Expectation Propagation and Transparent Propagation in Iterative Signal Estimation in the Presence of Impulsive Noise

Anoush Mirbadin, Armando Vannucci, and Giulio Colavolpe
Department of Engineering and Architecture
University of Parma
Parma, Italy
Email: anoush.mirbadin@studenti.unipr.it

***Abstract*—Expectation Propagation (EP) and Transparent Propagation (TP) are employed in iterative estimation of correlated Gaussian samples in the presence of bursty impulsive noise, modeled as Markov Middleton class A. The proposed estimation strategy is based on a message-passing approach in which a Kalman Smoother and a Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm work in parallel. Due to the correlation between signal samples and correlation of channel states, the corresponding factor graph includes cycles. Therefore, the message passing approach should be implemented iteratively. Furthermore, the presence of Gaussian observations, continuous random variables, and impulsive noise states, discrete random variables, produces Gaussian mixtures. We utilize the variational inference techniques such as EP and TP to approximate the Gaussian mixtures and to avoid exponentially increasing complexity of messages. The performance of EP and TP based estimators are evaluated by using computer simulations. The results show a considerable improvement in performance brought about by the estimation strategy.**

***Index Terms*—BCJR; Expectation Propagation; impulsive noise; Kalman Smoother; Transparent Propagation**

I. Introduction

Smart grid brings about an efficient power management system based on the feedbacks provided by consumer sensors. In a grid, communication between sensors and smart meters can be accomplished through a variety of environments. Power line communications (PLC), as the backbone of smart grids, attract attention of both industry and academia. The received signal at the output of PLC channels can be represented as the summation of correlated Gaussian signal, and impulsive noise, which is modelled as Markov Middleton class A [1].

This model has three advantages. First, Gaussian samples can be considered to represent multicarrier modulated signals (e.g., OFDM) [2]. Second, correlation among Gaussian samples varies signal statistics sample by sample [3]. Finally, the model also makes it possible to implement intelligent signal processing techniques [4]–[6]. Although the suitability of the Middleton class A noise model for impulsive noise of PLC channels was well investigated in [7], the Markov Middleton is still a more general model for bursty impulsive noise than the other ones [8].

A Bayesian minimum mean square error (MMSE) estimation strategy was proposed in [9] to estimate independent Gaussian samples in the Markov-Gaussian scenario where the channel states were detected by using a Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. The work later expanded to a more general scenario ones [4]. The joint estimation of the correlated Gaussian samples and detection of the channel states was obtained through a factor graph based approach in which a Kalman smoother and a BCJR work in parallel [4]. The same approach was employed in the Markov Middleton scenario [5].

Approximation of Gaussian mixtures, which are generated due to the presence of continuous Gaussian observations and discrete channel states, were proposed through a variety of techniques. The hard decision is the simplest method in which the mixture is approximated by one of its component with the highest weight [4], [5]. This technique was further replaced by Expectation propagation (EP) and its simplest version Transparent propagation (TP) [6]. These algorithms approximate each Gaussian mixture to a single Gaussian distribution by minimizing the Kullback-Leibler divergence.

The main objective of this work is to employ EP and TP techniques, as alternatives to hard decision, in iterative signal estimation in the presence of bursty impulsive noise, modeled as Markov Middleton class A. To this end, we use the same system model as in [5] and we adopt the same estimation strategy as in [6]. Since the 2-state noise scenario is fully analyzed in [6], we focus our attention to higher noise states, i.e., 4-state noise model.

II. System Model and Factor Graph

We briefly review the model, introduce the factor graph representation of the estimation strategy, and compute the corresponding messages, according to the well-known sum-product algorithm.

We consider a sequence of Gaussian samples $\{s_k\}_{k=0}^{K-1}$ obtained by using a single pole Infinite Impulse Response (IIR) digital filter, forming an autoregressive model of order one (AR(1))

$$s_k = a_1 s_{k-1} + \omega_k \tag{1}$$

978-1-7281-6376-5/20/$31.00 ©2020 IEEE

TABLE I
THE SUM-PRODUCT MESSAGES FOR THE FACTOR GRAPH DEPICTED IN FIG. 1 (SUBSCRIPTS DENOTE MESSAGE DIRECTION).

$$
\begin{aligned}
p_f(s_k) &= \int p_f(s_{k-1})p_u(s_{k-1})p(s_k \mid s_{k-1})ds_{k-1}\\
P_f(i_k) &= \sum_{i_{k-1}} P(i_k \mid i_{k-1})P_f(i_{k-1})P_d(i_{k-1}))\\
&\qquad\qquad (k = 1, \cdots, K-1)\\
p_b(s_k) &= \int p_b(s_{k+1})p_u(s_{k+1})p(s_{k+1} \mid s_k)ds_{k+1}\\
P_b(i_k) &= \sum_{i_{k+1}} P(i_{k+1} \mid i_k)P_b(i_{k+1})P_d(i_{k+1})\\
&\qquad\qquad (k = K-2, \cdots, 0)\\
p_u(s_k) &= \sum_{i_k} P_f(i_k)P_b(i_k)p(y_k \mid s_k, i_k)\\
P_d(i_k) &= \int p_f(s_k)p_b(s_k)p(y_k \mid s_k, i_k)ds_k\\
&\qquad\qquad (k = 0, \cdots, K-1)
\end{aligned}
$$

in which the feeding sequence is white Gaussian, $\omega_k \sim \mathcal{N}(0, (1-a_1^2)\sigma_s^2)$ where σ_s^2 is the variance of the signal sample. The noisy observations $\{y_k\}_{k=0}^{K-1}$ are obtained by observing the signal samples through a channel impaired by impulsive noise

$$y_k = s_k + n_k \quad (k = 0, 1, \cdots, K-1) \tag{2}$$

where n_k is a noise random variable whose statistics follow the Middleton class A model

$$p(n_k) = \sum_{i=0}^{\infty} \frac{p_i}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{n_k^2}{2\sigma_i^2}\right\} \tag{3}$$

in which $\sigma_i^2 = (1 + \frac{i}{A\Gamma})\sigma_0^2$ where A is the impulsive index, Γ describes the power of impulsive noise [1], σ_0^2 is the variance of the background Gaussian noise, and $p_i = \frac{e^{-A}A^i}{i!}$ follows the Poisson distribution and represents the probability of being in the i-th noise state. It should be noted that when i impulsive events occur at the same time, the noise sample seen by the receiver has a zero-mean Gaussian distribution with the largest variance σ_i^2 [1]. In order to take into account the noise memory, probability density function (pdf) of noise should be truncated to a finite number of terms, i.e., finite number of sources of interference

$$p(n_k) = \sum_{i=0}^{M-1} \frac{p'_i}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{n_k^2}{2\sigma_i^2}\right\}, \; p'_i = \frac{p_i}{\sum_{j=0}^{M-1} p_j} \tag{4}$$

This memory truncation allows us to consider Markovianity between the noise samples by defining the noise transition matrix as follows

$$\boldsymbol{P} = \begin{bmatrix} x + (1-x)p'_0 & (1-x)p'_1 & \cdots & (1-x)p'_{M-1} \\ (1-x)p'_0 & x + (1-x)p'_1 & \cdots & (1-x)p'_{M-1} \\ \vdots & \vdots & \vdots & \vdots \\ (1-x)p'_0 & (1-x)p'_1 & \cdots & x + (1-x)p'_{M-1} \end{bmatrix} \tag{5}$$

where x, a number between 0 and 1, is the probability of remaining in the same state. It is known as correlation parameter and determines the noise burst duration. As x increases, the duration of impulsive events increases. The average duration of an impulsive event can be defined as

$$T_i = \frac{1}{(1-x)(1-p'_i)} \tag{6}$$

An implementation of the MMSE or the maximum a posteriori (MAP) estimation strategy can be obtained through factorizing the joint probability of continuous signal samples $\boldsymbol{s} = \{s_k\}$ and discrete channel states $\boldsymbol{i} = \{i_k\}$ given the observation samples $\boldsymbol{y} = \{y_k\}$ and applying the sum-product algorithm to the corresponding factor graph, that is represented in Fig. 1 and is in a one-to-one correspondence with the following joint posterior distribution:

$$
\begin{aligned}
p(\boldsymbol{s}, \boldsymbol{i} \mid \boldsymbol{y}) &\propto p(\boldsymbol{y} \mid \boldsymbol{s}, \boldsymbol{i})p(\boldsymbol{s})P(\boldsymbol{i})\\
&= \left[\prod_{k=1}^{K-1} p(y_k \mid s_k, i_k)p(s_k \mid s_{k-1})P(i_k \mid i_{k-1})\right]\\
&\quad p(y_0 \mid s_0, i_0)p(s_0)P(i_0)
\end{aligned} \tag{7}
$$

According to (1) , (2) and (5), one can conclude that

$$p(s_k \mid s_{k-1}) = g(s_k - a_1 s_{k-1}, (1-a_1^2)\sigma_s^2) \tag{8}$$

$$p(y_k \mid s_k, i_k) = g(y_k - s_k, \sigma_{i,k}^2) \tag{9}$$

$$P(i_k \mid i_{k-1}) = \boldsymbol{P}(i_{k-1}+1, i_k+1) \tag{10}$$

where $g(x-\eta, \gamma^2)$ is the standard Gaussian pdf with given mean η and variance γ^2. The arguments of matrix $\boldsymbol{P}$ in (10) represent row and column indices. Fig. 1 shows a portion of the corresponding factor graph. The resulting sum-product messages are summarized in table I. The initial conditions are: $P(i_0) = \sum_{n=0}^{M-1} p'_n \delta(i_0 - n)$,$p(s_0) = g(s_0, \sigma_s^2)$, $p(s_{K-1}) = 1$, and $P(i_{K-1}) = 1$.

III. EXPECTATION PROPAGATION

EP is a message-passing algorithm to perform approximate variational inference [10]. It works by minimizing the Kull-backLeibler divergence over a distribution family. Suppose $p(x)$ is a distribution mixture. We are going to approximate it by $q(x)$ belonging to the chosen family $\mathcal{F}$. The optimal approximating $q(x)$ is formally expressed by the following projection operator.

$$\text{proj}[p(x)] = \arg\min_{q\in\mathcal{F}} \text{KL}\left(p(x)||q(x)\right) \tag{11}$$

in which

$$\text{KL}\left(p(x)||q(x)\right) = \int p(x)\log\frac{p(x)}{q(x)}dx + \int \left(p(x) - q(x)\right)dx \tag{12}$$

In the case of exponential family, i.e. $q(x) = \exp(\sum_j G_j(x)v_j)$ in which v_j are the so called *natural parameters* of the distribution and $G_j(x)$ are the *features* of the family, the projection operator simply reduces to matching the expectation of the features (hence the name "Expectation Propagation") of $p(x)$ with those of $q(x)$. Now, we are able to derive EP for the family of Gaussian pdfs whose features, $G_j(x)$, are defined by $\{1, x, x^2\}$. Let $p(x)$ be a weighted sum of Gaussian distributions,

$$p(x) = \sum_i \alpha_i g(x - \mu_i, \sigma_i^2) \tag{13}$$

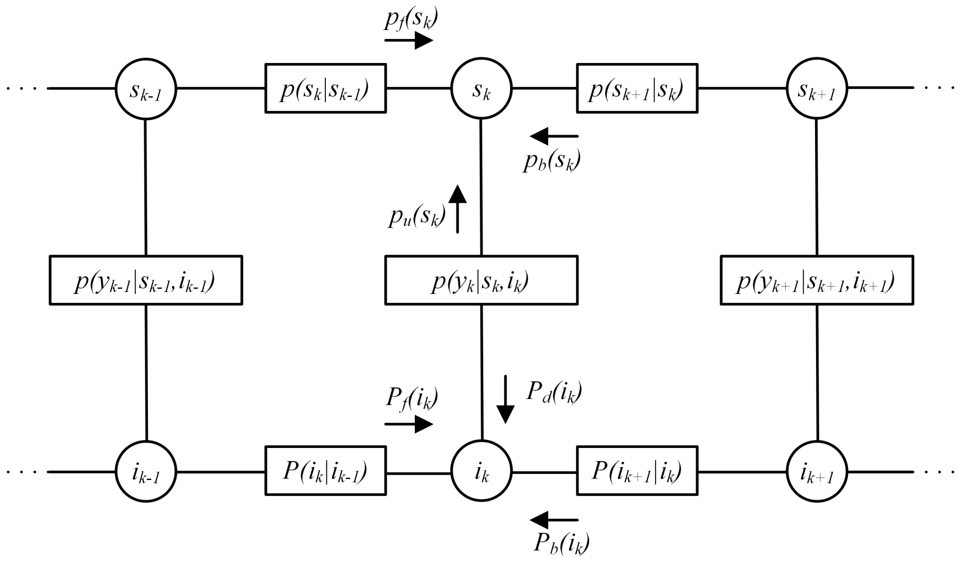

Fig. 1. A portion of the factor graph from (7).

where α_i is the weight of i-th Gaussian distribution and $\sum_i \alpha_i = 1$. The projection of the Gaussian mixture $p(x)$ into a single Gaussian pdf $q(x) = g(x-\mu, \sigma^2)$ can be obtained by matching their moments as follows

$$\mu = E_q[x] = E_p[x] = \sum_i \alpha_i \mu_i \tag{14}$$

$$\sigma^2 + \mu^2 = E_q[x^2] = E_p[x^2] = \sum_i \alpha_i (\sigma_i^2 + \mu_i^2) \tag{15}$$

Therefore,

$$\sigma^2 = E_q[(x-\mu)^2] = \sum_i \alpha_i (\sigma_i^2 + \mu_i^2) - \left|\sum_i \alpha_i \mu_i\right|^2 \tag{16}$$

IV. Iterative Estimation Strategy

As evident from Fig. 1, the factor graph has cycles; therefore, the message passing algorithm should be implemented iteratively. From the factor graph structure, the upper part of the factor graph estimates the correlated signal samples while the lower part of the graph is responsible of detecting the channel states. As a result, a parallel iterative schedule could be employed in which two halves of the graph work in parallel and they exchange their messages at every iteration [4].

Another important issue arising in this context is the presence of Gaussian mixtures as the messages (pdfs) passed along the factor graph edges, whose complexity increases exponentially at every time step. In particular, the messages passed between continuous distributions (s_k), Gaussian observations (y_k), and discrete random variables (the impulsive noise states, (i_k)), generates Gaussian mixtures at every vertical edge of the factor graph. One way to tackle this problem is to apply hard decisions to every Gaussian mixture such that the mixture is approximated by its Gaussian component with the highest weight [4], [5]. This simple approach extensively reduces the overall complexity at the expense of losing some information about channel states, decreasing the estimation accuracy.

Another approach is based on approximating every Gaussian mixture with a single Gaussian pdf minimizing the Kullback-Leibler divergence. This method can be categorized into two techniques: EP, proposed by Minka in [10], and TP, introduced by Vannucci *et al.* in [6]. Here, we use the same system model and estimation strategy as in [5] and we employ EP and TP techniques to reduce the complexity of messages. More precisely, we extend the work in [6], which is limited to the 2-state noise model, to a more realistic multilevel impulsive noise scenario.

To elaborate the estimation strategy in details, we should consider the equations of Tab. I and the corresponding messages on the factor graph. The pmf $P_d(i_k)$ has M different values, each in one-to-one correspondence with a different noise variance. Consequently, the messages at the bottom line of the graph could be computed by performing a M-step forward and backward recursions, forming a BCJR algorithm [9].

From another point of view, $p_u(s_k)$ is a mixture of M Gaussian distributions, each with its own weight and variance. If we approximate $p_u(s_k)$ with a single Gaussian pdf whose variance is not specified, since the noise is not stationary, all messages at the top line of the graph are Gaussian, thus reducing the computation of the "horizontal" messages (i.e., those sent forward and backward, with f and b subscripts) to the operation of a traditional Kalman smoother. At every iteration, the Kalman smoother provides the message $p_d(s_k) = p_f(s_k)p_b(s_k)$ and the BCJR computes the weights of components of observations, i.e., $P_u(i_k) = P_f(i_k)P_b(i_k)$.

Based on sum-product rules, the marginal pdf of s_k can be obtained by multiplicating all of the incoming messages to the

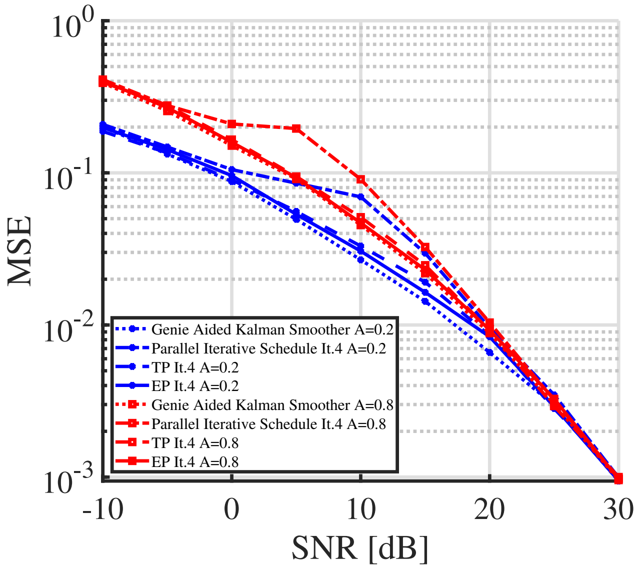

Fig. 2. MSE vs. SNR performance curves.

variable node s_k, which implies that

$$p(s_k|\boldsymbol{y}) = p_f(s_k)p_b(s_k)p_u(s_k) = p_d(s_k)p_u(s_k) \tag{17}$$

in which $p_d(s_k)$ as a product of two Gaussian messages is Gaussian and $p_u(s_k)$ is a Gaussian mixture. As a result, $p(s_k|\boldsymbol{y})$ is a Gaussian mixture. The mixture is then projected into a Gaussian pdf using EP.

$$\widetilde{p}(s_k|\boldsymbol{y}) = \text{proj}[p_d(s_k)p_u(s_k)] = g(s_k - \hat{s}_k, \hat{\sigma}_k^2) \tag{18}$$

where $\hat{s}_k$ is the signal estimate provided by EP. What remains is to compute $p_u(s_k)$ for the next iteration of the Kalman smoother. To this end, Gaussian division should be performed.

$$p_u^{EP}(s_k) = \frac{\widetilde{p}(s_k|\boldsymbol{y})}{p_d(s_k)} \tag{19}$$

This strategy is unstable. In situations where the variance of the denominator is larger than the variance of the numerator, improper distributions with negative variance are obtained and message rejection procedure should be performed [6]. A simpler and stable approach is the direct projection of the message $p_u(s_k)$ into the Gaussian family, called TP [6]. The lack of Gaussian division makes this strategy completely stable.

$$p_u^{TP}(s_k) = \text{proj}[p_d(s_k)p_u(s_k)] \tag{20}$$

In this case, the final estimation of signal samples can be obtained by

$$\widetilde{p}(s_k|\boldsymbol{y}) = p_d(s_k)p_u^{TP}(s_k) \tag{21}$$

V. Numerical Results

Fig. 2 shows the mean squared error (MSE) versus signal to noise ratio (SNR) for signal estimation in impulsive noise. The noise is modeled by a 4-state Markov Middleton class A. In order to simulate the bursty impulsive noise, a high value for correlation parameter x is considered. We use the same noise parameters as in [1]. We set $x = 0.98$, $A = 0.2, 0.8$, and $\Gamma = 0.01$, for impulsive noise. We consider $a_1 = 0.9$ and $\sigma_s^2 = 1$, for the AR(1) signal samples. 100 frames of 1000 samples each were transmitted for each SNR value.

The curve labeled "Genie Aided Kalman Smoother" is the lower bound for the performance in which Kalman smoother has exact knowledge of noise variances. The curve labeled "Parallel Iterative Schedule" is obtained using the proposed estimation strategy in [5] where Gaussian mixtures are approximated by hard decisions. The curves labeled by "TP" and "EP" address the estimation strategy in which approximation of Gaussian mixtures are performed by TP and EP, respectively. The final estimation is obtained after four iterations, denoted by "It.4" in the figures. Complete convergence is in fact observed after four iterations, after which the curves remain stable.

The comparison between simulation results at $A = 0.2$ and $A = 0.8$ reveals that as the value of A increases, i.e., the average number of active interferers increases, a worse estimation performance is obtained. It is evident that TP and EP based estimators perform close to the lower bound while the performance of the hard decision based estimator is degraded around the SNRs where the signal and the noise have almost equal power, i.e., where the signal estimate is neither dominated by noise, nor it is close to a noiseless scenario.

VI. Conclusion

This work addressed a factor graph based approach to estimate correlated Gaussian samples in bursty impulsive noise, in which the main limitation is the presence of Gaussian mixtures. Hence, this limitation has been overcome by employing EP and TP algorithms. It has been shown that the performances of EP and TP algorithms are close to the performance of the optimal strategy. It is worth noting that, unlike EP which is not necessarily stable, TP is inherently stable with lower complexity.

References

[1] G. Ndo, F. Labeau, and M. Kassouf, "A Markov-Middleton Model for Bursty Impulsive Noise: Modeling and Receiver Design," *IEEE Trans. Power Del.*, vol. 28, no. 4, pp. 2317-2325, Oct. 2013.

[2] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels," *IEEE Trans. Comm.*, vol. 48, no. 3, pp. 430-441, Mar. 2000.

[3] G. Doblinger, "Smoothing of noisy AR signals using an adaptive Kalman filter," *9th European Signal Processing Conference (EUSIPCO 1998)*, Rhodes, Greece, 1998, pp. 1-4.

[4] A. Vannucci, G. Colavolpe, R. Pecori, and L. Veltri, "Estimation of a Gaussian Source with Memory in Bursty Impulsive Noise," *2019 IEEE International Symposium on Power Line Communications and its Applications (ISPLC 2019)*, Prague, Czech Republic, Apr. 2019, pp. 1-6.

[5] A. Mirbadin, E. Kiani, A. Vannucci, and G. Colavolpe, "Estimation of Gaussian Processes in Markov-Middleton Impulsive Noise," *2019 1st Global Power, Energy and Communication Conference (GPECOM)*, Nevsehir, Turkey, 2019, pp. 68-73.

[6] A. Vannucci, G. Colavolpe, and L. Veltri, "Estimation of Correlated Gaussian Samples in Impulsive Noise," *IEEE Communications Letters*, vol. 24, no. 1, pp. 103-107, Jan. 2020.

[7] J. A. Cortes, A. Sanz, P. Estopinan, and J. I. Garcia, "On the suitability of the Middleton class A noise model for narrowband PLC," *2016 International Symposium on Power Line Communications and its Applications (ISPLC)*, Bottrop, 2016, pp. 58-63.

[8] T. Shongwe, A. J. H. Vinck, and H. C. Ferreira, "A Study on Impulse Noise and Its Models," *SAIEE Africa Research Journal*, vol. 106, no. 3, pp. 119-131, Sept. 2015.
[9] M. S. Alam, G. Kaddoum, and B. L. Agba, "Bayesian MMSE Estimation of a Gaussian Source in the Presence of Bursty Impulsive Noise," *IEEE Communications Letters*, vol. 22, no. 9, pp. 1846-1849, Sept. 2018.
[10] T. Minka, "Divergence measures and message passing," Tech. Report, MSR-TR-2005-173, Microsoft Research Ltd., Cambridge (UK), Jan. 2005.