Estimation of Correlated Gaussian Samples in Impulsive Noise

Estimation of Correlated Gaussian Samples in Impulsive Noise

Armando Vannucci, Giulio Colavolpe, and Luca Veltri

***Abstract*—We consider the estimation of correlated Gaussian samples in (correlated) impulsive noise, through message-passing algorithms. The factor graph includes cycles and, due to the mixture of Gaussian (samples and noise) and Bernoulli variables (the impulsive noise switches), the complexity of messages increases exponentially. We first analyze a simple but suboptimal solution, called Parallel Iterative Scheduling. Then we implement both Expectation Propagation — for which numerical stability must be addressed — and a simple variation thereof (called Transparent Propagation) that is inherently stable and simplifies the overall computation. Both algorithms reach a performance close to ideal, practically coinciding with the lower bound on the mean square estimation error.**

***Index Terms*—Factor graphs; Impulsive noise; Variational Bayesian inference.**

I. Introduction

The estimation of a Gaussian source from noisy observations is a classical task, yet a challenging one when the additive noise is non-Gaussian. In power-line communication (PLC) systems, as in other environments subject to electromagnetic interference, impulsive noise is often the dominant impairment, with strong power fluctuations occurring in bursts, such that the sequence of noise samples has memory. In a recent work [1], the simplified two-state Markov process of [2] is adopted to model bursty impulsive noise affecting a memoryless Gaussian source. From the system model viewpoint, the introduction of memory in the noise source brings about an augmentation of the state-space, that makes the signal estimation problem much more difficult. In [1], signal samples are estimated in a two-step process: a forward/backward message-passing algorithm (BCJR) [3] is applied after detecting the binary impulsive noise states of the channel. When the Gaussian source has memory, however, this strategy cannot be pursued. In [4], the observed signal sequence is assumed to follow a simple autoregressive model and the resulting estimation problem is modelled with a factor graphs (FG) that includes cycles, so that exact inference through the sum-product algorithm (SPA) [3] is not feasible. Moreover, the joint presence of discrete (the impulsive noise switches) and continuous (signal) variables generates messages with exponentially growing complexity, so that even approximate inference through the standard SPA is intractable [5].

We adopt here the same bursty impulsive noise model in [1], assuming in general that the signal source has memory, as in [4], and resorting to approximate variational inference techniques for the joint estimation of signal samples and impulsive noise channel states. First, we discuss an approximate solution, called *parallel iterative scheduling* (PISch) [4], that is based on hard decisions on the impulsive noise states, thus extending the results of [1] to a much broader scenario. The autoregressive model for the source signal that we consider here (or, even worse, the memoryless model of [1]) is of course too simplistic for a real PLC channel. Modulation formats like OFDM and long-memory channel taps might give rise to densely connected FGs [6] for which more powerful approximate inference algorithms are demanded, such as GAMP (generalized approximate message passing) [7]. The AR(1) model that we adopt in the following can however be straightforwardly generalized to AR(n), still using the same FG.

Although the considered system model belongs to the general class of *switching linear dynamic systems* (SDLS) [5], it falls within a special subclass for which the related FG has remarkable symmetry properties that can be exploited as we discuss. In fact, whenever either the noise sequence or the signal sequence is memoryless, the problem degenerates into the classical Kalman smoother or into the BCJR algorithm (the second case occurs in [1]). These are — in terms of FGs and involved messages — the mirror image of each other, so that PISch is a proper merging of them.

Taking hard decisions on impulsive noise states is however a suboptimal approach [8] that limits performance in both system models in [1], [4] as well as in a furher generalized Markov-Middleton noise model [9]. We then show that a smarter variational inference technique that exploits soft information, like expectation propagation (EP) [10], is able to reach a performance close to optimality, despite numerical stability problems. In order to overcome them, we propose an alternative to EP, based on a projection of individual messages, called transparent propagation (TP), that is shown to achieve the same results of EP, close to optimality, in a more effective way, involving less computations and ensuring numerical stability.

II. System model and Factor Graph

We consider the observation of a frame of K samples $\{y_k\}$ obtained by the following channel model:

$$\begin{aligned} y_k &= s_k + n_k^{\mathrm{G}} + i_k n_k^{\mathrm{I}} \qquad (k = 0, 1, \cdots, K-1) \\ s_k &= a_1 s_{k-1} + w_k \end{aligned} \tag{1}$$

where the signal sequence $\mathbf{s}$, obtained by filtering $\mathbf{w}$ through a stable single-pole filter, forms an *autoregressive model of*

The authors are with the Department of Engineering and Architecture, Università di Parma, 43124 Parma, Italy, email: armando.vannucci@unipr.it.

order one (AR(1)). The signal is observed through a sequence $\mathbf{y}$, that is affected by background noise n_k^{G} only or by extra additive impulsive noise n_k^{I}, depending on the binary values $i_k \in \{0,1\}$ of a *two-state Markov process* $\mathbf{i}$. The sequences $\mathbf{w}$, $\mathbf{n}^{\mathrm{G}}$ and $\mathbf{n}^{\mathrm{I}}$ are independent of each other and are made of independent and identically distributed (i.i.d.) real Gaussian samples with zero mean: $n_k^{\mathrm{G}} \sim \mathcal{N}(0,\sigma_{\mathrm{G}}^2)$; $n_k^{\mathrm{I}} \sim \mathcal{N}(0,\sigma_{\mathrm{I}}^2)$; $w_k \sim \mathcal{N}(0,(1-a_1^2)\sigma_s^2)$, where the signal variance σ_s^2 is taken as a reference.[1]

Depending on i_k, the impulsive noise channel can be in a "good" condition (only background noise n_k^{G}) or in a "bad" one. In the "bad" channel condition, the sum of background plus impulsive noise accounts for a single, more powerful, noise sequence $n_k^{\mathrm{B}} = n_k^{\mathrm{G}} + n_k^{\mathrm{I}} \sim \mathcal{N}(0,\sigma_{\mathrm{B}}^2)$, with variance $\sigma_{\mathrm{B}}^2 = \sigma_{\mathrm{G}}^2 + \sigma_{\mathrm{I}}^2 = (R+1)\sigma_{\mathrm{G}}^2$, where $R \triangleq \sigma_{\mathrm{B}}^2/\sigma_{\mathrm{G}}^2$ is often much larger than one.

The Markov process $\mathbf{i}$ is characterized by the $[2\times2]$ *one-step transition probabity matrix* Π, with entries $\pi_{r,c} = P\{i_{k+1} = c-1 \mid i_k = r-1\}$, $(r,c \in \{1,2\})$. All of the properties of the sequence $\{i_k\}$ can be derived from Π. In particular, the probability that the channel is in the "bad" condition is $p_{\mathrm{B}} = P\{i_k = 1\} = \gamma\pi_{12}$, where, in agreement with [2], we defined the parameter $\gamma \triangleq (\pi_{12}+\pi_{21})^{-1}$ that quantifies the memory of the Markov process. In fact, the average duration of 1's sequences, i.e., the duration of impulsive noise events, is $T_B = \gamma/p_{\mathrm{G}}$, whereas it would be $T_B = 1/p_{\mathrm{G}}$ for a memoryless process, where $p_{\mathrm{G}} = 1 - p_{\mathrm{B}}$.

Signal estimation through the SPA [3] requires the expression of the joint probability distribution function[2] of the signal samples $\mathbf{s} = \{s_k\}$ and parameters $\mathbf{i} = \{i_k\}$, given the observed samples $\mathbf{y} = \{y_k\}$. Thus, the conditional marginals of s_k and i_k can be employed for their minimum mean square error (MMSE) or maximum a posteriori (MAP) estimation.

Following the usual approach and disregarding the constant $p(\mathbf{y})$ of the observations, we seek to evaluate the marginals of

$$\begin{aligned} p(\mathbf{s},\mathbf{i}\mid\mathbf{y}) &\propto p(\mathbf{s},\mathbf{i},\mathbf{y}) = p(\mathbf{y}\mid\mathbf{s},\mathbf{i})p(\mathbf{s})P(\mathbf{i}) \qquad (2)\\ &= \left[\prod_{k=1}^{K-1} p(y_k\mid s_k,i_k)p(s_k\mid s_{k-1})P(i_k|i_{k-1})\right]\\ &\quad \times p(y_0\mid s_0,i_0)p(s_0)P(i_0) \end{aligned}$$

The FG representing the joint distribution in (2) is sketched in Fig. 1, where the k-th stage is highlighted along with the labels of the factor-to-variable node messages.

Tab. I shows the expression of messages according to the rules of the SPA [3], where subscripts "u,d,f,b" denote their (up, down, forward, backward) direction in Fig. 1. Denoting by $g(x-\eta,\sigma^2)$ a Gaussian pdf with mean η and variance σ^2, the conditional probabilities that appear in Tab. I are:

$$\begin{aligned} p(y_k\mid s_k,i_k) &= g(y_k - s_k, \sigma_{\mathrm{G}}^2 + i_k\sigma_{\mathrm{I}}^2) \qquad (3)\\ p(s_k\mid s_{k-1}) &= g\left(s_k - a_1 s_{k-1}, (1-a_1^2)\sigma_s^2\right) \qquad (4)\\ P(i_k\mid i_{k-1}) &= \pi_{(i_{k-1}+1),(i_k+1)} \qquad (5) \end{aligned}$$

[1]The extension to the case of complex processes and/or processes with possible nonzero mean value is straightforward.

[2]We use the notation $P(\cdot)$ to identify a probability mass function (pmf) for a discrete random variable and $p(\cdot)$ to denote a probability density function (pdf) or a continuous distribution with some discrete probability masses.

Table I

FACTOR- TO VARIABLE-NODE MESSAGES FOR THE SPA APPLIED TO THE FG IN FIG. 1 (SUBSCRIPTS DENOTE MESSAGE DIRECTION).

$$\begin{aligned} &p_{\mathrm{f}}(s_k) = \textstyle\int p_{\mathrm{f}}(s_{k-1})p_{\mathrm{u}}(s_{k-1})p(s_k\mid s_{k-1})ds_{k-1}\\ &P_{\mathrm{f}}(i_k) = \textstyle\sum_{i_{k-1}} P_{\mathrm{f}}(i_{k-1})P_{\mathrm{d}}(i_{k-1})P(i_k\mid i_{k-1})\\ &\qquad\qquad (k = 1,\cdots,K-1)\\ &p_{\mathrm{b}}(s_k) = \textstyle\int p_{\mathrm{b}}(s_{k+1})p_{\mathrm{u}}(s_{k+1})p(s_{k+1}\mid s_k)ds_{k+1}\\ &P_{\mathrm{b}}(i_k) = \textstyle\sum_{i_{k+1}} P_{\mathrm{b}}(i_{k+1})P_{\mathrm{d}}(i_{k+1})P(i_{k+1}\mid i_k)\\ &\qquad\qquad (k = K-2,\cdots,0)\\ &p_{\mathrm{u}}(s_k) = \textstyle\sum_{i_k} P_{\mathrm{f}}(i_k)P_{\mathrm{b}}(i_k)p(y_k\mid s_k,i_k)\\ &P_{\mathrm{d}}(i_k) = \textstyle\int p_{\mathrm{f}}(s_k)p_{\mathrm{b}}(s_k)p(y_k\mid s_k,i_k)ds_k\\ &\qquad\qquad (k = 0,\cdots,K-1) \end{aligned}$$

where (3) models the pdf of the observed samples, with variance $\sigma_{n,k}^2 = \sigma_{\mathrm{G}}^2 + i_k\sigma_{\mathrm{I}}^2 \in \{\sigma_{\mathrm{G}}^2,\sigma_{\mathrm{B}}^2\}$, while (4)-(5) model the one-step transition probabilities of the Markov processes s_k and i_k. In Fig. 1, the factor nodes $p(s_0) = p_{\mathrm{f}}(s_0) = g(s_0,\sigma_s^2)$ and $P(i_0) = P_{\mathrm{f}}(i_0) = p_{\mathrm{G}}\delta(i_0) + p_{\mathrm{B}}\delta(i_0-1)$ set the prior distributions of the initial samples, for $k=0$ in Tab. I, while $p_{\mathrm{b}}(s_{K-1}) = 1$ and $P_{\mathrm{b}}(i_{K-1}) = 1$ are the identity messages to the last variable nodes.

III. MESSAGE PASSING AND PARALLEL ITERATIVE SCHEDULING

Contrary to [1], the system considered here is modelled by a FG including cycles. Therefore, the estimation of the channel state i_k cannot be solved by a forward-backward message-passing algorithm. Indeed, the system (1) falls within the broad category of SDLS, for which exact inference is computationally intractable [5]. The basic reason is the joint presence of two time-series of latent variables of different nature, the discrete switches i_k and the continuous (Gaussian) samples s_k, which generates *Gaussian mixture* messages with complexity increasing exponentially with time [5]. In our case, for instance, each message $p_{\mathrm{u}}(s_k)$ is a linear combination of two Gaussians (with identical mean y_k and different variances $\sigma_{\mathrm{G,B}}^2$), weighted by the masses of the message/pmf $P_{\mathrm{f}}(i_k)P_{\mathrm{b}}(i_k)$. Propagation of $p_{\mathrm{u}}(s_k)$ along the upper line of the FG thus makes the number of Gaussian components, in the mixture messages $p_{\mathrm{f,b}}(s_k)$ of Tab. I, double at each time step. The general approach to perform approximate variational inference, in these cases, is to evaluate the message with increased (squared) complexity, at each time step, and collapse it into a simpler distribution (e.g., a mixture with a limited number of components), as done in the *assumed density filtering* methods [5], [10].

There is, however, one remarkable feature of our system model that makes it more tractable than the general SLDS: the impact of the switches i_k is only on the noise affecting the observations y_k and not on the evolution of the signal samples. This feature is clearly reflected onto the structure of the FG in Fig. 1, where the time-series $\mathbf{s}$ and $\mathbf{i}$ are *unconditionally independent* (although, as per (2), they are conditionally dependent, given $\mathbf{y}$). It is this feature that allowed the simple algorithmic solution known as *parallel iterative scheduling* (PISch) in [4] to be applied to the present problem.

The rationale behind PISch can be easily illustrated starting from the system discussed in [1], to which our system reduces

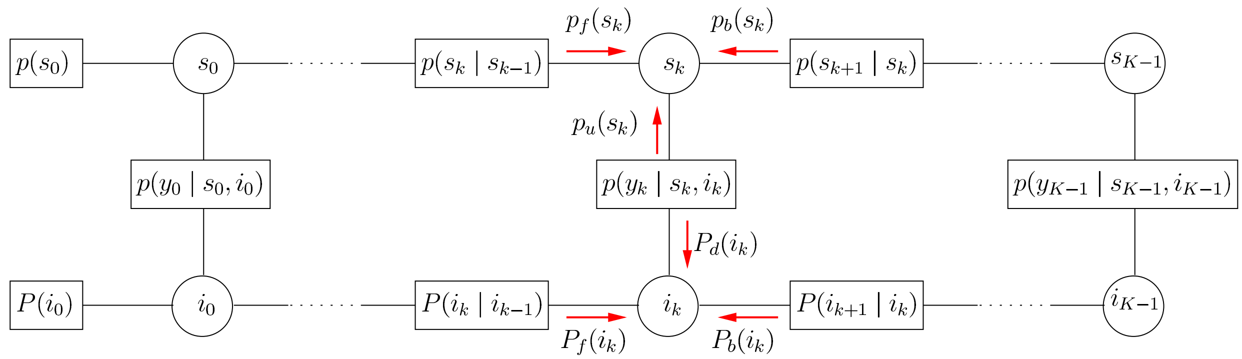

Figure 1. FG modelling the joint probability distribution function $p(\mathbf{s}, \mathbf{i} \mid \mathbf{y})$. The k-th stage is highlighted along with messages sent from factor nodes to variable nodes. The initial ($k = 0$) stage is driven by the priors, $P_{\mathrm{f}}(i_0) = P(i_0)$ and $p_{\mathrm{f}}(s_0) = p(s_0)$, while $P_{\mathrm{b}}(i_{K-1}) = p_{\mathrm{b}}(s_{K-1}) = 1$ for the last ($k = K - 1$) stage.

when $a_1 = 0$ in (1). In this case, the signal samples s_k are independent of each other and are characterized by their known prior distributions $\widetilde{p}(s_k)$. For the i.i.d. sequence $\mathbf{s}$ assumed in [1], these are all equal to the same Gaussian prior but, no matter if the priors $\widetilde{p}(s_k)$ are all equal or not, still an optimal estimation of the switching sequence $\mathbf{i}$ is realized by the BCJR algorithm [1]. This is known to be an instance of the SPA [3, Sec. IV-A], as applied to the FG in [3, Fig. 12(d)], which is equivalent to the lower half of the FG in Fig. 1, when the s_k are independent. After the forward-backward iteration of the BCJR is complete, a *hard decision* is taken on $i_k = m \quad (m \in \{0, 1\})$, so that its pmf is approximated as a deterministic $\widetilde{P}(i_k) = \delta(i_k - m)$. Suppose, on the contrary, to have a *genie aided receiver* that provides the true values of i_k as a *channel state information* (CSI). The system in (1) would thus reduce to a correlated sequence of Gaussian samples (s_k), to be estimated in the presence of (possibly non-stationary) additive white Gaussian noise (AWGN) with known variance per sample. This is the classical problem of Kalman smoothing, that is again implemented as an instance of the SPA [3, Sec. IV-C], as applied to the cycle-free FG in [3, Fig. 15], to which the FG in Fig. 1 reduces in the case of deterministic i_k. Its outcomes are the optimal Gaussian estimates $\widetilde{p}(s_k)$ for each sample.

The strategy of PISch relies on an iteration of the above two forward/backward algorithms: the BCJR and the Kalman smoother are implemented (in parallel) by the SPA passing messages 'horizontally' along the upper and lower halves of the FG in Fig. 1. After a forward/backward pass is completed, their outputs, i.e., the (temporary) hard decisions $P_{\mathrm{u}}(i_k) \triangleq P_{\mathrm{f}}(i_k)P_{\mathrm{b}}(i_k) = \widetilde{P}(i_k) = \delta(i_k - m)$ and the Gaussian estimates $p_{\mathrm{d}}(s_k) \triangleq p_{\mathrm{f}}(s_k)p_{\mathrm{b}}(s_k) = \widetilde{p}(s_k)$, are sent as 'vertical' messages (variable- to factor-node, not shown in Fig. 1) to the other FG half.

IV. Expectation Propagation (EP) and Transparent Propagation (TP)

The results of the PISch algorithm [4], as further shown in Sec. V, are however limited by the hard decisions on i_k, that impose a too strict constraint on the messages $P_{\mathrm{u}}(i_k)$. In fact, as pointed out in [8], the adoption of a hard decision on the state of impulsive noise, followed by the MMSE estimation of the source signal, results in a suboptimal strategy. This is true both when applying the single-pass solution in [1] and during the PISch iterations in [4]. In order to exploit the soft information provided by the lower half of the FG in Fig. 1, we can replace the marginal densities that arise from the Gaussian mixtures $p_{\mathrm{u,f,b}}(s_k)$ discussed above by their 'nearest' *projection* onto a given *approximating (exponential) family*. If the metrics to compute the 'nearest' pdf $q(s_k)$ is the Kullback-Leibler divergence $KL(p \parallel q)$, this approach reduces to the EP algorithm [10].

According to the rules of EP, the message $p_{\mathrm{u}}(s_k)$ in Tab. I (here renamed $p_{\mathrm{u}}^{\mathrm{SPA}}(s_k)$ for clarity), that is the one responsible for the increase in message complexity, is substituted by

$$p_{\mathrm{u}}^{\mathrm{EP}}(s_k) \quad = \quad \frac{1}{p_{\mathrm{d}}(s_k)} \mathrm{proj}\left[p_{\mathrm{d}}(s_k) p_{\mathrm{u}}^{\mathrm{SPA}}(s_k)\right] \tag{6}$$

where the projection operation is implemented as a *matching of the expectations* [10].

Since our problem involves Gaussian priors for s_k and Gaussian noise for the observations y_k, the Gaussian approximating family seems a natural choice, when computing the above projection. The resulting gaussianity of $p_{\mathrm{u}}^{\mathrm{EP}}(s_k)$ further implies that all of the messages $p_{\mathrm{f,b}}(s_k)$ in the upper half of the FG in Fig. 1 are themselves Gaussians, hence $p_{\mathrm{d}}(s_k) \triangleq p_{\mathrm{f}}(s_k)p_{\mathrm{b}}(s_k) = g(s_k - \eta_{s,k}, \sigma_{s,k}^2)$ is Gaussian too. Therefore, no other message than (6) needs projecting onto the approximating family, with consistent time savings during the algorithm. We shall limit the application of EP with a Gaussian approximating family to the upper half of the FG, that involves continuous variables s_k, while messages to and from the discrete variables i_k are still modelled by Bernoulli pmfs. Using standard Gaussian product rules, the projection operation in (6) is performed by matching the *moment parameters*, as prescribed by EP, to produce a Gaussian result, $\mathrm{proj}\left[p_{\mathrm{d}}(s_k)p_{\mathrm{u}}^{\mathrm{SPA}}(s_k)\right] = g(s_k - \widehat{s}_k, \widehat{\sigma}_k^2)$, where $\widehat{s}_k$ is the signal estimate produced by EP. Hence, $p_{\mathrm{u}}^{\mathrm{EP}}(s_k)$ is a ratio of Gaussian pdfs hence is Gaussian too, provided that the necessary condition $\widehat{\sigma}_k^2 < \sigma_{s,k}^2$ is satisfied.

However, when the variance $\widehat{\sigma}_k^2$ of the numerator in (6), i.e., that of the product between $p_{\mathrm{u}}^{\mathrm{SPA}}(s_k)$ and $p_{\mathrm{d}}(s_k)$, exceeds the variance $\sigma_{s,k}^2$ of $p_{\mathrm{d}}(s_k)$ alone (which could not occur

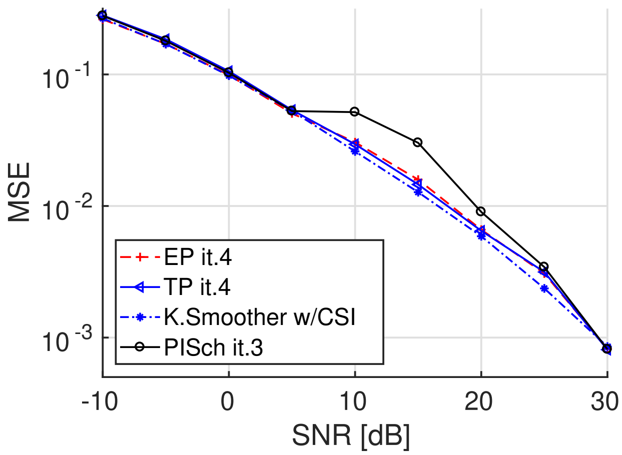

Figure 2. Comparison between Parallel Iterative Scheduling, Expectation Propagation and Transparent Propagation algorithms, at convergence. The experimental lower bound (Kalman smoother with CSI) is reported for comparison. 10^2 frames of 10^3 samples each have been considered, for each SNR value.

if $p_{\mathrm{u}}^{\mathrm{SPA}}(s_k)$ were a simple Gaussian), the division operation leads to an *improper distribution* for $p_{\mathrm{u}}^{\mathrm{EP}}(s_k)$, i.e., one with a negative variance. This is a well known problem of EP, that generates numerical instabilities and hinders the convergence of the algorithm [11].

An alternative solution, to circumvent the problem of numerical stability in EP, is to implement a message-passing algorithm that directly projects onto the selected Gaussian family only the mixture $p_{\mathrm{u}}^{\mathrm{SPA}}(s_k)$. This way, the EP computation rule (6) is substituted by a different one, i.e., by the following

$$p_{\mathrm{u}}^{\mathrm{TP}}(s_k) \quad = \quad \mathrm{proj}\left[\sum_{i_k} P_{\mathrm{f}}(i_k)P_{\mathrm{b}}(i_k)p(y_k \mid s_k, i_k)\right] \quad (7)$$

as if the message $p_{\mathrm{d}}(s_k)$ were extracted from the projection operation in (6) and simplified with the denominator, thus avoiding the critical division operation. This operation is not allowed in EP, since it is not the individual messages that are constrained to belong to the exponential approximating family, rather it is the estimated marginals $\widetilde{p}(s_k \mid \underline{y}) \propto p_{\mathrm{u}}(s_k)p_{\mathrm{d}}(s_k)$. In our problem, however, the Gaussian projection in (7) ensures the Gaussianity of $p_{\mathrm{d}}(s_k)$ as well as that of the estimated pdf $\widetilde{p}(s_k \mid \underline{y}) \propto p_{\mathrm{u}}^{\mathrm{TP}}(s_k)p_{\mathrm{d}}(s_k)= g(s_k - \widehat{s}_k, \widehat{\sigma}_k^2)$ (where $\widehat{s}_k$ is the related signal estimate, for the new TP algorithm), and of any other message sent through the edges of the upper half of the FG in Fig. 1.

This is a simple and novel approach, that we call *Transparent Propagation* (TP), that can be applied to a whole class of problems, including the present one. It should be noted that, despite the apparently minimal symbolic differences between (6) and (7) (and the fifth eq. in Tab. I), EP and TP implement different paradigms, due to the different application of the projection operator (we defer the reader to [12] for a thorough discussion of its properties and implementation). In EP, projection introduces an approximation onto the posterior conditional estimates of variables, while in TP it is the individual messages that are projected onto the selected (here Gaussian) approximating family. We show next that, at least in the present problem, TP achieves the same virtues as EP (constrained pdfs) with less computation, and guarantees numerical stability.

V. Simulation Results

Fig. 2 shows the MSE $E\left[e_k^2\right]$ of the signal estimate obtained by the algorithms discussed in Secs. III and IV, where the error signal is $e_k \triangleq s_k - \widehat{s}_k$. We considered the system parameters $a_1 = 0.9$ and $\sigma_s^2 = 1$, for the AR(1) signal samples, while we set $\gamma = 100$, $p_{\mathrm{B}} = 0.1$, and $R = 100$, for the impulsive noise with memory.

As seen in Fig. 2, the performance of PISch tends to follow a *waterfall* shape, with a breakpoint (around $SNR = 5\,dB$, for the present choice of system parameters) where the slope changes. This is not a numerical artifact, but rather an intrinsic feature of a two-states system, that is implied by the hard decisions taken on i_k.[3] As a result, the PISch algorithm in [4] significantly deviates from the ideal performance.

As a reference, Fig. 2 reports the MSE of a Kalman Smoother with perfect CSI on the impulsive noise states, which implements an optimal *genie aided estimator* that knows the noise statistics at each time epoch, hence acts as an experimental lower bound.

As opposed to the simpler PISch algorithm, which relies on 'hard messages', i.e., hard decisions on the impulsive noise swithes i_k, both EP and TP rely on a Gaussian projection of the (mixture) messages containing soft information on the channel state, so that their performance is inherently superior. This is especially true at moderate-to-high SNR values, where the benefits of exchanging soft information are more evident. In fact, as seen in Fig. 2 , the results of both EP and TP practically coincide with the lower bound, hence are close to optimality, with a maximum deviation of 1.25 dB (at $SNR = 25$ dB).

In addition to the results in Fig. 2, we implemented other simple two stages estimation strategies (e.g., a threshold detection of impulsive noise, followed by a Kalman filter or smoother). Results are not reported here since, far from being optimal, they are outperformed by EP and TP. In addition, their performance is worse than the PISch algorithm at its second iteration, which corresponds to a BCJR impulsive noise state estimation followed by a Kalman smoother for signal estimation (see Sec. III). When the signal process is memoryless, this coincides with [1].

Regarding convergence, PISch quickly converged in 3 iterations while, as discussed in Sec. IV, numerical instabilities caused by improper distribution prevented the convergence of the EP algorithm. At the first iteration, negative variances occurred for about 2% of samples s_k and propagated their destructive effect on signal estimates so as to spoil the convergence of the algorithm. To ensure the numerical stabilization of EP, we implemented the simple accept/reject strategy suggested in [11]. This strategy prescribes that the estimation updates of a message/pdf are rejected, if the *moment*

[3] A theoretical justification can be given for this behavior, which however falls outside the scope of the present work.

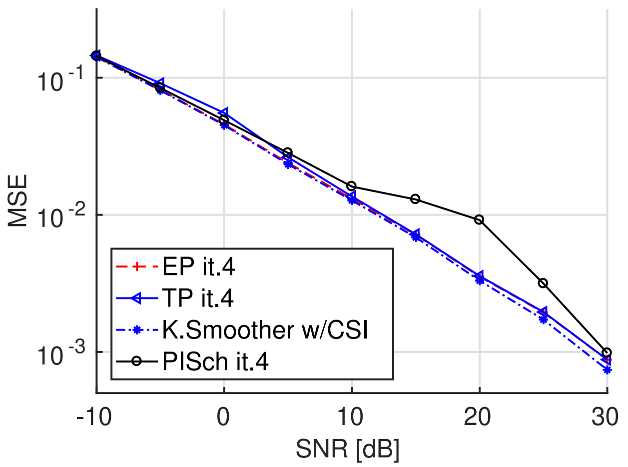

Figure 3. Comparison between the investigated algorithms, at convergence, in a *strongly impulsive* noise scenario ($p_B = 0.03$, $R = 1000$).

parameters fall outside their allowed region, and substituted with a message identical to that of the previous iteration (an identity message $p_{\rm u}^{\rm EP}(s_k) = 1$ is sent if rejection occurs at the first iteration). The accept/reject strategy thus implemented made the percentage of improper message rejections drop from 2.22%, at the first iteration, to 0.46% at the second, then settling to 0.08%, at the following iterations, possibly due to the stabilization of messages, since the algorithm converged in 4 iterations. The TP algorithm, instead, is inherently stable and spontaneously converged in 4 iterations. All of the algorithms are stable, and the MSE curves did not change appreciably after convergence (as we numerically checked, until the $10th$ iteration).

Since a careful application of EP to the present problem yields a (close-to-)optimal performance, no further improvement in performance can be expected from the TP algorithm. The purpose of introducing TP here is to demonstrate a novel, viable alternative to EP, that is inherently stable and that performs equally well, at least with the system considered here.

A. Further Simulation Results

In order to support our conclusions with more numerical examples, we considered two other scenarios, with different system parameters, that are opposite with respect to the one in Fig. 2. The first is that of *weakly impulsive* noise, characterized by frequent ($p_B = 1/3$) "bad" channel states with noise power comparable to that of background noise ($R = 4$); the other system parameters are as in Sec. V. We found that all of the considered algorithms reached an optimal performance in four iterations (despite EP still needed improper message rejections in 0.29% of cases), hence we do not report the MSE graphically. As a matter of fact, when noise is *weakly impulsive*, any decent estimation algorithm performs well, although not designed for impulsive noise.

The other scenario is instead that of *strongly impulsive* noise, characterized by rare ($p_B = 0.03$) "bad" states with much larger power ($R = 1000$). The MSE reported in Fig. 3 shows a behavior qualitatively similar to that of Fig. 2, so that the same conceptual conclusions apply. The PISch algorithm took 4 iterations to converge and its loss in performance is still considerable. As expected, EP and TP converged in four iterations to the optimal performance of a genie-aided estimator, where EP still showed instabilities to be handled (0.08% improper messages) while TP required less computation and was inherently stable.

VI. Conclusions

The estimation of a correlated Gaussian sequence affected by bursty impulsive noise, considered in this work, is a problem of remarkable complexity and symmetry. As typical of systems with mixed discrete and continuous variables, the direct implementation of the sum-product algorithm is computationally intractable, so that approximate solutions are demanded. We implemented the recently proposed Parallel Iterative Scheduling (PISch) algorithm [4], that achieves complexity reduction by taking a hard decision on the impulsive noise state, as in [1], that is shown to lead to suboptimal performance. Expectation Propagation (EP), instead, is shown to achieve a performance close to optimality, although with numerical stability problems. We finally proposed a variation on EP, called Transparent Propagation (TP), that achieves the same performance as EP, with lower complexity and avoiding numerical instabilities.

References

[1] M. S. Alam, G. Kaddoum and B. L. Agba, "Bayesian MMSE estimation of a Gaussian source in the presence of bursty impulsive noise," in *IEEE Communications Letters*, vol. 22, no. 9, pp. 1846-1849, Sept. 2018.

[2] D. Fertonani and G. Colavolpe, "On reliable communications over channels impaired by bursty impulse noise," *IEEE Trans. Commun.*, vol. 57, no 7, pp. 2024-2030, July 2009.

[3] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Tr. on Inf. Th.*, vol. 47, no 2, pp. 498-519, Feb. 2001.

[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), Praha, Czech Republic, Apr. 2019, pp. 1-6.

[5] D. Barber and A. T. Cemgil, "Graphical models for time series," *IEEE Signal Process. Mag.*, vol. 27, pp. 18-28, Nov. 2010.

[6] M. Nassar, P. Schniter and B. L. Evans, "A Factor Graph Approach to Joint OFDM Channel Estimation and Decoding in Impulsive Noise Environments," *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1576-1589, Mar. 2014.

[7] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *2011 IEEE International Symposium on Information Theory Proceedings*, St. Petersburg, 2011, pp. 2168-2172.

[8] B. Dulek, "Comment on "Bayesian MMSE estimation of a Gaussian source in the presence of bursty impulsive noise"," in *IEEE Communications Letters*, vol. 23, no. 4, pp. 772-772, Apr. 2019.

[9] A. Mirbadin, E. Kiani, A. Vannucci and G. Colavolpe, "Estimation of Gaussian processes in Markov-Middleton impulsive noise," to appear in *2019 IEEE Global Power, Energy and Communication Conference* (GPECOM), Nevsehir, Turkey, June 2019.

[10] T. P. Minka, "Expectation propagation for approximate Bayesian inference," *Proc. of the Seventeenth conference on Uncertainty in artificial intelligence* (UAI'01), Seattle (USA), pp. 362–369, Aug. 2001.

[11] M. Seeger, "Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations", *PhD thesis, University of Edinburgh*, July 2003.

[12] T. Minka, "Divergence measures and message passing", Tech. Report, Microsoft Research Ltd., Cambridge (UK), Jan. 2005.