Joint Robust Linear Regression and Anomaly Detection in Poisson noise using Expectation-Propagation

D. Yao, Y. Altmann*, S. McLaughlin
School of Engineering and Physical Sciences
Heriot-Watt University
Edinburgh, UK

M. E. Davies
School of Engineering
University of Edinburgh
Edinburgh, UK

***Abstract*—In this paper, we propose a new Expectation-Propagation (EP) algorithm to address the problem of joint robust linear regression and sparse anomaly detection from data corrupted by Poisson noise. Adopting an approximate Bayesian approach, an EP method is derived to approximate the posterior distribution of interest. The method accounts not only for additive anomalies, but also for destructive anomalies, i.e., anomalies that can lead to observations with amplitudes lower than the expected signals. Experiments conducted with both synthetic and real data illustrate the potential benefits of the proposed EP method in joint spectral unmixing and anomaly detection in the photon-starved regime of a Lidar system.**

***Index Terms*—Linear regression, Poisson noise, Anomaly detection, Approximate Bayesian inference, Expectation-Propagation**

I. Introduction

Anomaly detection plays an important role in robust linear regression problems. It is essential to identify and compensate for potential anomalies in real-world measurements since such anomalies can contain valuable information [1] and/or jeopardize the accurate estimation of the coefficients of the linear regression. The presence of random observation noise makes the task more challenging, especially when the noise is non-Gaussian. For instance, multispectral Lidar data detected in the low photon count regime [2], [3] is primarily contaminated by Poisson noise and robust spectral unmixing and anomaly detection from this kind of data remains difficult. The amplitude of anomalies can be higher or lower than the expected signals and the anomalies can be structured. In this work, we consider the case of sparse anomalies, which only corrupt a fraction of the observations.

Bayesian approaches enable uncertainty quantification about the unknown signals to be recovered as well as the sparse anomalies, and in addition to providing their support. This is classically achieved by defining a likelihood modelling the observation process and assigning prior distributions to the unknown model parameters. A full Bayesian treatment for imaging problems with Poisson data is challenging, due to the non-negativity constraint and high-dimensionality [4]. Markov chain Monte Carlo (MCMC) algorithms have been traditionally proposed to exploit the resulting posterior distribution. However, the sampling process implies a high computational cost and MCMC-based algorithms are not (yet) scalable for fast inference. Approximate Bayesian methods [5], [6] are attractive state-of-the-art alternative solutions which aim at approximating the actual posterior distribution by a simpler distribution whose moments are easier to compute with a much reduced computational cost compared with MCMC. This is the approach adopted here.

Expectation-Propagation (EP) [6] is an approximate Bayesian inference algorithm that often provides a faster and more accurate alternative than variational Bayes (VB) approaches. The approximation of EP is achieved by minimizing the information loss between the exact yet complex posterior distribution and the approximation, which is measured using a reversed Kullback-Leibler (KL) divergence, compared to VB [7]. The approximate distributions in EP are generally chosen from the exponential family to approximate factor nodes of the likelihood and of the prior model. Gaussian distributions are commonly chosen as approximations in EP (for continuous variables), while here we also use Bernoulli distributions for the binary variables. EP has already been successfully applied in the context of linear regression and more generally linear inverse problems in the presence of Poisson noise [4], [8], [9]. Approximating the Poisson observation model or likelihood in EP is more challenging than probit and finite Gaussian mixture likelihoods due to more challenging integrals [10]. The key to solving this problem is to transform the intractable integrals into integrals that are easier to solve. Data augmentation schemes by the introduction of appropriate auxiliary variables in EP [9] allow transforming multidimensional integrals over multivariate random variables into lower-dimensional problems that can be solved either analytically or efficiently using numerical integration. Note that while a generalized vector approximate message passing (V-AMP) [11], [12] could also be used here, the robust linear regression problem we consider here does not fit the large system limit assumption made by V-AMP and we therefore wish to preserve some posterior

This work was supported by the Royal Academy of Engineering under the Research Fellowship scheme RF201617/16/31.

correlations between parameters (e.g., the coefficients of the linear regression), which motivates the use of EP instead.

The remainder of this paper is organized as follows. In section II we describe the main difference between the proposed anomaly model and the anomaly model used in our previous work. The exact Bayesian model for joint robust linear regression and anomaly detection is constructed in Section III. Section IV presents the augmented factor graph used in the EP algorithm and the associated update rules for each factor. Experiments conducted on synthetic data and a multispectral Lidar data unmixing application are reported in Section V and the conclusions are finally reported in Section VI.

II. Related Work

In our previous work [13], we proposed an EP method for joint robust linear regression and anomaly detection in the presence of Poisson noise. The model used in [13] relies on the consideration of sparse, *additive* anomalies corrupting the signal, whereby anomalies, if present, contribute positively to the mean of the Poisson observation noise. More precisely, let $\boldsymbol{y} = [y_1, ..., y_M]^T$ be a random vector of M observations, with mean $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_M]^T$ and corrupted by Poisson noise such that $\boldsymbol{y}|\boldsymbol{\lambda} \sim \mathcal{P}(\boldsymbol{\lambda})$. The model in [13] assumes that $\boldsymbol{\lambda} = \mathbf{A}\boldsymbol{x} + \boldsymbol{v}$ where $\mathbf{A}$ is a positive and known $M \times N$ matrix with $N \leq M$, and $\boldsymbol{x} \in \mathbb{R}_+^N$ is the vector of coefficients of interest. Moreover, $\boldsymbol{v} = [v_1, ..., v_M]^T$ is a sparse and non-negative vector that represents the anomalies.

This sparse anomaly model has been used for robust non-negative matrix factorization (e.g., in [14]) and robust spectral unmixing (e.g., in [3]) as it can capture components present in the mixture, but not originally represented by the columns of $\mathbf{A}$. Moreover, this model naturally enforces the positivity of $\boldsymbol{\lambda}$ which is bounded below by $\mathbf{A}\boldsymbol{x}$. However, this model only accounts for additive anomalies which locally increase the mean of the Poisson observations. In specific applications, anomalies can violate the elementwise inequalities $\boldsymbol{\lambda} \geq \mathbf{A}\boldsymbol{x}$. Such a scenario with *destructive* anomalies might occur for instance in the presence of faulty detectors (e.g., a sensor providing completely erroneous readings at random times or some faulty sensors in an array) or errors when specifying $\mathbf{A}$, i.e., the mean of the Poisson distribution can be $\mathbf{A}\boldsymbol{x}$ or anomalies $\boldsymbol{v}$. In such cases, the model used in [13] (which uses a Bernoulli-Exponential prior model for $\boldsymbol{v}$) is not expected to yield accurate regression nor anomaly detection results.

To address this issue, we consider an alternative model and associated approximate Bayesian estimation procedure based on EP is described in the next section.

III. Bayesian Model

In this work, we model $\boldsymbol{\lambda}$, the mean of Poisson likelihood $f(\boldsymbol{y}|\boldsymbol{\lambda})$, as

$$\boldsymbol{\lambda} = (\mathbf{1} - \boldsymbol{z}) \odot (\mathbf{A}\boldsymbol{x}) + \boldsymbol{z} \odot \boldsymbol{r}, \tag{1}$$

where $\boldsymbol{z} = [z_1, ..., z_M]^T \in (0;1)^N$ is a binary vector encoding the support of the anomalies, $\boldsymbol{r} = [r_1, ..., r_M]^T$ represents the (positive) amplitudes of anomalies and $\odot$ is the Hadamard (element-wise) product. With this model, $\lambda_m = r_m$ if an anomaly is corrupting the m-th observation ($z_m = 1$), and $\lambda_m = \boldsymbol{a}_m\boldsymbol{x}$ if $z_m = 0$, where $\boldsymbol{a}_m$ is the m-th row of $\mathbf{A}$. Our goal here is to recover $(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z})$ from $\boldsymbol{y}$, while also incorporating prior knowledge available about these parameters.

A. Poisson likelihood

Using (1), the likelihood can be expressed as

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z}) &= \mathcal{P}((\mathbf{1} - \boldsymbol{z}) \odot (\mathbf{A}\boldsymbol{x}) + \boldsymbol{z} \odot \boldsymbol{r}) \\ &= \prod_{m=1}^{M} \mathcal{P}((1 - z_m)(\boldsymbol{a}_m\boldsymbol{x}) + z_m r_m), \end{aligned} \tag{2}$$

where

$$\begin{cases} y_m|\boldsymbol{a}_m\boldsymbol{x}, r_m, z_m = 0 \sim \mathcal{P}(\boldsymbol{a}_m\boldsymbol{x}) \text{ (signal only)}, \\ y_m|\boldsymbol{a}_m\boldsymbol{x}, r_m, z_m = 1 \sim \mathcal{P}(r_m) \quad \text{(anomaly only)}. \end{cases} \tag{3}$$

B. Prior distributions

To complete our Bayesian model, we specify a prior model for $(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z})$. Such a model can be application specific and here we simply assume prior independence between $\boldsymbol{x}$, $\boldsymbol{r}$, and $\boldsymbol{z}$, i.e., $f(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z}) = f_{\boldsymbol{x}}(\boldsymbol{x}) f_{\boldsymbol{r}}(\boldsymbol{r}) f_{\boldsymbol{z}}(\boldsymbol{z})$. We further assume mutual prior independence between the elements of $\boldsymbol{x}$, those in $\boldsymbol{r}$, and those in $\boldsymbol{z}$. Without loss of generality, we use a product of independent exponential priors for $\boldsymbol{x}$ and $\boldsymbol{r}$, with mean $\lambda_{\boldsymbol{x}}^{-1}$ and $\lambda_{\boldsymbol{r}}^{-1}$, respectively. However, any separable model ensuring the positivity of $\boldsymbol{x}$ and $\boldsymbol{r}$ (e.g., gamma or truncated Gaussian distributions) could be used.

A product of M independent Bernoulli prior distributions with mean $\sigma(p_0) = \frac{1}{1+\exp(-p_0)}$ is assigned to $f_{\boldsymbol{z}}(\boldsymbol{z})$, i.e., $\sigma(p_0)$ is the prior probability of anomaly presence in each observation. However, more complex models, e.g., enforcing group-sparsity as in [2], [13] could also be used. This case is not further detailed here due to space constraints.

C. Posterior distribution

Using the likelihood (2) and the prior distributions $f_{\boldsymbol{x}}(\boldsymbol{x})$, $f_{\boldsymbol{r}}(\boldsymbol{r})$ and $f_{\boldsymbol{z}}(\boldsymbol{z})$, joint posterior distribution becomes

$$f(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z}) f_{\boldsymbol{x}}(\boldsymbol{x}) f_{\boldsymbol{r}}(\boldsymbol{r}) f_{\boldsymbol{z}}(\boldsymbol{z}). \tag{4}$$

Exploiting the joint posterior distribution in (4) is challenging, mainly because of the Poisson likelihood, the positivity constraints on $(\boldsymbol{x}, \boldsymbol{r})$ and the discrete nature of $\boldsymbol{z}$. As mentioned above, while MCMC methods could be used to sample from (4) (e.g., as in [3] which uses the model discussed in Section II), such methods still suffer from high computational cost and potentially slow convergence speed. An approximate Bayesian inference based technique is proposed in this paper to overcome this issue and achieve fast and scalable joint robust linear regression and anomaly detection.

IV. EP METHODS FOR JOINT ROBUST LINEAR REGRESSION AND ANOMALY DETECTION

A. Data augmentation in EP

EP methods aim at approximating a complex joint posterior distribution as in (4) by a simpler distribution, by approximating each of the individual factors $f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{r},\boldsymbol{z})$, $f_{\boldsymbol{x}}(\boldsymbol{x})$, $f_{\boldsymbol{r}}(\boldsymbol{r})$ and $f_{\boldsymbol{z}}(\boldsymbol{z})$ using simpler distributions. The families of approximating distributions can be arbitrary, but Gaussian (resp. Bernoulli) distributions are usually considered for continuous (resp. discrete) variables. Moreover, to simplify the estimation process, it is convenient to introduce auxiliary variables in a data augmentation fashion. In our problem, the main difficulty is the likelihood term which includes a Poisson distribution and the linear operator $\mathbf{A}$. This combination complicates the computation of means and covariance matrices during the EP algorithm. Thus, we introduce an auxiliary vector $\boldsymbol{u} = \mathbf{A}\boldsymbol{x}$, which is assigned a degenerate prior distribution $f(\boldsymbol{u}|\boldsymbol{x}) = \delta(\boldsymbol{u}-\mathbf{A}\boldsymbol{x}) = \prod_m \delta(u_m - \boldsymbol{a}_m\boldsymbol{x})$, where $\delta(\cdot)$ denotes the Dirac delta function. The resulting factor graph is depicted in Fig. 1 and the resulting posterior distribution with data augmentation becomes $f(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},\boldsymbol{z}|\boldsymbol{y})$. This factor graph also shows the approximation factors in green. For instance, the factor node $\delta(\boldsymbol{u}-\mathbf{A}\boldsymbol{x})$ is approximated by a product of two (unnormalized) multivariate Gaussian densities ($q_{\boldsymbol{u},0}(\boldsymbol{u})$ and $q_{\boldsymbol{x},1}(\boldsymbol{x})$). For completeness, we also include, in Fig. 2, the factor graph of the model from [13], which uses a different auxiliary vector $\boldsymbol{\lambda} = \mathbf{A}\boldsymbol{x} + \boldsymbol{v}$.

B. EP updates

EP aims at obtaining an approximation $Q(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},z)$ of the exact joint posterior distribution $f(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},z|\boldsymbol{y})$ by minimizing the reversed KL divergence

$$\min_{Q(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},\boldsymbol{z})} KL(f(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},\boldsymbol{z}|\boldsymbol{y})||Q(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},\boldsymbol{z})). \tag{5}$$

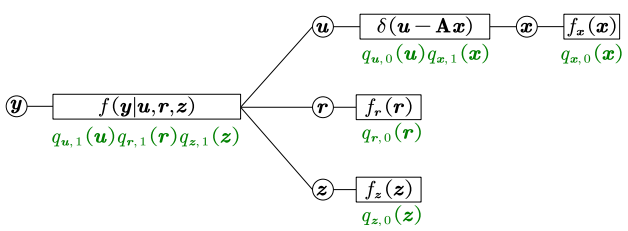


Fig. 1. Factor graph used to perform EP-based joint robust linear regression and anomaly detection using variable splitting. The rectangular boxes (resp. circles) represent the factor (resp. variable) nodes and the approximation factors are shown in green.

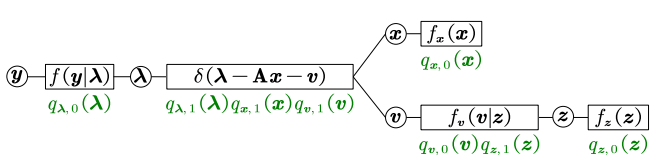


Fig. 2. Factor graph for EP approximation with data augmentation for additive anomaly model in [13].

However, computing the KL divergence in (5) directly is intractable as computing expectations with respect to the exact joint posterior is not feasible. Instead, EP methods aim at approximately solving (5) by iteratively minimizing the KL divergence between so-called tilted distribution and the global approximation, which corresponds to moment matching between the distributions when the approximation is a Gaussian. More specifically, using the factor graph from Fig. 1, the sequential EP updates become

$$\begin{cases} \min\limits_{Q_1(\boldsymbol{u},\boldsymbol{r},\boldsymbol{z})} & KL(f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{r},z)Q_0(\boldsymbol{u},\boldsymbol{r},z)||Q(\boldsymbol{u},\boldsymbol{r},z)), \\ \min\limits_{q_{\boldsymbol{u},0}(\boldsymbol{u})q_{\boldsymbol{x},1}(\boldsymbol{x})} & KL(\delta(\boldsymbol{u}-\mathbf{A}\boldsymbol{x})q_{\boldsymbol{u},1}(\boldsymbol{u})q_{\boldsymbol{x},0}(\boldsymbol{x})||Q(\boldsymbol{u})Q(\boldsymbol{x})), \\ \min\limits_{q_{\boldsymbol{x},0}(\boldsymbol{x})} & KL(f_{\boldsymbol{x}}(\boldsymbol{x})q_{\boldsymbol{x},1}(\boldsymbol{x})||Q(\boldsymbol{x})), \\ \min\limits_{q_{\boldsymbol{r},0}(\boldsymbol{r})} & KL(f_{\boldsymbol{r}}(\boldsymbol{r})q_{\boldsymbol{r},1}(\boldsymbol{r})||Q(\boldsymbol{r})), \\ \min\limits_{q_{\boldsymbol{z},0}(\boldsymbol{z})} & KL(f_{\boldsymbol{z}}(\boldsymbol{z})q_{\boldsymbol{z},1}(\boldsymbol{z})||Q(\boldsymbol{z})), \end{cases} \tag{6}$$

where $q_{\boldsymbol{u},i}(\boldsymbol{u})$, $q_{\boldsymbol{x},i}(\boldsymbol{x})$, $q_{\boldsymbol{r},i}(\boldsymbol{r})$ ($\forall i \in (0;1)$) are multivariate Gaussian densities, and $q_{\boldsymbol{z},i}(\boldsymbol{z})$ ($\forall i \in (0;1)$) are Bernoulli distributions. To simplify notations, we used

$$\begin{cases} Q_i(\boldsymbol{u},\boldsymbol{r},\boldsymbol{z}) = q_{\boldsymbol{u},i}(\boldsymbol{u})q_{\boldsymbol{r},i}(\boldsymbol{r})q_{\boldsymbol{z},i}(\boldsymbol{z}), \forall i \in (0;1), \\ Q_{\boldsymbol{u}}(\boldsymbol{u}) = q_{\boldsymbol{u},0}(\boldsymbol{u})q_{\boldsymbol{u},1}(\boldsymbol{u}), \\ Q_{\boldsymbol{x}}(\boldsymbol{x}) = q_{\boldsymbol{x},0}(\boldsymbol{x})q_{\boldsymbol{x},1}(\boldsymbol{x}), \\ Q_{\boldsymbol{r}}(\boldsymbol{r}) = q_{\boldsymbol{r},0}(\boldsymbol{r})q_{\boldsymbol{r},1}(\boldsymbol{r}), \\ Q_{\boldsymbol{z}}(\boldsymbol{z}) = q_{\boldsymbol{z},0}(\boldsymbol{z})q_{\boldsymbol{z},1}(\boldsymbol{z}), \\ Q(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},z) = Q_0(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},z)Q_1(\boldsymbol{u},\boldsymbol{x},\boldsymbol{r},z). \end{cases} \tag{7}$$

Note that $f_{\boldsymbol{z}}(\boldsymbol{z})$ and the approximation $q_{\boldsymbol{z},0}(\boldsymbol{z})$ are both products of Bernoulli distributions. Thus, $q_{\boldsymbol{z},0}(\boldsymbol{z}) = f_{\boldsymbol{z}}(\boldsymbol{z})$ and this factor only needs to be updated when initializing the algorithm.

To capture as much a posterior correlation as possible between the elements of $(\boldsymbol{x},\boldsymbol{u},\boldsymbol{r},z)$ while minimizing convergence issues of the algorithm, it is important to tailor the approximating factors, i.e., the structure of the covariance matrices of the Gaussian approximations $q_{\boldsymbol{u},i}(\boldsymbol{u})$, $q_{\boldsymbol{x},i}(\boldsymbol{x})$, and $q_{\boldsymbol{r},i}(\boldsymbol{r})$ ($\forall i \in (0;1)$). For $q_{\boldsymbol{r},0}(\boldsymbol{r})$, the covariance matrix is kept isotropic since $f_{\boldsymbol{r},0}(\boldsymbol{r})$ is a product of independent and identical distributions. A full covariance matrix is used for $q_{\boldsymbol{x},1}(\boldsymbol{x})$ to preserve as much information as possible in the data augmentation term. The covariance matrices of the remaining approximation factors are constrained to be diagonal.

EP algorithm is an iterative algorithm and each iteration consists of solving sequentially the five KL minimization problems in (6), subject to the constraints on the covariance matrices mentioned above. These constraints are easy to ensure at each EP update, which consist here of matching the first and second moments of the tilted distributions (on the left-hand side of KL divergences in (6)) and of the global approximation.

The update of $Q_1(\boldsymbol{u},\boldsymbol{r},\boldsymbol{z})$ relies on the marginals of the tilted distribution $f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{r},\boldsymbol{z})Q_0(\boldsymbol{u},\boldsymbol{r},\boldsymbol{z})$, since $Q_1(\boldsymbol{u},\boldsymbol{r},\boldsymbol{z})$ and $Q_0(\boldsymbol{u},\boldsymbol{r},\boldsymbol{z})$ are separable (with respect to $\boldsymbol{u},\boldsymbol{r}$ and $\boldsymbol{z}$). The marginal tilted distribution of $\boldsymbol{u}$ and $\boldsymbol{r}$ are mixtures of two distributions, i.e., one Gaussian distribution and one non-standard distribution whose density is proportional to the

product of the Poisson likelihood $f(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z}|\boldsymbol{y})$ by a Gaussian distribution. Although the latter distribution is not standard, its mean and variance can be computed numerically as in [8].

The update of $q_{\boldsymbol{u},0}(\boldsymbol{u})q_{\boldsymbol{x},1}(\boldsymbol{x})$ relies on computing the moments of tilted distribution $\delta(\boldsymbol{u} - \mathbf{A}\boldsymbol{x})q_{\boldsymbol{u},1}(\boldsymbol{u})q_{\boldsymbol{x},0}(\boldsymbol{x})$. Since both $q_{\boldsymbol{u},0}(\boldsymbol{u})$ and $q_{\boldsymbol{u},1}(\boldsymbol{u})$ have diagonal covariance matrices, it is sufficient to compute the marginal means and variances (with respect to $\boldsymbol{u}$) of the tilted distribution, which can be done in closed form. The factor $q_{\boldsymbol{x},1}(\boldsymbol{x})$ has a full covariance matrix and this factor is updated by computing the moments of $\int \delta(\boldsymbol{u} - \mathbf{A}\boldsymbol{x})q_{\boldsymbol{u},1}(\boldsymbol{u})q_{\boldsymbol{x},0}(\boldsymbol{x})\mathrm{d}\boldsymbol{u}$, which is a multivariate Gaussian distribution.

The factor $q_{\boldsymbol{x},0}(\boldsymbol{x})$ is updated by sequentially computing the marginal means and variances of $f_{\boldsymbol{x}}(\boldsymbol{x})q_{\boldsymbol{x},1}(\boldsymbol{x})$ for each element in $\boldsymbol{x}$. These sequential updates are necessary as the covariance matrix of $q_{\boldsymbol{x},1}(\boldsymbol{x})$ is full and $f_{\boldsymbol{x}}(\boldsymbol{x})$ imposes positivity constraints on $\boldsymbol{x}$. The update of $q_{\boldsymbol{r},0}(\boldsymbol{r})$ is performed by computing the mean and variance of tilted distribution $f_{\boldsymbol{r}}(\boldsymbol{r})q_{\boldsymbol{r},1}(\boldsymbol{r})$, which is a Truncated Gaussian distribution or a Gamma distribution depending on the prior $f_{\boldsymbol{r}}(\boldsymbol{r})$ and has a diagonal covariance. A gradient-based method [15] minimizing the fourth line of (6) is used to update the isotropic covariance of $q_{\boldsymbol{r},0}(\boldsymbol{r})$.

The algorithm is stopped when the variation of the parameters of the approximating factors falls below a threshold. While damping strategies [15] can be used to reduce convergence issues, we did not implement such methods here as our algorithm did not suffer from severe convergence issues in the scenarios considered. After convergence, we use as point estimates the marginal means of $(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z})$, which are the means of the approximate marginals $Q_{\boldsymbol{x}}(\boldsymbol{x}), Q_{\boldsymbol{r}}(\boldsymbol{r})$, and $Q_{\boldsymbol{z}}(\boldsymbol{z})$, respectively. The marginal posterior covariance of $\boldsymbol{x}$ is approximated by the covariance of $Q_{\boldsymbol{x}}(\boldsymbol{x})$.

V. Experiments

A series of experiments were conducted using synthetic and real data to illustrate the performance of the proposed EP method. While the synthetic data was generated in a similar fashion as the simulations in [13], a real multispectral photon-starved dataset [3] was also used to demonstrate the effectiveness of the proposed method.

A. Synthetic experiments

Two synthetic datasets were generated according to the models used in [13] and the model considered here, respectively, with $(M, N) = (500, 20)$ and the elements of $\mathbf{A}$ were generated independently from a uniform distribution over $(0, 1)$. The prior variance of $\boldsymbol{x}$ was set to $\lambda_{\boldsymbol{x}}^{-2} = 1$ and the prior variance of $\boldsymbol{r}$ was set to $\lambda_{\boldsymbol{r}}^{-2} = 150$ ($\boldsymbol{r}$ in Fig. 1 and $\boldsymbol{v}$ in Fig. 2). The prior probability $\frac{1}{1+\exp(-p_0)}$ of anomaly presence in each observation was set to 10%. The first synthetic dataset was generated with additive anomalies. The ground truth of signals $\mathbf{A}\boldsymbol{x}$ and anomalies $\boldsymbol{z} \odot \boldsymbol{r}$ are shown in Fig. 3 (1a). The second synthetic dataset was generated with destructive and additive anomalies and the original matrix $\mathbf{A}$ was multiplied by 30 to increase the mean amplitude of the signal of interest and allow

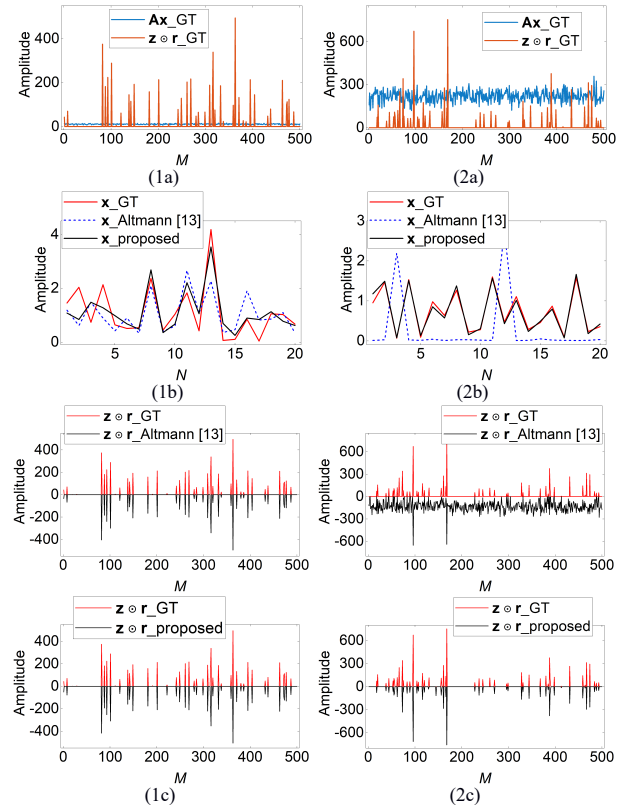
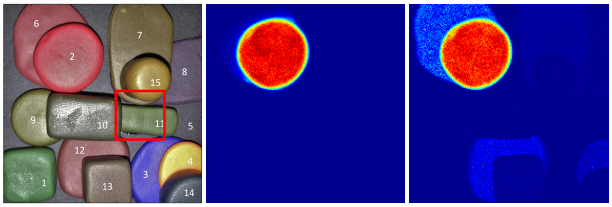


Fig. 3. Top row: ground truth signals (blue) and anomalies (orange). Second row: ground truth of $\boldsymbol{x}$ (red) and estimated posterior mean of $\boldsymbol{x}$ by the method in [13] (dashed blue) and the proposed EP method (black). Bottom row: ground truth of $\boldsymbol{z} \odot \boldsymbol{r}$ and estimated posterior mean of $\boldsymbol{z} \odot \boldsymbol{r}$ by the method in [13] and the proposed EP method.

destructive anomalies using $\lambda_{\boldsymbol{r}}^{-2} = 150$. The ground truth of signals $\mathbf{A}\boldsymbol{x}$ and anomalies $\boldsymbol{z} \odot \boldsymbol{r}$ are shown in Fig. 3 (2a). It can be seen in Fig. 3 (1b) and (1c) that both methods provide very similar results when the anomalies are additive, illustrating the ability of the new proposed EP method to capture the additive anomalies. On the other hand, the results in Fig 3. (2b) and (2c) show that when anomalies are destructive, the three posterior means are successfully estimated using the new proposed EP method, while the method in [13] fails to estimate the regression coefficients and the anomalies accurately.

B. Real data experiment

Real data experiments were conducted using the multispectral data used in [3] and the RGB image of the scene of interest is depicted in Fig. 4 (a). The scene consists of $N = 15$ main materials (endmembers in the spectral unmixing context), whose known spectral signatures are stored in $\mathbf{A}$ and whose abundances are the linear regression parameters in $\boldsymbol{x}$. Anomalies here are primarily due to additional materials (glue) and limitations of the imaging system which introduce non-uniform illumination, spatially and spectrally. The data consists of 190×190 pixels with $M = 33$ bands. The proposed EP method was applied pixel-wise and we used the



(a) (b) (c)

Fig. 4. (a): RGB image of the scene of interest. The expected anomalies are mainly located in the red box. (b)-(c): Spectral unmixing results for materials #2 by method from [3] (b) and by the proposed method (c).

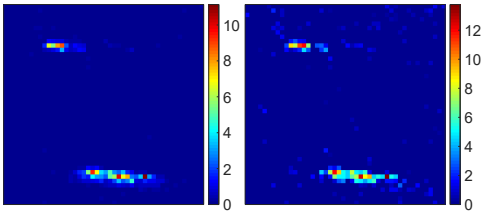


Fig. 5. ℓ_1-norm of anomalies detected by method in [3] (left) and proposed method (right).

results obtained by the MCMC method proposed in [3] as reference. Fig. 4 (b) and (c) depict the unmixing results for the material #2 labelled in Fig. 4 (a). We deliberately chose the abundance maps of this material as it is one of the few that present some differences. While the abundances estimated via the proposed EP method are generally in good agreement with the reference, they are not perfectly matched mainly due to the fact that, in contrast to [3], our prior model for $\boldsymbol{x}$ does not enforce abundance sparsity and spatial correlation. This limitation highlights potential improvements of the proposed method using more informative models for $\boldsymbol{x}$. Nonetheless, one advantage of EP is that it is very easy to parallelise, here the proposed EP method only takes 0.4s per pixel using Matlab R2018b implementation running on a HP Z2 Tower G4 Workstation of 16 RAM and a 3.7GHz Intel Core i7 processor. For comparison, the MCMC method from [3] takes more than 6 hours for the whole image, although most of the updates are performed in parallel. Finally, Fig. 5 depicts the map of the (pixel-wise) ℓ_1-norm of the detected anomalies in red box highlighted in Fig. 4 (a). The results obtained are in good agreement with the reference and confirm the presence of glue used to stick the objects to the dark cardboard.

VI. Conclusion

We proposed a new EP method for joint robust linear regression and anomaly detection in the presence of Poisson noise, assuming that anomalies can be additive or destructive. By using an extended Bayesian model, the proposed EP method yields a computationally efficient solution to approximate the complex joint posterior distribution. The experiment conducted with synthetic data illustrated the potential benefits of the model over the previous anomaly model and the simulations conducted with real data pointed at several routes for improvements. While accounting for the structured sparsity of anomalies (e.g., as in [13]) is possible, significant gains can be expected in spectral unmixing applications where more complex models can be used for $\boldsymbol{x}$.

References

[1] P. J. Rousseeuw and M. Hubert, "Anomaly detection by robust statistics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 2, p. e1236, 2018.

[2] G. E. Newstadt, A. O. Hero, and J. Simmons, "Robust spectral unmixing for anomaly detection," in *IEEE Workshop on Statistical Signal Processing (SSP)*, 2014, pp. 109–112.

[3] Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. S. Buller, S. McLaughlin, and A. Hero, "Robust spectral unmixing of sparse multispectral lidar waveforms using gamma markov random fields," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 658–670, 2017.

[4] C. Zhang, S. Arridge, and B. Jin, "Expectation propagation for poisson data," *Inverse Problems*, vol. 35, no. 8, p. 085006, 2019.

[5] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[6] T. Minka, "Expectation propagation for approximate Bayesian inference," in *Uncertainty in Artificial Intelligence*, vol. 17, 2001, pp. 362–369.

[7] A. S. Kim, M. Wand *et al.*, "The explicit form of expectation propagation for a simple statistical model," *Electronic Journal of Statistics*, vol. 10, no. 1, pp. 550–581, 2016.

[8] Y.-J. Ko and M. W. Seeger, "Expectation propagation for rectified linear poisson regression," in *Asian Conference on Machine Learning*, 2016, pp. 253–268.

[9] Y. Altmann, A. Perelli, and M. E. Davies, "Expectation-propagation algorithms for linear regression with poisson noise: application to photon-limited spectral unmixing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 5067–5071.

[10] A. Gelman, A. Vehtari, P. Jylänki, T. Sivula, D. Tran, S. Sahai, P. Blomstedt, J. P. Cunningham, D. Schiminovich, and C. Robert, "Expectation propagation as a way of life: A framework for bayesian inference on partitioned data," *arXiv preprint arXiv:1412.4869*, 2017.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[12] P. Schniter, S. Rangan, and A. K. Fletcher, "Vector approximate message passing for the generalized linear model," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 1525–1529.

[13] Y. Altmann, D. Yao, S. McLaughlin, and M. E. Davies, "Robust linear regression and anomaly detection in the presence of poisson noise using expectation-propagation," in *Proceedings of the Second World Congress on Contition Monitoring (WCCM)*, 2019.

[14] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4810–4819, 2015.

[15] J. M. Hernández-Lobato, D. Hernández-Lobato, and A. Suárez, "Expectation propagation in linear regression models with spike-and-slab priors," *Machine Learning*, vol. 99, no. 3, pp. 437–487, 2015.