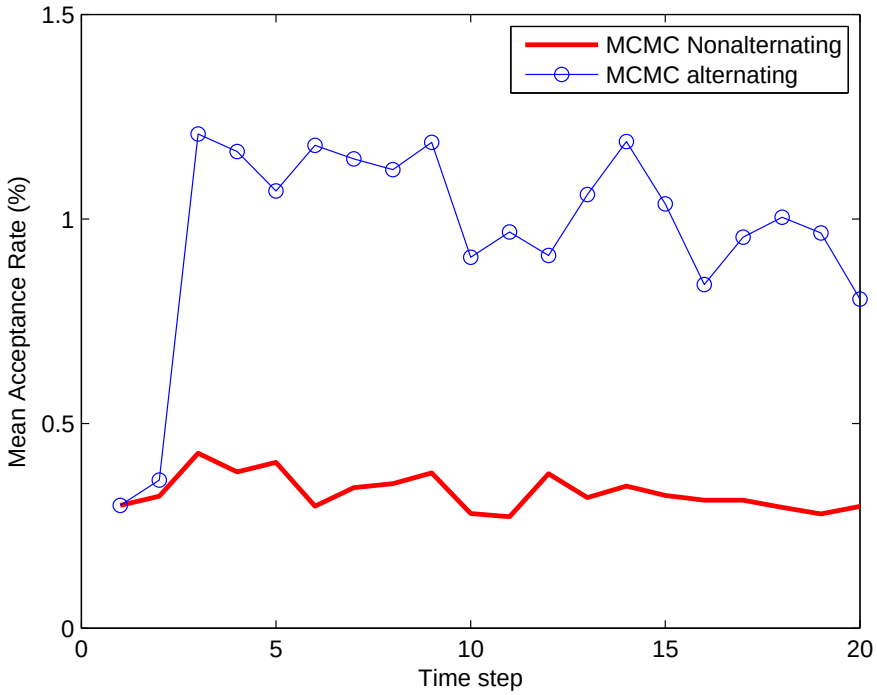

Figure 2. Mean acceptance rate computed based on 50 Monte Carlo runs of the subgradient-based MCMC algorithm with Metropolis Hastings step. Nonlinear system with 100 states.

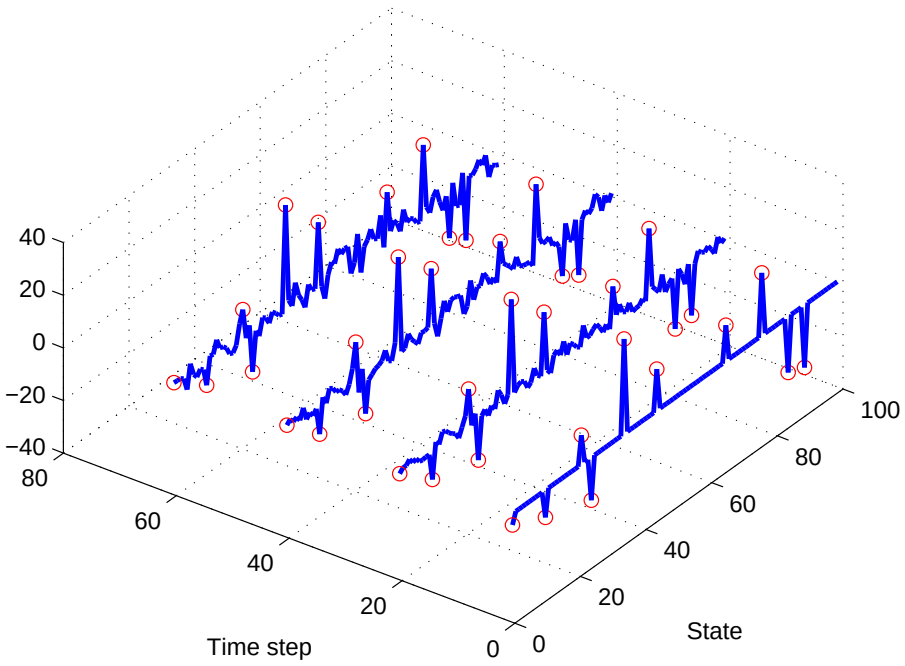

Figure 3. Dynamic compressed sensing example. A typical signal realization is shown over time. System with 100 states.

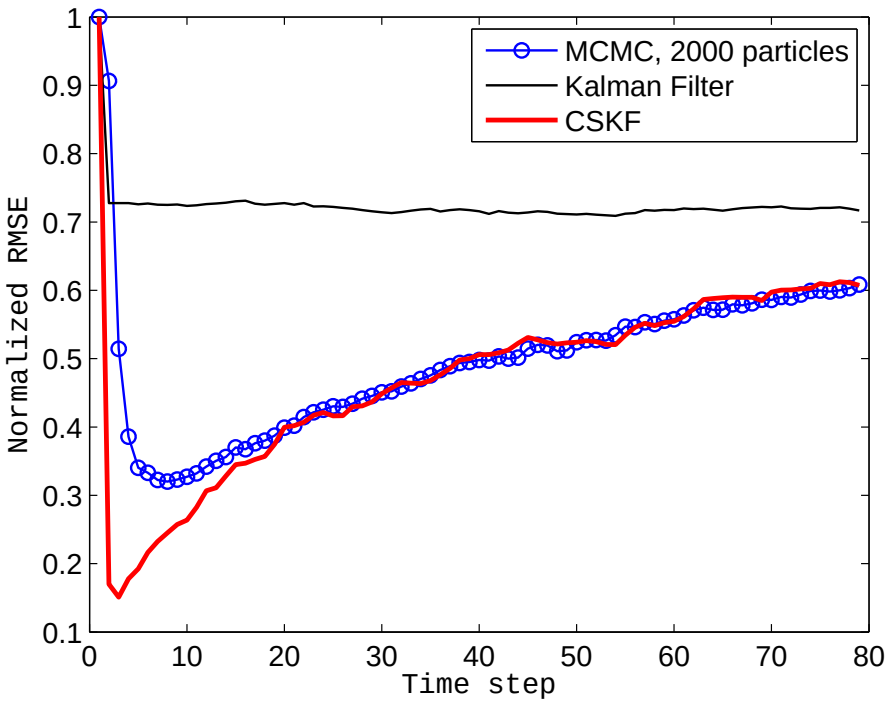

Figure 4. The normalised RMSEs of the Kalman filter, the CSKF, and the MCMC are shown in the bottom panel. System with 100 states.