CONVERGENCE ANALYSIS FOR SEQUENTIAL MONTE CARLO RECEIVERS IN COMMUNICATIONS APPLICATIONS

Soner Özgür, Volkan Cevher, Douglas B. Williams, James H. McClellan

School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, Georgia 30332–0250
E-mail: {ozgur, volkan, dbw, jim.mcclellan}@ece.gatech.edu

ABSTRACT

Recently, sequential Monte Carlo methods have been applied in the telecommunications field and found application in receiver design. The properties of these receivers make these design approaches very attractive. These receivers do not require channel state information or training. Therefore, they are bandwidth efficient and no communication bandwidth needs to be wasted on training. The receivers are optimal in the sense that they achieve minimum symbol error rate regardless of the noise distribution, nonlinearities in the system, and distribution of the transmitted symbol. Moreover, these receivers are capable of producing soft-information outputs, which enables the designer to utilize iterative receiver architectures for near-optimal performance. In this work we investigate the convergence properties of these algorithms when utilized in various types of receivers. We quantify the convergence rate. We describe how various parameters (e.g. noise power, channel fading rate, etc) and factors (e.g. state-space model mismatch) affect the convergence rate and point out the factors that should be improved first to gain speed and accuracy in the convergence.

1. INTRODUCTION

Monte Carlo methods allow sampling from complex, intractable distributions by simulating a Markov chain, whose invariant distribution approaches the desired *stationary* distribution almost surely [1]. In Monte Carlo terms, distributions are represented by discrete state realizations, called particles, that are distributed according to the underlying distribution after convergence. Sequential Monte Carlo (SMC) methods, also known as particle filters, employ a different approach to sample recursively from dynamically varying distributions: they reuse the current particle support to help reconstruct the new particles needed to represent the evolving system.

There are many factors that affect the performance of SMC algorithms. The choice for the number of particles is a trade-off between computational complexity and estimation bias [2]. A similar trade-off exists in integer implementations where likelihood values have to be quantized. One must determine whether a finer quantization or more particles result in better algorithm performance for the same computational power. The choice of proposal function poses a trade-off between computational complexity and estimation accuracy [3]. The use of alternative resampling strategies may impact the convergence of the algorithm. The algorithm performance is also affected by system model parameters. For example, in a communications application, the amount of transmitted power affects the observation signal to noise ratio (SNR). In such cases, the engineer may decide to use fewer particles in high SNR cases. The goal is to choose the design parameters so that the algorithm gives acceptable performance for reasonable computational requirements.

We propose to evaluate the particle filter convergence rate and quantify the effect of system parameters. In communications systems, the transmitted symbols usually are from a discrete and finite space. Therefore, the approach is similar to convergence analysis methods that have been applied to discrete Markov chains [4]. We relate the convergence speed of the chain to the magnitude of the second largest eigenvalue of the Markov chain transition probability matrix. It is possible to compare the effectiveness of different proposal functions by monitoring the temporal evolution of this eigenvalue.

2. CONVERGENCE OF SEQUENTIAL MONTE CARLO ALGORITHM

The trajectory of each particle in SMC algorithms can be viewed as a realization of a Markov chain. At every iteration of the SMC algorithm, the particles are moved to a new position by the proposal distribution. The proposal distribution constitutes a Markov transition kernel. As the algorithm iterates, the trajectory of each particle reflects the state changes of the system. However, observation noise, model mismatches, and other factors cause deviations from the ideal trajectory [1]. For the algorithm to continue tracking a dynamic state space, the Markov chain should quickly converge around the mode of the posterior. When this convergence rate is slow, the resulting estimates are inaccurate.

Non-homogeneous Markov chains do not have an invariant state. However, the transmitted symbols in communications applications are from a discrete and finite state. We propose to relabel the states so that one state represents the "correct" transmitted symbol and the others the "wrong" ones. By applying this procedure, we can convert a non-homogeneous Markov chain to a homogeneous Markov chain. In the sections below, we repeat the sufficient conditions for convergence of homogeneous Markov chains for clarity reasons. We also propose a method to construct an estimate of the probability transition matrix of the chain using some intermediate results obtained by the running SMC algorithm.

2.1. Basics of Markov Chains

Let us denote a trajectory with $\{\mathbf{x}_0^{(i)}, \mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_N^{(i)}\}$, which is also a realization of a Markov chain defined on a state space E [2, 4]. Let $\mathcal{E}$ be a countably generated σ-algebra on E. Let $P(A|\mathbf{x}) =$

$P(\mathbf{x}_{n+1}^{(i)} \in A|\mathbf{x}_n^{(i)} = \mathbf{x})$ for all $A \in \mathcal{E}$ and $\mathbf{x} \in E$, which is the probability of transition from state $\mathbf{x}$ to state A ($\mathbf{x}$ is in the state space and A is in the σ-algebra). Let also $P^n(A|\mathbf{x})$ denote a n-step transition function $P^n(A|\mathbf{x}) = P(\mathbf{x}_n^{(i)} \in A|\mathbf{x}_0^{(i)} = \mathbf{x})$.

The property of irreducibility is a first measure of the sensitivity of the Markov chain to the initial conditions. Irreducibility is crucial because it leads to a guarantee of convergence, without detailed study of the transition function.

Basically, a chain is irreducible if all of its states communicate [5]. The formal definition is a little more involved. Given a measure ϕ, the Markov chain $\mathbf{x}_n$ with transition kernel $P(A|\mathbf{x})$ is *ϕ-irreducible* if, for every $A \in \mathcal{E}$ with $\phi(A) > 0$ there exists n such that $P^n(A|\mathbf{x}) > 0$ for all $\mathbf{x} \in E$.

Another property of the Markov chain that must be specified is its aperiodicity. A period of a state $\mathbf{x} \in E$ is defined as

$$d(\mathbf{x}) = \text{g.c.d}\{n \geq 1; P^n(\mathbf{x}, \mathbf{x}) > 0\} \tag{1}$$

The value of the period for all states that communicate with $\mathbf{x}$ is the same. If the chain is irreducible, the period is the same for all states. If the period is one for the state $\mathbf{x}$ and the chain is irreducible, then the chain is called aperiodic.

Here, we introduce the L^1-distance between two measures P and Q, defined as

$$\begin{aligned} \|P - Q\|_{\text{var}} &= \sup_{A \in \mathcal{E}} |P(A) - Q(A)| \\ &= \frac{1}{2} \sum_{\mathbf{x} \in E} |P(\mathbf{x}) - Q(\mathbf{x})| \equiv \frac{1}{2} \|P - Q\|_{L^1} \end{aligned} \tag{2}$$

When the two measures have density functions $p(\mathbf{x})$ and $q(\mathbf{x})$, the distance can be written as

$$\|P - Q\|_{\text{var}} = \frac{1}{2} \int |P(\mathbf{x}) - Q(\mathbf{x})| \text{d}\mathbf{x} \tag{3}$$

Theorem 2.1. *[6]: Suppose the state space* E *of a Markov chain is finite. The transition function of this chain is irreducible and aperiodic; then, $P^n(A|\mathbf{x}) = P^n(\mathbf{x}_n^{(i)} \in A|\mathbf{x}_0^{(i)} = \mathbf{x})$ as a probability measure on A converges to its invariant distribution $\pi(\cdot)$ geometrically in variation distance; namely, there exists a $0 < r < 1$ and $c > 0$ such that*

$$\|P^n(A|\mathbf{x}) - \pi(A)\|_{\text{var}} \leq cr^n \tag{4}$$

Theorem 2.2. *(Tierney [6]) : Suppose A is π-irreducible and $\pi A = \pi$. Then A is positive recurrent and π is the unique invariant distribution of A. If A is also aperiodic, then, for almost all $\mathbf{x}$ (for all $\mathbf{x}$ except a subset whose measure under π is zero),*

$$\|P^n(A|\cdot) - \pi(\cdot)\|_{\text{var}} \to 0, \tag{5}$$

where $\|\cdot\|_{\text{var}}$ denote the total variation distance.

The above two theorems prove that all Markov chains that are irreducible and aperiodic converge geometrically to their invariant distribution. This convergence proof is achieved without the explicit analysis of the transition kernel.

2.2. Convergence Rate

Consider a Markov chain on a finite discrete state space with κ states. We define a state transition probability matrix $\mathbf{P}$ whose entry p_{ij} indicates the transition probability from state i to state j. By definition, $\sum_j p_{ij} = 1$ for all i. Hence, $\mathbf{P}$ has an eigenvalue equal to 1.

Consider an arbitrary function g on the state-space E. Because the state-space is discrete, the function can be represented by a vector $\mathbf{g}_0$. The vector after the state transition is $\mathbf{g}_1 = \mathbf{P}\mathbf{g}_0$, which is simply the conditional expectation function:

$$g_1(\mathbf{x}) = E\{g(\mathbf{x}_1)|\mathbf{x}_0 = \mathbf{x}\} \tag{6}$$

We also note that

$$\text{var}\{E\{g_1(\mathbf{x})\}\} \leq \text{var}\{g(\mathbf{x}_1)\}. \tag{7}$$

Hence, the eigenvalues of the matrix $\mathbf{P}$ must be less than or equal to 1 in absolute value [3, 5]. Moreover, Berrman and Plemmons show that all of $\mathbf{P}$'s eigenvalues are real and can be diagonalized [7]

$$\mathbf{P} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}^{-1}, \tag{8}$$

where $\mathbf{\Lambda} = diag([1, \lambda_2, \ldots, \lambda_\kappa])$, with the eigenvalues sorted in descending order of their absolute value, $1 \geq |\lambda_2| \geq \ldots |\lambda_\kappa|$. Therefore, as $n \to \infty$,

$$\mathbf{P}^n = \mathbf{B}\mathbf{\Lambda}^n\mathbf{B}^{-1} \overset{n\to\infty}{\longrightarrow} \mathbf{B} \begin{bmatrix} 1 & 0 & & 0 \\ 0 & 0 & & 0 \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix} \mathbf{B}^{-1} \tag{9}$$

The above is true if and only if $|\lambda_2| < 1$. Because $\pi\mathbf{P}^n = \pi$ and the limit of $\mathbf{P}^n$ is of rank 1, every row of $\mathbf{P}^\infty$ must be the same as π.

3. CONVERGENCE ANALYSIS OF AN SMC BLIND EQUALIZER FOR FREQUENCY SELECTIVE FADING CHANNELS

Details of receiver design and computational complexity improvements are thoroughly discussed in [8].

3.1. System Model

Consider a communications system that transmits a sequence of symbols $\mathbf{s}_{0:n} = \{s_0, s_1, \ldots, s_n\}$ randomly chosen from some constellation $\mathcal{A}$. At the receiver end, the receiver observes the signal $\mathbf{r}_{0:n}$ which is a random process that depends on the symbols $\mathbf{s}_{0:n}$ and channel parameters $\mathbf{h}_{0:n}$. Typically the channel state information $\mathbf{h}_{0:n}$ is a multidimensional continuous-valued quantity. The system can be described with the following state-space representation:

$$\begin{aligned} \mathbf{h}_n &\sim q_1(\mathbf{h}_n|\mathbf{h}_{0:n-1}) \\ s_n &\sim q_2(s_n|s_{n-1}) \\ \mathbf{r}_n &\sim f(\mathbf{r}_n|\mathbf{s}_{0:n}, \mathbf{h}_{0:n}). \end{aligned} \tag{10}$$

Based on the state-space representation, we can formulate two inference problems:

- Symbol Sequence Detection: Assuming the channel $\mathbf{h}_{0:n}$ is known, find

$$\hat{\mathbf{s}}_{0:n} = \arg \max_{\mathbf{s}_{0:n} \in \mathcal{A}^{n+1}} p(\mathbf{s}_{0:n}|\mathbf{r}_{0:n}, \mathbf{h}_{0:n}) \tag{11}$$

- Channel Estimation: Assuming the transmitted symbols $\mathbf{s}_{0:n}$ are known, find

$$\hat{\mathbf{h}}_{0:n} = \arg\max_{\mathbf{h}_{0:n}} p(\mathbf{h}_{0:n}|\mathbf{r}_{0:n}, \mathbf{s}_{0:n}) \tag{12}$$

The channel is usually not known and needs to be estimated. For this purpose either a training sequence of symbols is required or a blind approach must be utilized. Using Monte Carlo processing, the channel parameters can be marginalized out

$$\hat{\mathbf{s}}_{0:n} = \arg\max_{\mathbf{s}_{0:n}\in\mathcal{A}^{n+1}} E\{p(\mathbf{s}_{0:n}|\mathbf{r}_{0:n}, \mathbf{h}_{0:n})|\mathbf{h}_{0:n}\}, \tag{13}$$

where the expectation is taken with respect to the distribution of $\mathbf{h}_{0:n}$.

We represent the posterior with a discretized approximation. We obtain a set of samples and assign weights to it. The notation $\{\mathbf{s}_n^{(i)}, w^{(i)}\}_{i=0}^{N_s}$ is called a random measure that characterizes the posterior distribution $p(\mathbf{s}_{0:n}|\mathbf{r}_{0:n})$. A maximum a posteriori estimate (MAP) is calculated as follows:

$$\hat{s}_n = \arg\max_{s_n\in\{\mathcal{A}\}} \sum_{i=1}^{N_s} w^{(i)}\delta(s_n - s_n^{(i)}), \tag{14}$$

Table 1. **Blind equalizer using nested SMC**

- Initialization
- for $i = 1, \ldots, N_s$ /*for each particle*/
  - Draw $\tilde{s}_{n+k}^{(i)} \sim \beta(\cdot)$ as described in Table 2
  - Update

$$\mathbf{h}_n^{(i)} = \mathbf{h}_{n-1}^{(i)} + \mathbf{K}_n^{(i)}(\mathbf{r}_n - \mu_n^{(i)})$$

$$\mathbf{\Sigma}_n^{(i)} = (1 - K_n^{(i)})\mathbf{s}_n^{(i)^T}\mathbf{\Sigma}_{n-1}^{(i)},$$

    where the Kalman gain is:

$$K_n^{(i)} = \mathbf{\Sigma}_{n-1}^{(i)}\mathbf{s}_n^{(i)}(P_n^{(i)})^{-1}$$

  - Update weight $w_n^{(i)} \propto w_{n-1}^{(i)}\beta_{n:n+L-1}^{(i)}(\cdot)$
- Obtain symbol estimate as

$$\hat{s}_n = \arg\max_{s_n\in\{\pm 1\}} \sum_{i=1}^{N_s} w^{(i)}\delta(s_n - s_n^{(i)}),$$

- Resample if needed
- Perturb channel particles and update based on likelihoods.

3.2. Convergence

We utilize Rao-Blackwellization in the algorithm, which means that we proceed with analytical solutions rather than resorting

Table 2. **Symbol draw using nested SMC**

- Set $N_{partilces} = 1$
- for k=0,...,m-1 /*start with first sub-symbol*/
  - Compute

$$\mathbf{h}_{n+k} = \zeta\mathbf{h}_{n-1+k} + (1-\zeta)\eta_{n-1} \tag{15}$$

$$\mu_{n+k}^{(i)} = \mathbf{s}_{n+k}^{(i)^T}\mathbf{h}_{n+k}^{(i)}$$

$$P_{n+k}^{(i)} = \sigma^2 + \mathbf{s}_{n+k}^{(i)^T}\mathbf{\Sigma}_{n+k}^{(i)}\mathbf{s}_{n+k}^{(i)}$$

    ,for each possible $s_{n+k}^{(i)} \in \{\pm 1\}$
  - Compute $p(r_{n+k}|\mathbf{h}_{n+k}^{(i)}, \mathbf{s}_{n+k}) = -(r_{n+k} - \mu_{n+k}^{(i)})^2/P_{n+k}^{(i)}$
  - For $j = 1, \ldots, N_{particles}$
    - Expand each trajectory to $|\{\pm 1\}| = 2$ (alphabet_size) distinct paths

$$\beta^{(2*j)}(r_{n+k}|\mathbf{h}_{n+k}^{(i)}, \mathbf{s}_{n+k}) = \beta^{(j)}(r_{n+k-1}|\mathbf{h}_{n+k-1}^{(i)}, \mathbf{s}_{n+k-1})\, p(r_{n+k}|\mathbf{h}_{n+k}^{(i)}, \mathbf{s}_{n+k-1}, s_{n+k} = -1)$$

$$\beta^{(2*j+1)}(r_{n+k}|\mathbf{h}_{n+k}^{(i)}, \mathbf{s}_{n+k}) = \beta^{(j)}(r_{n+k-1}|\mathbf{h}_{n+k-1}^{(i)}, \mathbf{s}_{n+k-1})\, p(r_{n+k}|\mathbf{h}_{n+k}^{(i)}, \mathbf{s}_{n+k-1}, s_{n+k} = +1)$$

  - Update $N_{particles} = N_{particles} * alphabet\_size$
  - if $N_{particles} > N_{partilces}\_max$ Truncate the trajectory list to $N_{partilces}\_max$ and set $N_{particles} = N_{partilces}\_max$. Use residual resampling when choosing the subset to be retained.
- Choose $\tilde{s}_n^{(i)}$ based on the trajectory with highest likelihood.

to Monte Carlo whenever possible and feasible. This results in improved efficiency and convergence. In this case we Rao-Blackwellize the channel coefficient part of the problem. As it is easily seen the results resemble Kalman filtering equations [8]. The symbol estimation part, however, is a non-Gaussian problem. The symbol distribution is non-Gaussian. Therefore, we retain Monte Carlo methodology. The symbols are drawn from the posterior distribution that can be analytically computed, which also results in increased convergence and efficiency.

The properties of the transition distribution $p(\mathbf{s}_{0:n}|\mathbf{s}_{0:n-1}, \mathbf{r}_{0:n})$ determine the convergence rate of the algorithm. Let us factorize the posterior distribution:

$$p(\mathbf{s}_{0:n}|\mathbf{r}_{0:n}) = \frac{p(\mathbf{r}_n|\mathbf{s}_{0:n}, \mathbf{r}_{0:n-1})p(\mathbf{s}_n|\mathbf{s}_{0:n-1}, \mathbf{r}_{0:n-1})}{p(\mathbf{r}_n)} \\ p(\mathbf{s}_{0:n-1}|\mathbf{r}_{0:n-1}) \tag{16}$$

The transition kernel turns out to be

$$p(\mathbf{s}_{0:n}|\mathbf{s}_{0:n-1}, \mathbf{r}_{0:n}) = \frac{p(\mathbf{s}_{0:n}|\mathbf{r}_{0:n})}{p(\mathbf{s}_{0:n-1}|\mathbf{r}_{0:n-1})} \\ \propto p(\mathbf{r}_n|\mathbf{s}_{0:n}, \mathbf{r}_{0:n-1}). \tag{17}$$

The second term in the numerator in (16) gets absorbed in the proportionality because the transmitted symbols are independent and uniformly distributed. The denominator in (16) is just a proportionality constant.

The transition kernel turns out to be the incremental weight update factor, when the posterior distribution is used for the symbol draws.

To analyze the convergence we construct an estimate of the transition probability matrix. We do this by computing the weighted average across all particles (18). Since our transmitted symbols are from the set $\{\pm 1\}$, we have a 2×2 probability transition matrix. The size of the matrix grows proportionally, for problems of higher dimesionality. Each element contains the following transition probability:

$$p(\mathbf{s}_n|\mathbf{s}_{n-1}, \mathbf{r}_n) = \sum_{i=1}^{N_s} p(\mathbf{s}_n^{(i)}|\mathbf{s}_{n-1}^{(i)}, \mathbf{r}_n) w_n^{(i)} \tag{18}$$

As was described in the previous section, the second eigenvalue of this matrix determines the rate of convergence.

4. NUMERICAL RESULTS

In our simulations we use frequency selective channels with $d = 3$ and 4 channel coefficients. We implement the SMC algorithm with temporal partitioning of the symbol space [8]. The transmitted symbols are BPSK. The receiver noise is additive white Gaussian noise.

We display the distribution of the second largest eigenvalue with a box-plot and with a plot of the cumulative distribution function. In the box-plot, the horizontal axis indicates the SNR, and the vertical axis is the magnitude of the second largest eigenvalue. The eigenvalue is obtained from the simulations and depends on the particular realizations of the channel, received signal and transmitted data. The line in the middle of the box indicates the median value for the collected samples of λ_2. The upper and the lower edge of the box indicate the upper and the lower quartile of the data. The vertical lines, extending from each end of the boxes, show the extent of the rest of the data. The horizontal lines at the ends of the extending vertical lines indicate an interval of 1.5 times the interquartile range. The red + markers indicate outliers.

As expected, the median of λ_2 decreases as we increase the SNR as seen in the box-plots in Figures 1 and 3. Similar observations can be made about the cdf plots as well (Figures 2 and 4).

Comparing Figure 2 to Figure 4, one may conclude more particles are needed to track a 4 dimensional channel state vector relative to the $d = 3$ case.

There is another interesting observation that can only be seen in the cdf plot, but not in the box-plot (Figures 6 and 7). For a fixed SNR, as the particle count for the channel state is increased, the eigenvalue distribution tends to get bi-modal. At a low number of channel particles, there tends to be a peak close to $|\lambda_2| = 1$ (look at the slope). This is an indication of an estimation bias. When the number of particles is insufficient, the wrong transmitted symbol is selected.

Similar conclusions can be drawn from the results of the fading channel case.The ability of the SMC algorithm to track time-varying channels is illustrated in Figure 5. The channel is a slowly fading Rayleigh channel with Doppler coefficient $f_m T = 10^{-3}$. The channel coefficients are symbol spaced and mutually independent. In all simulations there were used 300 particles.

Figures 1 and 2 display the box-plot and the cdf of $|\lambda_2|$ for the 3-tap fading channel. As SNR increases, the eigenvalue gets smaller, which indicates that the number of used particles to estimate the channel may be reduced at high SNR. Similar behavior is observed at the channel with delay spread of 4 symbol intervals (Figures 3 and 4).

The dependence between the second largest eigenvalue and used particles is illustrated in Figures 6 and 7. The bias as the particle number decreases is evident.

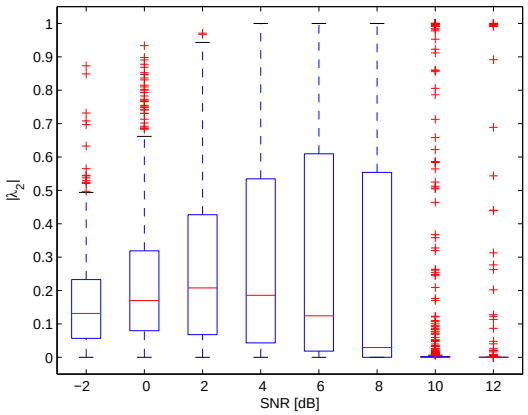

Fig. 1. Box-plot of convergence rate for temporally partitioned SMC on a frequency selective fading channel with $d = 3$ taps, 300 channel particles and a Doppler shift of $f_m T = 10^{-3}$.

5. REFERENCES

[1] J. S. Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, Sep. 1998.

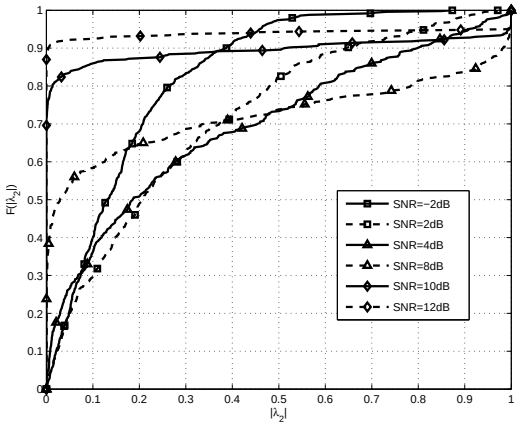

Fig. 2. CDF of convergence rate for temporally partitioned SMC on a frequency selective fading channel with $d = 3$ taps, 300 channel particles and a Doppler shift of $f_m T = 10^{-3}$.

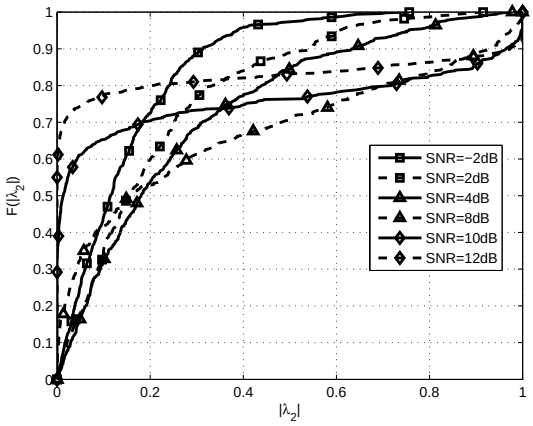

Fig. 4. CDF of convergence rate for temporally partitioned SMC on a frequency selective fading channel with $d = 4$ taps, 300 channel particles and a Doppler shift of $f_m T = 10^{-3}$.

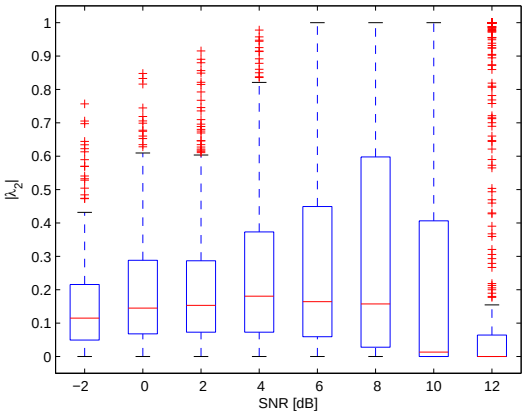

Fig. 3. Box-plot of convergence rate for temporally partitioned SMC on a frequency selective fading channel with $d = 4$ taps, 300 channel particles and a Doppler shift of $f_m T = 10^{-3}$.

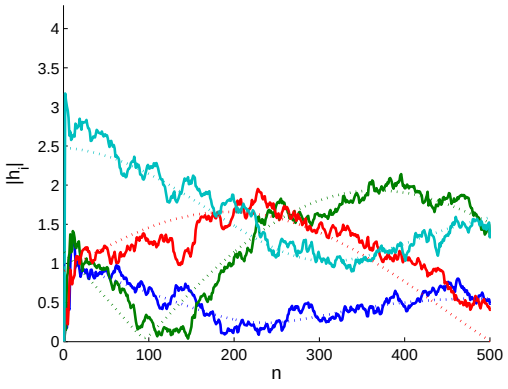

Fig. 5. Channel coefficient magnitude estimates. Random initialization. Rayleigh fading frequency selective channel with $d = 4$ taps, SNR=$10dB$, $fmT = 1e - 3$.

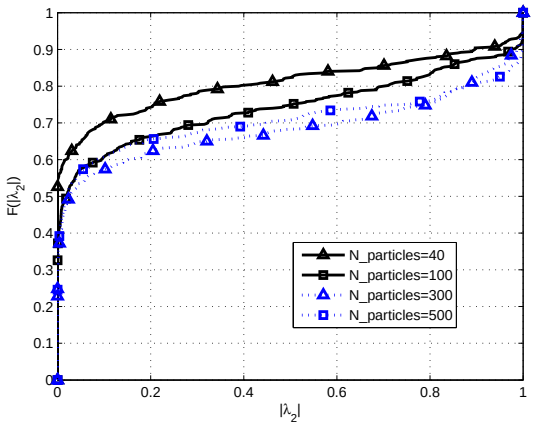

Fig. 6. CDF of convergence rate for temporally partitioned SMC on a frequency selective fading channel with $d = 3$ taps, $SNR = 8dB$, a Doppler shift of $f_m T = 10^{-3}$ and varying number of channel particles.

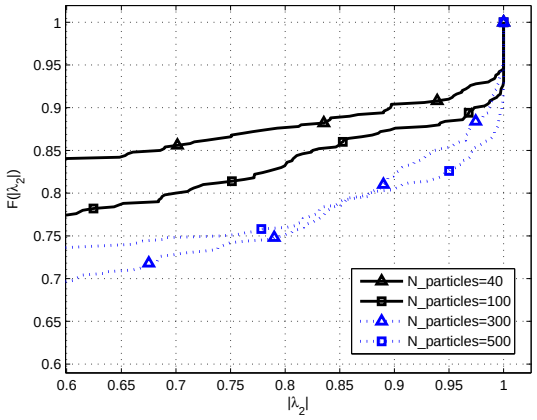

Fig. 7. CDF of convergence rate for temporally partitioned SMC on a frequency selective fading channel with $d = 3$ taps, $SNR = 8dB$, a Doppler shift of $f_m T = 10^{-3}$ and varying number of channel particles. The display is zoomed at the second mode.

[2] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag, New York, 2001.

[3] M. A. Tanner, *Tools for Statistical Inference*, Springer-Verlag, New York, 1991.

[4] R. Chen, J. S. Liu, and X. Wang, "Convergence analyses and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, February 2002.

[5] G. Casella C. P. Robert, *Monte Carlo Statistical Methods*, Springer-Verlag, New York, 2004.

[6] L. TIerney, "Markov chains for exploring posterior distributions," *Annals of Statistics*, vol. 22, pp. 1701–1728, 1994.

[7] A. Berrmon and R. J. Plemmons, "Non-negative matrices in the mathematical sciences," *Classics in applied mathematics 9, Society for industrial and applied mathematics*, 1994.

[8] S. Ozgur and D. B. Williams, "Temporal partition particle filtering for multiuser detectors with mutually orthogonal sequences," *ICASSP 2005*, Mar. 2005.