Advanced Signal Processing Techniques

For Wireless Communications

An application:

Blind Phase Noise Estimation and Data Detection

E. Panayırcı
Kadir Has University
Department of Electronics Engineering,
Istanbul, Turkey

Outline

- Introduction
- Key Knowledge Gaps
- EM and SAGE Algorithms
- Sequential Monte Carlo (SMC) Method
- An example: Blind Phase Noise Estimation and Data Detection with SMC technique
- Simulation Results
- Conclusions

Introduction

Future generation wireless communication systems are confronted with new challenges mainly due to

- Hostile channel characteristics
- Limited bandwidth
- Very high data rates

Introduction (Cont.)

Advanced Statistical Signal Processing techniques such as

- The Expectation-Maximization (EM) algorithm
- The SAGE algorithm
- Sequential Monte Carlo Techniques
- Kalman filters and their extensions

In collaboration with

- Inexpensive and
- Rapid

computational power provide powerful tools to overcome the limitations of current technologies.

Introduction (Cont.)

Applications of advanced signal processing algorithms, include, but are not limited to

- Joint/Blind/Adaptive sequence (data) detection
- Frequency, phase, timing synchronization
- Equalization
- Channel Estimation techniques.

Introduction (Cont.)

These techniques are employed in advanced wireless communication systems such as

- OFDM/OFDMA
- CDMA/MC-CDMA
- MIMO
- Space-time-frequency Coding
- Multi-User detection

Introduction (Cont.)

Especially, development of the suitable algorithms for wireless multiple access systems in

- Non-stationary
- Interference-rich

environments presents major challenges to us.

Introduction (Cont.)

Optimal solutions for these problems, mostly, can not be implemented in practice mainly due to

$\Rightarrow$ high computational complexity

Introduction (Cont.)

Advanced signal processing tools, I mentioned before, have provided a promising route for the design of low complexity algorithms with performances approaching the theoretical optimum for

- Fast, and
- Reliable

communication in highly severe and dynamic wireless environment

Key Knowledge Gaps

The Key Knowledge Gaps

- Theoretical performance and convergence analysis of these Algorithms
- Some new efficient algorithms need to be worked out and developed
- Computational complexity problems of these algorithms, when applied to on-line implementations in the digital receivers, must be handled
- Also, implementation of these algorithms based on batch processing and sequential (adaptive) processing depending on how the data are processed and the inference is made has not been completely solved.

Key Knowledge Gaps(cont.)

- Some class of algorithms requires efficient generation of random samples from an arbitrary target probability distribution
- So far two basic types of algorithms, *Metropolis algorithm* and *Gibbs sampler* have been widely used in diverse fields.
- But they are very complex and difficult to apply for on-line applications like wireless communications.
- There are gaps for devising new types of more efficient algorithms that can be effectively employed in wireless applications.

The EM Algorithm

The Expectation-Maximization (EM) Algorithm

- The EM algorithm was popularized in 1977
- An iterative algorithm for obtaining ML parameter estimates
- Based on definition of *complete* and *incomplete data*

The EM Algorithm (cont.)

Main References:

- L. E. Baum, T. Petrie, G. Soules and N. Weiss, A Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Chains, *Annals of Mathematical Statistics*, pp. 164-171, 970.
- A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum-Likelihood from Incomplete Data Via the EM Algorithm, em Journal, Royal Statistical Society, Vol. 39, pp. 1-17, 1977.
- C. F. Wu, On the Convergence Properties of the EM Algorithm, *Annals of Statistics*, Vol. 11, pp. 95-103, 1983.

The EM Algorithm (cont.)

The Essential EM Algorithm

Consider estimating parameter vector $\mathbf{s}$ from data $\mathbf{y}$ (incomplete data):

$$\mathbf{y} = F(\mathbf{z}, \mathbf{s}) + \mathbf{n}$$

where

$\mathbf{s}$ is the parameter vector to be estimated, and
$\mathbf{z}$ is the random nuisance parameter vector.

The EM Algorithm (cont.)

Then, the ML estimate of $\mathbf{s}$ is:

$$\widehat{\mathbf{s}}_{ML} = \arg\max_{\mathbf{s}} p(\mathbf{y}|\mathbf{s}) = E_{\mathbf{z}}[p(\mathbf{y}|\mathbf{s},\mathbf{z})]$$

Thus, obtaining ML estimates may require:

An Expectation $\Rightarrow$ Often analytically intractable

A Maximization $\Rightarrow$ Computationally intensive

The EM Algorithm (cont.)

The EM Iteration

Define the complete data $\mathbf{x}$

$\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x}) \Rightarrow$ Many-to-one mapping

Having the conditional density $f(\mathbf{x}|\mathbf{s})$

The EM iteration at the i-th step:

E-Step: $Q(\mathbf{s}|\widehat{\mathbf{s}}^i) \equiv E\left[\log f(\mathbf{x}|\mathbf{s})|\mathbf{y}, \widehat{s}^i)\right]$

M-Step: $\widehat{\mathbf{s}}^{i+1} = \arg\max_{\mathbf{s}} Q(\mathbf{s}|\widehat{\mathbf{s}}^i)$

The EM Algorithm (cont.)

Convergence Properties

- At each iteration the likelihood-function is monotonically non-decreasing
- If the likelihood-function is bounded, then the algorithm converges to the ML estimate

The EM Algorithm (cont.)

The SAGE Algorithm

- The SAGE algorithm is an extension of EM algorithm
- It provides much faster convergence than EM
- Algorithm alternates several hidden data spaces rather than just using one complete data space, and
- Updates only a subset of elements of the parameters in each iteration

The EM Algorithm (cont.)

Some Application Areas

- Positron-Emission-Tomography (PET)
- Genetics
- Neural Networks
- Radar Imaging
- Image / Speech processing Communications
- Channel Estimation / Equalization
- Multiuser detection
- Sequence estimation
- Interference rejection

Sequential Monte Carlo Method

- Emerged in the field of Statistics (J. S. Liu and R. Chen, *Sequential Monte Carlo Methods for Dynamics Systems*, J. American Stat. Assoc., Vol. 93, pp. 1032-1044, 1998)
  - provides a promising paradigm for the design of low-complexity signal processing techniques
  - successfully applied to optimal signal reception problems encountered in wireless communications
    - ◇ blind equalization
    - ◇ blind detection/decoding in fading channels
    - ◇ synchronization
    - ◇ multiuser detection

Sequential Monte Carlo Method (Cont.)

- In communications problems, on-line estimation can be obtained from the marginal posterior distribution
- computation of the marginal distribution involves very high dimensional integration → infeasible in practice
- Monte Carlo methods provide a viable alternative to the required computation
- It is basically based on approximating the expectation operation by means of sequentially generated Monte Carlo samples from either unknown state variables or system parameters
- Main advantages
  - SMC is self adaptive and no training/pilot symbols or decision feedback are needed
  - Channel noise can be either Gaussian on Non-Gaussian

Sequential Monte Carlo Method (Cont.)

- Main advantages:
  - It is suitable for MAP receiver design
  - If the system employs channel coding, the coded signal structure can be easily exploited to improve the accuracy of both channel and data estimation
  - SMC is suitable for high-speed parallel implementation using VLSI
  - Does not require iterations like in EM algorithm
  - Updating with new data can be done more efficiently

Phase Noise Model

- We investigate the use of the SMC method to the problem of jointly detecting the data and estimating the phase noise
- We consider a communication system in the presence of random phase noise and the additive Gaussian noise
- Each data symbol is then transmitted through a channel whose input-output relationship is given by

$$y_t = s_t e^{i\theta_t} + n_t, \qquad t = 0, 1, \cdots$$

where y_t, s_t, θ_t, are the received signal, the transmitted symbols and the phase noise, respectively, and n_t the additive complex Gaussian noise with mean zero and the variance $\sigma_n^2 = E[|n_t|^2]$

Phase Noise Model (Cont.)

- The phase noise process θ_t is defined as a Wiener process determined as

$$\begin{aligned} \theta_t &= \theta_{t-1} + u_t, \qquad t = 1, 2, \cdots \\ \theta_0 &\sim \text{uniform}(-\pi, +\pi) \end{aligned}$$

where $\{u_t\}$ is a sequence of independent and identically distributed (i.i.d.) zero-mean random variables with variance equal to σ_u^2.

- Our main objective is to solve the problem of online detection of the symbols s_t and estimation of the phase noise θ_t, completely blindly, based on the received signals.

Phase Noise Model (Cont.)

- Since we are interested in jointly estimating the symbol s_t and the phase noise θ_t, at time t based on the observation $\mathbf{Y}_t$, from Bayes perspective all necessary information is contained in the joint posterior PDF

$$p(s_t, \theta_t | \mathbf{Y}_t) = \int p(\theta_t | \mathbf{Y}_t, \mathbf{S}_t) p(\mathbf{S}_t | \mathbf{Y}_t) d\mathbf{S}_{t-1}.$$

- Unfortunately, posterior PDF analytically intractable
- Particle filter can be used to approximate the desired posterior PDF
- Since the phase noise process is a Gaussian

$$p(\theta_t | \mathbf{S}_t, \mathbf{Y}_t) \sim N(\mu_{\theta_t}(\mathbf{S}_t), \sigma^2_{\theta_t}(\mathbf{S}_t)), \tag{1}$$

- $\mu_{\theta_t}(\mathbf{S}_t), \sigma^2_{\theta_t}(\mathbf{S}_t))$ can be obtained by a Kalman filter algorithm.

Unscented Filtering (UF)

- UF is a technique for calculating the mean and covariance of a random variable which undergoes a nonlinear transformation.
- Suppose $\mathbf{x}$ is an $n \times 1$ random vector with mean $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance $\mathbf{P}_{\mathbf{xx}}$. $\mathbf{y}$ is related to $\mathbf{x}$ through the nonlinear function

$$\mathbf{y} = \boldsymbol{\psi}(\mathbf{x}).$$

The mean $\boldsymbol{\mu}_{\mathbf{y}}$ and covariance $\mathbf{P}_{\mathbf{yy}}$ of $\mathbf{y}$ can be calculated as follows: The n-dimensional random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance $\mathbf{P}_{\mathbf{xx}}$ is approximated by $2n+1$ points $\boldsymbol{\chi}_i$, called the *sigma-points*, and the weights W_i given by

$$\begin{aligned}
\boldsymbol{\chi}_0 &= \boldsymbol{\mu}_{\mathbf{x}}, & W_0 &= \kappa/(n+\kappa) \\
\boldsymbol{\chi}_i &= \boldsymbol{\mu}_{\mathbf{x}} + \left(\sqrt{(n+\kappa)\mathbf{P}_{\mathbf{xx}}}\right)_i, & W_i &= 1/2(n+\kappa) \\
\boldsymbol{\chi}_{i+n} &= \boldsymbol{\mu}_{\mathbf{x}} - \left(\sqrt{(n+\kappa)\mathbf{P}_{\mathbf{xx}}}\right)_i, & W_{i+n} &= 1/2(n+\kappa)
\end{aligned}$$

Unscented Filtering (Cont.)

The transformation procedure is as follows:

1. Obtain the set of transformed sigma-points,

$$\mathcal{Y}_{\mathbf{i}} = \boldsymbol{\psi}(\boldsymbol{\chi}_i), \qquad i = 0, 1, \cdots, 2n$$

2. Compute the mean given by the weighted average of the transformed points

$$\boldsymbol{\mu}_{\mathbf{y}} = \sum_{i=0}^{2n} W_i \mathcal{Y}_{\mathbf{i}}$$

3. Compute the covariance by the weighted outer product of the transform points,

$$\mathbf{P}_{\mathbf{yy}} = \sum_{i=0}^{2n} W_i (\mathcal{Y}_{\mathbf{i}} - \boldsymbol{\mu}_{\mathbf{y}})(\mathcal{Y}_{\mathbf{i}} - \boldsymbol{\mu}_{\mathbf{y}})^{\dagger}$$

Tracking of Phase Noise with UF

- Since the phase noise process is a Gaussian

$$p(\theta_t|\mathbf{S}_t,\mathbf{Y}_t) \sim N(\mu_{\theta_t}(\mathbf{S}_t), \sigma^2_{\theta_t}(\mathbf{S}_t)), \quad (2)$$

- Denoting $\widehat{\theta}_{t|t-1}$ as the estimator of θ_t based on the observations $\mathbf{Y}_{t-1} = (y_0, y_1 \cdots, y_{t-1})$, the mean $\mu_{\theta_t}(\mathbf{S}_t)$ and the variance $\sigma^2_{\theta_t}(\mathbf{S}_t)$:

$$\mu_{\theta_t}(\mathbf{S}_t) \stackrel{\triangle}{=} \widehat{\theta}_{t|t} \quad \text{and} \quad \sigma^2_{\theta_t}(\mathbf{S}_t) \stackrel{\triangle}{=} M_{t|t},$$

$\widehat{\theta}_{t|t}$ and $M_{t|t}$ can be calculated recursively by using the UF as:

$$\begin{aligned}
\widehat{\theta}_{t|t} &= \widehat{\theta}_{t|t-1} + K_t \nu \\
M_{t|t} &= M_{t|t-1} - K_t M^{\nu\nu}_{t|t-1} K^*_t \\
\nu &= y_t - \widehat{y}_{t|t-1} \\
K_t &= M^{\theta\nu}_{t|t-1} (M^{\nu\nu}_{t|t-1})^{-1},
\end{aligned}$$

Tracking of Phase Noise with UF

- In order to implement the UF algorithm, one needs to compute
  - Prediction of the new state of the phase noise $\widehat{\theta}_{t|t-1}$ and its variance $M_{t|t-1}$
  - Prediction of the expected observation $\widehat{y}_{t|t-1}$ and the innovation variance $M^{\nu\nu}_{t|t-1}$
  - Prediction of the cross correlation $M^{\theta\nu}_{t|t-1}$.
- From the state equation, we can easily obtain $\widehat{\theta}_{t|t-1}$ and $M_{t|t-1}$ as

$$\begin{aligned} \widehat{\theta}_{t|t-1} &= \widehat{\theta}_{t-1|t-1} \\ M_{t|t-1} &= M_{t|t-1} + \sigma_u^2 \end{aligned}$$

- Furthermore, it can be easily shown that

$$\widehat{y}_{t|t-1} = E\{\exp(j\theta_t)|\mathbf{Y}_{t-1}\}$$

Tracking of Phase Noise with UF (Cont.)

$$M^{\nu\nu}_{t|t-1} = 1 + \sigma_w^2 - |\widehat{y}_{t|t-1}|^2$$

$$M^{\theta\nu}_{t|t-1} = E\{\theta_t \exp(-j\theta_t)|\mathbf{Y}_{t-1}\} - \widehat{y}^*_{t|t-1}\widehat{\theta}_{t|t-1}$$

- The expectations above can be computed by UF
- Since the θ_t's are one-dimensional and Gaussian, three sigma-points would be sufficient to implement the algorithm.
- The three sigma points and the corresponding weights are chosen according to the general formulation as

$$\begin{aligned}
\Theta^{(0)}_{t|t-1} &= \widehat{\theta}_{t|t-1}, & W_0 &= \kappa/(1+\kappa) \\
\Theta^{(1)}_{t|t-1} &= \widehat{\theta}_{t|t-1} + \sqrt{(1+\kappa)M_{t|t-1}}, & W_1 &= 1/2(1+\kappa) \\
\Theta^{(2)}_{t|t-1} &= \widehat{\theta}_{t|t-1} - \sqrt{(1+\kappa)M_{t|t-1}}, & W_2 &= 1/2(1+\kappa).
\end{aligned}$$

SMC for Blind Detection and Estimation

- We can now make timely estimates of θ_t and detection of s_t
- With the Bayes theorem, we realize that the optimal solution to this problem is

$$\widehat{\theta_t} = E\{\theta_t|\mathbf{Y}_t\} = \int_{\mathbf{S}_t} \underbrace{\left[\int_{\theta_t} \theta_t p(\theta_t|\mathbf{S}_t, \mathbf{Y}_t) d\theta_t\right]}_{\mu_{\theta_t}(\mathbf{S}_t)} p(\mathbf{S}_t|\mathbf{Y}_t) d\mathbf{S}_t.$$

- It then follows that

$$\widehat{\theta_t} = E\{\theta_t|\mathbf{Y}_t\} = \int_{\mathbf{S}_t} \mu_{\theta_t}(\mathbf{S}_t) p(\mathbf{S}_t|\mathbf{Y}_t) d\mathbf{S}_t \ .$$

- Similarly, the data can be detected by the hard decisions on the symbol s_t by

$$\widehat{s_t} = \arg\max_{a_i \in A} P(s_t = a_i|\mathbf{Y}_t)$$

SMC for Blind Detection and Estimation (Cont.)

- if we can draw m independent random samples $\{\mathbf{S}_t^{(j)}\}_{j=1}^m$ from the distribution $p(\mathbf{S}_t|\mathbf{Y}_t)$, then we can approximate the quantities of interest $E\{\theta|\mathbf{Y}_t\}$ and $E\{1(s_t = a_i)|\mathbf{Y}_t\}$ respectively, by

$$E\{\theta|\mathbf{Y}_t\} \cong \frac{1}{m}\sum_{j=1}^{m} \mu_{\theta_t}(\mathbf{S}_t^{(j)})$$

$$E\{1(s_t = a_i)|\mathbf{Y}_t\} \cong \frac{1}{m}\sum_{j=1}^{m} 1(s_t^{(j)} = a_i)$$

- But, usually drawing samples from $p(\mathbf{S}_t|\mathbf{Y}_t)$ directly is usually difficult. Instead, sample generation from some *trial distribution* may be easier

- In this case, the idea of *importance sampling* can be used

SMC for Blind Detection and Estimation (Cont.)

- By associating the weight

$$w_t^{(j)} = \frac{p(\mathbf{S}_t^{(j)}|\mathbf{Y}_t)}{q(\mathbf{S}_t^{(j)}|\mathbf{Y}_t)}$$

to the samples, the quantities of interest, $E\{1(s_t = a_i)|\mathbf{Y}_t\}$ and $E\{\theta_t|\mathbf{S}_t\}$ can be approximated as follows.

$$E\{\theta|y_t\} \cong \frac{1}{W_t}\sum_{j=1}^{m} \mu_t(\mathbf{S}_t^{(j)})w_t^{(j)}$$

$$E\{1(s_t = a_i)|\mathbf{Y}_t\} \cong \frac{1}{W_t}\sum_{j=1}^{m} 1(s_t^{(j)} = a_i)w_t^{(j)}, \quad i = 1, 2, .., |A|$$

with $W_t = \sum w_t^{(j)}$

SMC for Blind Detection and Estimation (Cont.)

- For this trial distribution, the importance weight is updated according to

$$w_t^{(j)} = w_{t-1}^{(j)} p(y_t | \mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}), \quad t = 0, 1, \cdots$$

- The predictive distribution is given by

$$p(y_t | \mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}) = \sum_{a_i \in A} p(y_t | \mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}, s_t = a_i) P(s_t = a_i | \mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)})$$

$$= \sum_{a_i \in A} p(y_t | \mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}, s_t = a_i) P(s_t = a_i)$$

SMC for Blind Detection and Estimation (Cont.)

- Furthermore, it can be shown from the state and observation equations,respectively, that

$$p(y_t|\mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}, s_t = a_i) \sim N(\mu_{y_t}^{(j)}(i), \sigma_{y_t}^{2(j)}(i))$$

with mean and variance given by

$$\begin{aligned}
\mu_{y_t}^{(j)}(i) &= E\{y_t|\mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}, s_t = a_i\} \\
&= a_i(H_t\mu_{\theta_{t-1}}^{(j)} + Q_t) \\
\sigma_{y_t}^{2(j)}(i) &= \text{Var}\{y_t|\mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}, s_t = a_i\} \\
&= \sigma_{\theta_{t-1}}^{2(j)} + \sigma_n^2 + \sigma_p^2
\end{aligned}$$

where the quantities $\mu_{\theta_t}^{(j)}$ and $\sigma_{\theta_t}^{2(j)}$ can be computed recursively with UF equations

SMC for Blind Detection and Estimation (Cont.)

- The trial distribution can be computed as follows:

$$\begin{aligned} p(s_t = a_i|\mathbf{Y}_t, \mathbf{S}_{t-1}^{(j)}) &= p(y_t|\mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}, s_t = a_i) \\ &\quad \times P(s_t = a_i|\mathbf{Y}_{t-1}, \mathbf{S}_{t-1}^{(j)}) \\ &\triangleq \xi_{t,i}^{(j)} \end{aligned}$$

where

$$\xi_{t,i}^{(j)} = \frac{1}{\pi \sigma_{y_t}^{2(j)}(i)} \exp\left(-\frac{||y_t - \mu_{y_t}^{(j)}(i)||^2}{\sigma_{y_t}^{2(j)}(i)} \right) P(s_t = a_i).$$

Summary of the Proposed Approach

Step 1- *Initialization:*

- Initialize the Kalman filter

$$\begin{aligned} \mu_{\theta_{-1}}^{(j)} &= \widehat{\theta}_{-1|-1}^{(j)} = 0 \\ \sigma_{\theta_{-1}}^{2(j)} &= M_{-1|-1}^{(j)} = \pi^2/12, \quad j = 1, 2, \cdots, m. \end{aligned}$$

- Initialize the importance weights: $w_{-1}^{(j)} = 1, j = 1, 2, \cdots, m$

Step 2- *Compute* $\xi_{t,i}^{(j)}$, *and For each* $a_i \in A$ *compute the* $\mu_{y_t}^{(j)}(i), \sigma_{y_t}^{2(j)}(i)$ *and* $\xi_{t,i}^{(j)}$

Step 3- *Draw samples* $s_t^j, j = 1, 2, \cdots, m$ from the set A with probabilities

$$P(s_t^{(j)} = a_i) \propto \xi_{t,i}^{(j)}, \quad a_i \in A.$$

Summary of the Proposed Approach

Step 4- *Compute the importance weights:*

$$w_t^{(j)} = w_{t-1}^{(j)} \sum_{a_i \in A} \xi_{t,i}^{(j)}.$$

Step 5- *Detect the symbol* s_t

Step 6- *Update the a posteriori mean and variance of the phase noise*

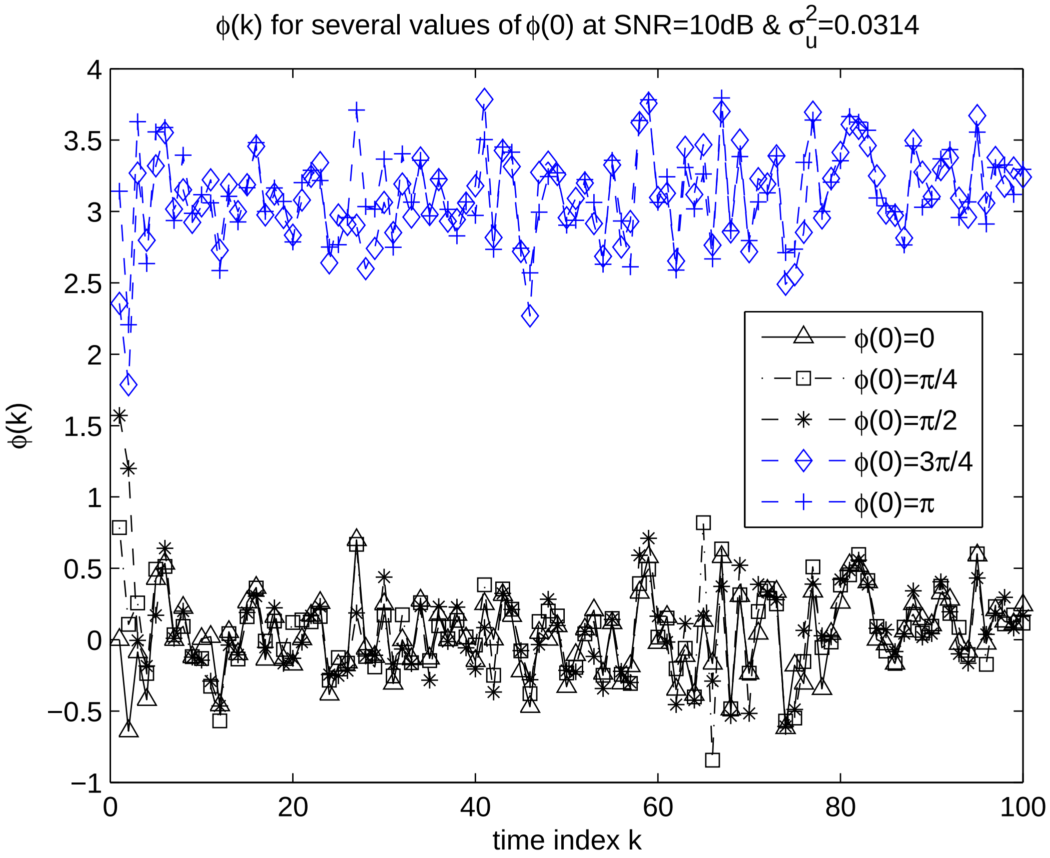
$$\begin{aligned} \mu_{\theta_t}(\mathbf{S}_t^{(j)}) &\stackrel{\Delta}{=} \mu_{\theta_t}^{(j)} = \widehat{\theta}_{t|t}^{(j)} \\ \sigma_{\theta_t}^{2(j)}(\mathbf{S}_t^{(j)}) &\stackrel{\Delta}{=} \sigma_{\theta_t}^{2(j)} = M_{t|t}^{(j)} \quad j = 1, 2, \cdots, m. \end{aligned}$$

according to the Kalman equations

Step 7- *Do the resampling*

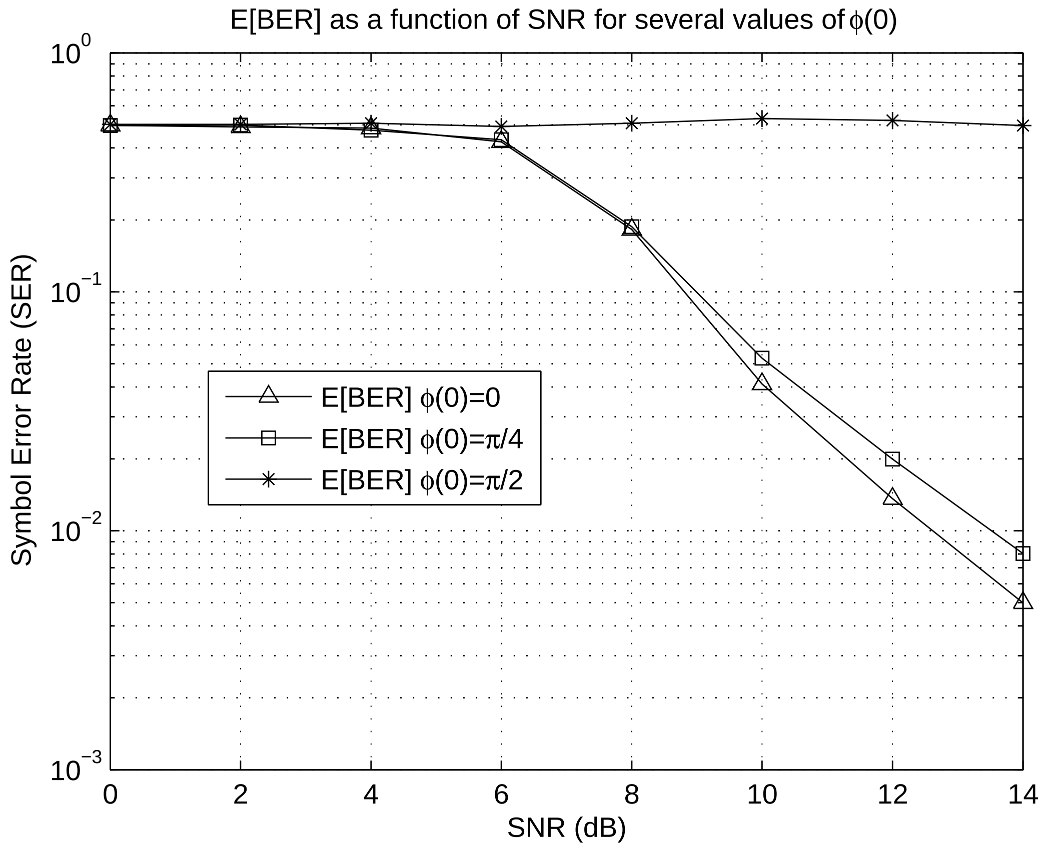
Simulation Results

Tracking Performance of the Proposed SMC Method (Phase Ambiguity)



Simulation Results (Cont.)

BER Performance of the Proposed SMC Method - Bit error rate



Conclusions

- We have developed a new adaptive Bayesian approach for blind phase noise estimation and data detection based on sequential monte carlo methodology
- The optimal solutions to joint symbol detection and phase noise estimation problem is computationally prohibitive to implement by conventional methods
- Thus the proposed sequential approach offers an novel and powerful approach to tackling this problem at a reasonable computational cost
  - as the initial phase error $\phi(0)$ approaches π, the probability that the phase error converges to the dual equilibrium point becomes very high
  - as the initial phase error $\phi(0)$ approaches π, the BER increases, for $\phi(0) = \pi/2$, the BER is almost equal to 1 (due to ambiguity).