Blind Identification Based on Expectation-Maximization Algorithm Coupled With Blocked Rhee–Glynn Smoothing Estimator

Wenhao Chen, *Student Member, IEEE*, Lu Ma, and Xuwen Liang

***Abstract*— In this letter, we consider blind estimation of channel parameters over a frequency-selective channel. We use a blocked Rhee–Glynn smoothing estimator to derive E-step in the expectation-maximization (EM) algorithm. The proposed algorithm copes with the curse of dimensionality of a forward–backward algorithm; meanwhile, it is easy to parallelize, which is amenable to a modern computing hardware and speeds up the estimation of channel parameters. The experiment results show that the proposed algorithm is close to the Baum–Welch algorithm in terms of convergence of channel coefficients and outperforms the EM algorithm coupled with a joined two-filter smoothing algorithm in terms of convergence of channel coefficients and running time.**

***Index Terms*— Maximum likelihood estimation, multipath channels, hidden Markov models, Rhee-Glynn estimator, expectation-maximization algorithm.**

I. Introduction

We consider blind estimation of channel parameters over an unknown frequency-selective channel, modeled as a finite impulse response filter, in additive Gaussian complex noise. The problem can be seen as parameter estimation in Hidden Markov Models (HMM), so the expectation-maximization (EM) algorithm coupled with Forward-Backward (FB) algorithm can be used to deal with it. However, when the state space is large (because the size of the modulation constellation and/or the channel order are large), the computational complexity of FB algorithm explodes. Since the derivation of the exact smoothing distributions is the most computation-intensive part of the EM algorithm when the state space is large, [1] replaced it by particle smoothing. However, all the particle smoothing algorithms in [1] do not actually sample their particle positions from the smoothing distribution, but from the filtering distribution - even if it is from the forward as well as the backward filtering distribution for the Joined Two-Filter smoothing algorithms. If the channel is maximum phase, the forward particle filter is less likely to identify the states with high smoothing probabilities, so the channel estimation will not be completely accurate. Meanwhile, the computation complexity of particle smoothing algorithms proposed in [1] increases significantly when the modulation order increases, and due to propagation steps and resampling steps, they are hard to parallelize.

Recently, Rhee-Glynn smoothing estimator (RGSE) [2] was proposed, which is the first unbiased estimator of smoothing expectations and allows for a complete parallel implementation. However, when the channel is maximum phase and the length of the observed data series is long, we need more particles to maintain the average meeting time of RGSE in a few steps. So in this letter, we split the observed data series into disjoint blocks to make the length of each block short, then in each block, we run RGSE easily no matter what channel type (Minimum/Maximum phase) is. We can adjust the number of blocks according to signal-to-noise ratio (SNR) to make blocked Rhee-Glynn smoothing estimator (blocked-RGSE) approximate RGSE well. The combination of EM algorithm and blocked-RGSE is used to deal with the blind estimation of the channel parameters.

The rest of this letter is organized as follows. In section II, the channel model are described. In section III, the details of proposed algorithm are presented. Simulation results are provided in Section IV. Finally, concluding remarks are given in Section V.

II. Model Description

Let M be the set of linear modulation constellation points of size $K = |M|$. The transmitted symbols $\{a_t\}_{t\geq 1}$ are uniformly and independently drawn from M. After preprocessing, the baseband complex envelope of the received signal sampled at one sample per symbol at the output of a matched filter can be written as

$$y_t = \sum_{l=0}^{L-1} a_{t-l} h_l + \varepsilon_t, \varepsilon_t \sim \mathbb{C}\mathrm{N}(0, \sigma^2), \tag{1}$$

where L be the channel order and $\boldsymbol{h} = [h_0, ..., h_{L-1}]^{\mathrm{T}}$ be the channel coefficients, superscript “T” means matrix transpose. $\{\varepsilon_t\}_t$ is a sequence of i.i.d. complex circular Gaussian variables with variance σ^2.

The received signal y_t can be modeled as a probabilistic function of a hidden state at time t, the hidden states form a first order Markov chain. So we can model the received signals by a first order HMM with the following characteristics:

1) The state of the HMM at the t-th time instant is $\boldsymbol{s}_t = [a_t, ..., a_{t-L+1}]^{\mathrm{T}}$, which is in the state space X of size K^L.

Manuscript received June 6, 2018; accepted July 9, 2018. Date of publication July 13, 2018; date of current version September 8, 2018. This work was supported by the Shanghai Natural Science Foundation (Grant No. shkw15zr01). The associate editor coordinating the review of this paper and approving it for publication was M. Oner. *(Corresponding author: Xuwen Liang.)*

W. Chen is with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China, and with the Shanghai Engineering Center for Microsatellites, Chinese Academy of Sciences, Shanghai 201210, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: 18221505463@163.com).

L. Ma is with Shanghai SpaceOK Aerospace Technology Co., Ltd., Shanghai 201802, China (e-mail: e_wqs@hotmail.com).

X. Liang is with the Shanghai Engineering Center for Microsatellites, Chinese Academy of Sciences, Shanghai 201210, China (e-mail: 18217631362@163.com).

Digital Object Identifier 10.1109/LCOMM.2018.2855969

2) The initial state is sampled uniformly from X.
3) The transition probability distribution from $\boldsymbol{s} = [a_1, ..., a_L]^{\mathrm{T}}$ to $\boldsymbol{s}' = [a'_1, ..., a'_L]^{\mathrm{T}}$ is

$$q(\boldsymbol{s}'|\boldsymbol{s}) = \begin{cases} \frac{1}{K} & \text{if } [a'_2, ..., a'_L]^{\mathrm{T}} = [a_1, ..., a_{L-1}]^{\mathrm{T}}, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

4) The probability of the observation y conditional to the current state $\boldsymbol{s}$ can be expressed as

$$f(y; \boldsymbol{\theta}|\boldsymbol{s}) = \frac{1}{\pi\sigma^2} exp(-\frac{1}{\sigma^2}|y - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{s}|^2), \quad (3)$$

where $\boldsymbol{\theta} = (\boldsymbol{h}, \sigma)$ denotes the vector of unknown parameters, which is to be estimated in the following part.

Finally, for $1 \leq i \leq j \leq T$, we introduce the short-hand notations $i : j$ for the set $\{i, ..., j\}$ and $v_{i:j}$ for the vector $(v_i, ..., v_j)$.

III. Proposed Method

The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ maximizes the log-likelihood $\log p(y_{1:T}; \boldsymbol{\theta})$. In HMM, we can use EM algorithm to estimate $\hat{\boldsymbol{\theta}}$. In this specific model, we take $\boldsymbol{s}_{1:T}$ as the hidden data, $y_{1:T}$ as the observed data, and $\boldsymbol{\theta}^{(i)} = (\boldsymbol{h}^{(i)}, \sigma^{(i)})$ as the current estimate of $\boldsymbol{\theta}$. the new reestimates are derived as follows [3]:

1) E-step. The conditional expectation of the log-likelihood of the complete data (the combination of the observed data and hidden data) is defined as

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\boldsymbol{s}_{1:T}; \boldsymbol{\theta}^{(i)}|y_{1:T}}[log(p(y_{1:T}, \boldsymbol{s}_{1:T}; \boldsymbol{\theta}))], \quad (4)$$

where $\mathbb{E}_{\boldsymbol{s}_{1:T}; \boldsymbol{\theta}^{(i)}|y_{1:T}}[\cdot]$ stands for expectation about $\boldsymbol{s}_{1:T}$ conditioned on $y_{1:T}$.

2) M-step. The new reestimates of unknown parameters can be obtained by calculating

$$\boldsymbol{\theta}^{(i+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i)}). \quad (5)$$

So, by taking the derivative of (4) with respect to $\boldsymbol{h}$ and σ^2, we can obtain the following equations:

$$\mathbb{E}_{\boldsymbol{s}_{1:T}; \boldsymbol{\theta}^{(i)}|y_{1:T}}[g(\boldsymbol{s}_{1:T})](\boldsymbol{h}^*)^{(i+1)} = \mathbb{E}_{\boldsymbol{s}_{1:T}; \boldsymbol{\theta}^{(i)}|y_{1:T}}[z(\boldsymbol{s}_{1:T})], \quad (6)$$

$$(\sigma^2)^{(i+1)} = \frac{1}{T}\mathbb{E}_{\boldsymbol{s}_{1:T}; \boldsymbol{\theta}^{(i)}|y_{1:T}}[d(\boldsymbol{s}_{1:T})], \quad (7)$$

where

$$g(\boldsymbol{s}_{1:T}) \triangleq \sum_{t=1}^{T} \boldsymbol{s}_t \boldsymbol{s}_t^{\mathrm{H}}, \quad z(\boldsymbol{s}_{1:T}) \triangleq \sum_{t=1}^{T} \boldsymbol{s}_t y_t^*,$$
$$d(\boldsymbol{s}_{1:T}) \triangleq \sum_{t=1}^{T} |y_t - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{s}_t|^2.$$

Superscript "$*$" denotes complex conjugate and superscript "H" represents Hermitian tranpose.

Since the derivation of the smoothing distribution $P(\boldsymbol{s}_{1:T}; \boldsymbol{\theta}^{(i)}|y_{1:T})$ is the most difficult and expensive part when the state space is large, we resort to RGSE which combines a generic debiasing technique for Markov chains [4] and the conditional particle filter for smoothing [5]. Through coupling two conditional particle filter, the combination of the two methods delivers the smoothing estimator. To understand the details about how to couple two conditional particle filters, the readers can refer to [6]. In our proposed method, we use index-coupled resampling to sample pairs of ancestor indices whose computational cost is linear in N, and take advantage of ancestor sampling technique [7] to improve the mixing of the underlying conditional particle filter kernel.

When we follow the steps in [6] to obtain RGSE, we will find that in all the bootstrap particle filters (BPF), conditional particle filters with ancestor sampling (CPF-AS) and coupled conditional particle filters with ancestor sampling (CCPF-AS), the process-generating variables are independent and identically distributed, because the random variables used to initialize the particle filters are uniform distribution between 0 and 1, and the random variables used to propagate the particles are uniformly and independently drawn from M. The steps to achieve two Markov chains $(\boldsymbol{s}_{1:T}^{(n)})_{n\geq 0}$ and $(\tilde{\boldsymbol{s}}_{1:T}^{(n)})_{n\geq 0}$ on the space X^T of trajectories can be summarized as: At first, we draw two trajectories $\boldsymbol{s}_{1:T}^{(0)}$ and $\tilde{\boldsymbol{s}}_{1:T}^{(0)}$ from two independent BPFs, which we denote by $\boldsymbol{s}_{1:T}^{(0)} \sim$BPF$(\boldsymbol{U}^{(0)})$ and $\tilde{\boldsymbol{s}}_{1:T}^{(0)} \sim$BPF$(\tilde{\boldsymbol{U}}^{(0)})$, with $\boldsymbol{U}^{(0)} \sim \varphi$ and $\tilde{\boldsymbol{U}}^{(0)} \sim \varphi$ denoting the process-generating variables. Then, we apply one step of the CPF-AS to the first trajectory: sample process-generating variables $\boldsymbol{U}^{(1)} \sim \varphi$ and write $\boldsymbol{s}_{1:T}^{(1)} \sim$CPF-AS$(\boldsymbol{s}_{1:T}^{(0)}, \boldsymbol{U}^{(1)})$. At last, for all $n \geq 2$, we apply the CCPF-AS to the pair of trajectories, which is written as $(\boldsymbol{s}_{1:T}^{(n)}, \tilde{\boldsymbol{s}}_{1:T}^{(n-1)}) \sim$CCPF-AS$(\boldsymbol{s}_{1:T}^{(n-1)}, \tilde{\boldsymbol{s}}_{1:T}^{(n-2)}, \boldsymbol{U}^{(n)})$, where $\boldsymbol{U}^{(n)} \sim \varphi$. Until $\boldsymbol{s}_{1:T}^{(n)} = \tilde{\boldsymbol{s}}_{1:T}^{(n-1)}$, we stop the iteration. Then the two chains satisfy the following conditions [2] and can be used to estimate the smoothing expectations:

- The two chains admit the smoothing distribution as invariant distribution both.
- For all $n \geq 0$, $\boldsymbol{s}_{1:T}^{(n)}$ and $\tilde{\boldsymbol{s}}_{1:T}^{(n)}$ have the same marginal distribution.
- There exists a time τ, termed the meeting time, such that $\boldsymbol{s}_{1:T}^{(n)} = \tilde{\boldsymbol{s}}_{1:T}^{(n-1)}$ almost surely for all $n \geq \tau$.

Here we take estimation of $\mathbb{E}_{\boldsymbol{s}_{1:T}; \boldsymbol{\theta}^{(i)}|y_{1:T}}[d(\boldsymbol{s}_{1:T})]$ as an example to illustrate the basic RGSE. Using the chains obtained above, the smoothing expectation estimator can be written as:

$$D = d(\boldsymbol{s}_{1:T}^{(0)}) + \sum_{n=1}^{\tau-1} d(\boldsymbol{s}_{1:T}^{(n)}) - d(\tilde{\boldsymbol{s}}_{1:T}^{(n-1)}). \quad (8)$$

To reduce the variance of the estimator, we use $D_{m,\infty}$ estimator, exploit Rao-Blackwellization to take all the particles generated by the conditional particle filters into account, and repeat R times RGSE independently to calculate the average, the details about these ways to reduce variance of estimator can be found in [2].

From [6] we know that when T increases, we should increase the number of particles N to make the meeting time and variance of estimator stable. So in this specific channel blind estimation problem, we propose an approximate way to estimate the smoothing expectations. As SNR increases, the observation likelihood is more informative. So we divide the observed data $y_{1:T}$ into B disjoint blocks, the length of every block is $l = \frac{T}{B}$ (here choose B to make T be divisible by B). Then, in the i-th block, i.e. $\Delta_i \triangleq (i-1)\times l+1 : i\times l$, we use RGSE to estimate smoothing expectations of this block. Specificly, $\mathbb{E}_{\boldsymbol{s}_{\Delta_i};\boldsymbol{\theta}^{(i)}|y_{\Delta_i}}[d(\boldsymbol{s}_{\Delta_i})]$ can be estimated by

$$D_i = d(\boldsymbol{s}_{\Delta_i}^{(0)}) + \sum_{n=1}^{\tau-1} d(\boldsymbol{s}_{\Delta_i}^{(n)}) - d(\tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)}). \tag{9}$$

Finally, the smoothing expectation estimate D for the whole observed data can be approximated by

$$D \approx \sum_{i=1}^{B} D_i. \tag{10}$$

As SNR increases, we can increase B accordingly. There are two benefits using the approximation, the first one, in each block, the length of the observed data series is comparatively short, so the CCPF-AS will meet easily; the second one, we can run RGSE for each block parallelly, which can accelerate the algorithm.

The procedure of $H_{m,\infty}$ estimator proposed in [2] for the smoothing expectations of the i-th block is summarized in Algorithm 1. Likewise, we can exploit Rao-Blackwellization and sample R RGSEs independently to further reduce the variance of estimator for the smoothing expectations of the i-th block.

Algorithm 1 $H_{m,\infty}$ Estimator for the Smoothing Expectations of the i-th Block

Input: m.
Draw $\boldsymbol{s}_{\Delta_i}^{(0)} \sim \mathrm{BPF}(\boldsymbol{U}^{(0)})$ and $\tilde{\boldsymbol{s}}_{\Delta_i}^{(0)} \sim \mathrm{BPF}(\tilde{\boldsymbol{U}}^{(0)})$.
Draw $\boldsymbol{s}_{\Delta_i}^{(1)} \sim \text{CPF-AS}(\boldsymbol{s}_{\Delta_i}^{(0)}, \boldsymbol{U}^{(1)})$.
for $n = 1 : m$ **do**
  Draw $(\boldsymbol{s}_{\Delta_i}^{(n)}, \tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)}) \sim \text{CCPF-AS}(\boldsymbol{s}_{\Delta_i}^{(n-1)}, \tilde{\boldsymbol{s}}_{\Delta_i}^{(n-2)}, \boldsymbol{U}^{(n)})$.
end for
Set $G_i = g(\boldsymbol{s}_{\Delta_i}^{(m)})$, $Z_i = z(\boldsymbol{s}_{\Delta_i}^{(m)})$, $D_i = d(\boldsymbol{s}_{\Delta_i}^{(m)})$.
if $\boldsymbol{s}_{\Delta_i}^{(n)} \neq \tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)}$ **then**
  for $n = m+1, m+2, \ldots$ **do**
    Draw $(\boldsymbol{s}_{\Delta_i}^{(n)}, \tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)}) \sim \text{CCPF-AS}(\boldsymbol{s}_{\Delta_i}^{(n-1)}, \tilde{\boldsymbol{s}}_{\Delta_i}^{(n-2)}, \boldsymbol{U}^{(n)})$.
    if $\boldsymbol{s}_{\Delta_i}^{(n)} = \tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)}$ **then**
      $\tau = n$, exit the loop.
    end if
    Set $G_i = G_i + g(\boldsymbol{s}_{\Delta_i}^{(n)}) - g(\tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)})$.
    Set $Z_i = Z_i + z(\boldsymbol{s}_{\Delta_i}^{(n)}) - z(\tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)})$.
    Set $D_i = D_i + d(\boldsymbol{s}_{\Delta_i}^{(n)}) - d(\tilde{\boldsymbol{s}}_{\Delta_i}^{(n-1)})$.
  end for
end if

The computational cost of sampling one RGSE for the i-th block is of order $Nl \times max\{m, \tau\}$. Because of the highly parallelizable nature of this algorithm, the running time of the algorithm is only related to the smoothing estimators for one block, which imply that using the proposed algorithm can greatly accelerate the procedure of blind channel estimation.

TABLE I
CHANNEL TYPE'S IMPACT ON AVERAGE MEETING TIME

Channel type	Minimum phase	Maximum phase
$\mathbb{E}(\tau)$	3.226	3.186

TABLE II
CHANNEL ORDER'S IMPACT ON AVERAGE MEETING TIME

Channel order	L=2	L=3	L=4
$\mathbb{E}(\tau)$	2.6611	3.226	3.8533

TABLE III
MODULATION ORDER'S IMPACT ON AVERAGE MEETING TIME

Modulation type	64QAM	128QAM	256QAM
$\mathbb{E}(\tau)$	3.226	3.2730	3.3044

IV. NUMERICAL RESULTS

The performance of the proposed smoothing estimator is tied to the meeting time τ, so in the specific model, we firstly research how channel condition (includes the minimum/maximum phase channel and the channel order) and modulation order influence on τ. In the experiments related to τ below, we keep SNR$= 12dB, T = 10$ and modulation type is 64QAM except for the experiments researching modulation order's influence on τ. In minimum phase channel, we set $H = [0.63, 0.05, -0.3]^{\mathrm{T}}$ and $N = 500$, in maximum phase channel, we set $H = [-0.3, 0.05, 0.8]^{\mathrm{T}}$ and $N = 30000$, the average meeting time is shown in Table I. In maximum phase channel, we need more particles compared to minimum phase channel to maintain $\mathbb{E}(\tau)$ in a few steps. We set $H = [0.63, -0.3]^{\mathrm{T}}$, $H = [0.63, 0.05, -0.3]^{\mathrm{T}}$ and $H = [0.63, 0.05, 0.05, -0.3]^{\mathrm{T}}$ respectively to research how channel order influences on $\mathbb{E}(\tau)$, $N = 500$ for each experiment. From Table II we can see that keeping N and channel type unchanged, with the increase of L, the average meeting time is stable. At last, we choose modulation type as 64QAM/128QAM/256QAM to see how modulation order impacts on $\mathbb{E}(\tau)$, likewise, N is fixed to 500 and $H = [0.63, 0.05, -0.3]^{\mathrm{T}}$ for each experiment. From Table III we can find that keeping N unchanged, with the increase of modulation order, the average meeting time is stable too.

Then, the performance of blocked-RGSE is analysed. We choose RGSE for the whole observed data (whole-RGSE) as reference. Here we take estimation of $\mathbb{E}_{\boldsymbol{s}_{1:T};\boldsymbol{\theta}^{(i)}|y_{1:T}}[d(\boldsymbol{s}_{1:T})]$ as an example, we choose 64QAM as modulation constellation, set SNR=$12dB$, $T = 300$ and $H = [0.63, 0.05, -0.3]^{\mathrm{T}}$, in whole-RGSE $N = 30000$, $m = 5$ while in blocked-RGSE $N = 500$, $m = 5$ and $B = 60$. Fig.1 shows the distributions of estimates of whole-RGSE and blocked-RGSE. When SNR is comparatively high, the blocked-RGSE can approximate whole-RGSE well, meanwhile, it can reduce the computation complexity compared to whole-RGSE and can be implentmented parallelly.

Finally, we study on the performance of EM algorithm coupled with Blocked-RGSE (EM-Blocked-RGSE),

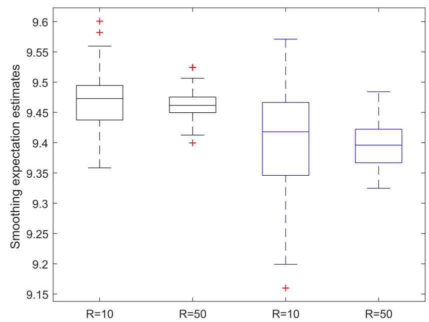

Fig. 1. The distribution of smoothing expectation estimates, the left two box plots are whole-RGSE and right two are blocked-RGSE.

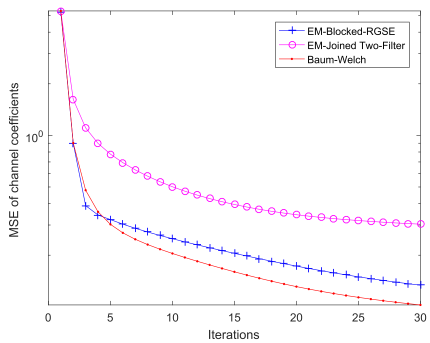

Fig. 2. MSE of channel coefficients in each iteration.

here we choose EM algorithm coupled with Joined Two-Filter (EM-Joined Two-Filter) and Baum-Welch algorithm as reference. The convergence of the EM is measured in terms of the mean squared error $\mathrm{MSE}(\hat{\boldsymbol{h}}) = \mathbb{E}[\sum_{l=0}^{L-1} |\hat{h}_l - h_l|^2]$ of the current estimate of the channel coefficients $\hat{\boldsymbol{h}}$ compared to the true coefficients $\boldsymbol{h}$. Note that blind channel estimation problem has phase ambiguity, the potential true coefficients may be one of $\boldsymbol{h}$ with phase rotation of $\{0, \pi/2, \pi, 3\pi/2\}$. The initial values of the channel coefficients for the EM algorithm are estimated by higher-order statistics (HOS) of the received signal [8]. We choose 64QAM as modulation constellation, set SNR=$15dB$, $T = 300$ and $H = [-0.3, 0.05, 0.8]^{\mathrm{T}}$ which is maximum phase channel. In each Monte Carlo experiment, the three methods mentioned above run with the same observed data and initial values. For EM-Blocked-RGSE, we set $N = 30000, m = 5, B = 60, R = 10$; in EM-Joined Two-Filter, we choose the number of forward filter particles $N_F = 500$ and the number of backward filter particles $N_B = 500$ too. From Fig.2 and Fig.3 we can find that in terms of convergence of channel coefficients, EM-Blocked-RGSE is close to Baum-Welch algorithm and outperforms EM-Joined Two-Filter; in terms of running time, due to propagation steps and resampling steps, Joined Two-Filter smoothing expectations estimator is hard to parallelize, and the computation complexity is $\mathcal{O}(K(N_F + N_B)T)$, while EM-Blocked-RGSE is easy to parallelize, the running time is only related to one block smoothing expectations estimator, whose computation complexity is $\mathcal{O}(Nl \times max\{m, \tau\})$, so EM-Blocked-RGSE is faster than EM-Joined Two-Filter. Furthermore, as modulation order increases, the computation complexity of Joined Two-Filter smoothing expectations estimator will increase too, but Blocked-RGSE will stay stable because the average meeting time will be stable when N keeps unchanged.

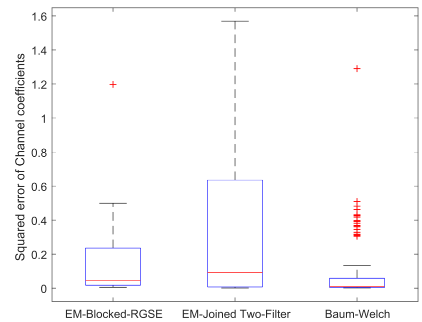

Fig. 3. The distribution of the squared error of channel coefficients after 30 iterations.

V. Conclusion

From experiments above, we can conclude that the proposed algorithm approximates Baum-Welch algorithm well but has a lower computational complexity compared to Baum-Welch algorithm when the state space is large. In addition, in terms of the accuracy of estimation, it outperforms EM-Joined Two-Filter when the channel is maximum phase. The computation complexity of proposed method is stable when modulation order and/or channel order change, meanwhile because it is easy to parallelize, its running time is less than EM-Joined Two-Filter. In the future, we will study on coupling of the other MCMC kernels of smoothing distribution to alleviate the channel type's impact on the average meeting time further, and the convergence properties of EM algorithm coupled with proposed approximate smoothing expectation estimator need to be researched in theory.

References

[1] S. Barembruch, A. Garivier, and E. Moulines, "On approximate maximum-likelihood methods for blind identification: How to cope with the curse of dimensionality," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4247–4259, Nov. 2009.

[2] P. E. Jacob, F. Lindsten, and T. B. Schon. (2016). "Coupling of particle filters." [Online]. Available: https://arxiv.org/abs/1606.01156

[3] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2406–2413, Jul. 1994.

[4] P. W. Glynn and C.-H. Rhee, "Exact estimation for Markov chain equilibrium expectations," *J. Appl. Probab.*, vol. 51, no. A, pp. 377–389, Dec. 2014.

[5] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *J. Roy. Statist. Soc. B, Statist. Methodol.*, vol. 72, no. 3, pp. 269–342, 2010.

[6] P. E. Jacob, F. Lindsten, and T. B. Schön. (2017). "Smoothing with couplings of conditional particle filters." [Online]. Available: https://arxiv.org/abs/1701.02002

[7] F. Lindsten, M. I. Jordan, and T. B. Schön, "Particle Gibbs with ancestor sampling," *J. Mach. Learn. Res.*, vol. 15, no. 15, pp. 2145–2184, 2014.

[8] A. Puengnim, T. Robert, N. Thomas, and J. Vidal, "Hidden Markov models for digital modulation classification in unknown ISI channels," in *Proc. Eur. Signal Process. Conf.*, Sep. 2015, pp. 1882–1886.