



# Sequential Image Recovery Using Joint Hierarchical Bayesian Learning

Yao Xiao<sup>1</sup> · Jan Glaubitz<sup>2</sup> 

Received: 3 August 2022 / Revised: 2 May 2023 / Accepted: 3 May 2023 /  
Published online: 18 May 2023  
© The Author(s) 2023

## Abstract

Recovering temporal image sequences (videos) based on indirect, noisy, or incomplete data is an essential yet challenging task. We specifically consider the case where each data set is missing vital information, which prevents the accurate recovery of the individual images. Although some recent (variational) methods have demonstrated high-resolution image recovery based on jointly recovering sequential images, there remain robustness issues due to parameter tuning and restrictions on the type of sequential images. Here, we present a method based on hierarchical Bayesian learning for the joint recovery of sequential images that incorporates prior intra- and inter-image information. Our method restores the missing information in each image by “borrowing” it from the other images. More precisely, we couple sequential images by penalizing their pixel-wise difference. The corresponding penalty terms (one for each pixel and pair of subsequent images) are treated as weakly-informative random variables that favor small pixel-wise differences but allow occasional outliers. As a result, *all* of the individual reconstructions yield improved accuracy. Our method can be used for various data acquisitions and allows for uncertainty quantification. Some preliminary results indicate its potential use for sequential deblurring and magnetic resonance imaging.

**Keywords** Sequential image recovery · Hierarchical Bayesian learning · Uncertainty quantification · Fourier data · Image deblurring

**Mathematics Subject Classification** 15A29 · 62F15 · 65F22 · 65K10 · 68U10

---

JG acknowledges support from AFOSR #F9550-18-1-0316, ONR #N00014-20-1-2595, and the US Department of Energy, SciDAC program, under grant DE-SC0012704.

---

✉ Jan Glaubitz  
glaubitz@mit.edu

Yao Xiao  
yao.xiao.gr.dartmouth@gmail.com

<sup>1</sup> Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA

<sup>2</sup> Department of Aeronautics and Astronautics, MIT, Cambridge, MA 02139, USA

### 1 Introduction

Many applications rely on recovering temporal image sequences from noisy, indirect, and incomplete data [6, 28, 42, 43]. We can often formulate this task as a sequence of *linear inverse problems*,

$$\mathbf{y}^{(j)} = F^{(j)}\mathbf{x}^{(j)} + \mathbf{e}^{(j)}, \quad j = 1, \dots, J, \tag{1}$$

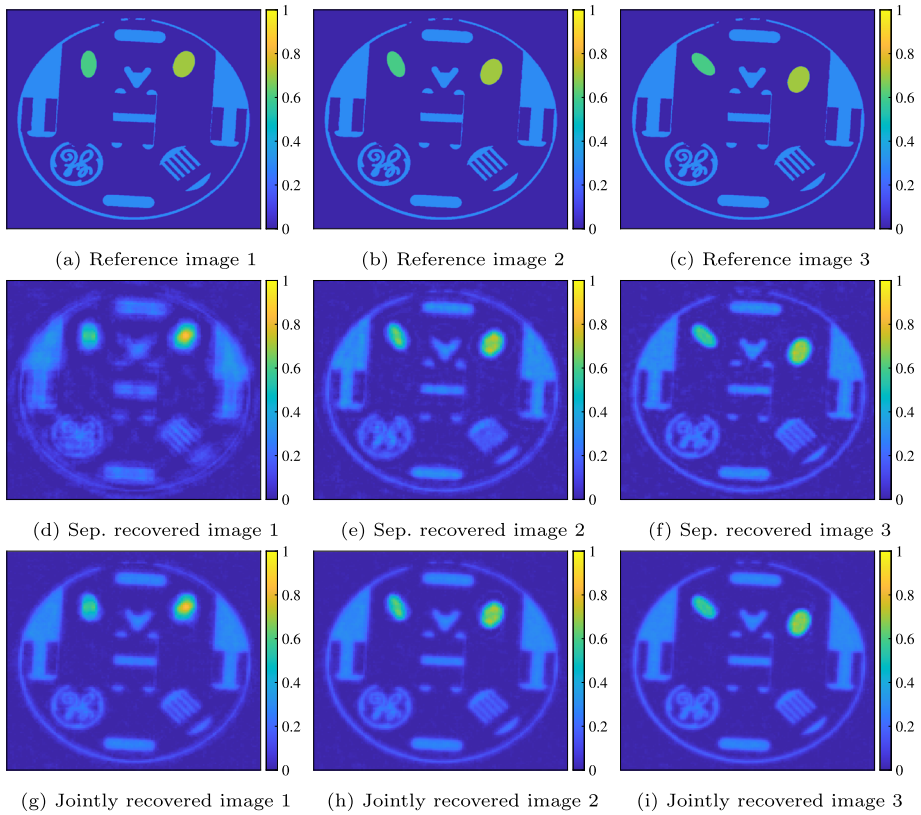
where  $\mathbf{y}^{(j)}$  is a given data vector,  $\mathbf{x}^{(j)}$  is the (unknown) vectorized image,  $F^{(j)}$  is a known linear forward operator, and  $\mathbf{e}^{(j)}$  corresponds to unknown noise. The individual recovery of each image by separately solving the linear inverse problems (1) is a well-studied, although challenging problem by itself [25, 27, 40]. A prominent approach is to replace (1) with a nearby regularized inverse problem that promotes some prior belief about the unknown image. In imaging applications, it is often reasonable to assume that some linear transform of the unknown image  $\mathbf{x}$ , say  $R\mathbf{x}$ , is sparse. This prior belief yields the  *$\ell^1$ -regularized inverse problems*

$$\min_{\mathbf{x}^{(j)}} \left\{ \|F^{(j)}\mathbf{x}^{(j)} - \mathbf{y}^{(j)}\|_2^2 + \lambda_j \|R\mathbf{x}^{(j)}\|_1 \right\}, \quad j = 1, \dots, J, \tag{2}$$

where  $R$  is the regularization operator and  $\lambda_j > 0$  are regularization parameters. Usual choices for  $R$  are discrete (high-order) total variation (TV) [37], total generalized/directional variation [8]/ [35, 36], and polynomial annihilation [2, 3, 23] operators. The rationale behind considering (2) is that the  $\ell^1$ -norm,  $\|\cdot\|_1$ , serves as a convex surrogate for the  $\ell^0$ -“norm”,  $\|\cdot\|_0$ ; an observation that lies in the heart of compressed sensing [16, 18, 21]. Another prominent approach, closely related to regularization, is Bayesian inverse problems [11, 29, 38]. In this setting, we express our lack of information about some of the quantities in (1) by modeling them as random variables, which (relations) are characterized by certain density functions. The fidelity and regularization term corresponds to the negative logarithm of the likelihood and prior density, respectively, and the regularized solution (2) corresponds to the maximizer of the posterior density. The class of conditionally Gaussian priors [9, 10, 12, 13, 39] is particularly suited to promote sparsity.

Here, we consider the case that each data set  $\mathbf{y}^{(j)}$  in (1) is missing vital information, which prevents the accurate individual recovery of the images  $\mathbf{x}^{(j)}$ . A common strategy is to restore the missing information in each image by “borrowing” it from the other images. [1, 15, 17, 41, 45] considered deterministic and Bayesian methods of such a flavor that rely on a *common sparsity assumption*, meaning that  $R\mathbf{x}^{(1)}, \dots, R\mathbf{x}^{(J)}$  have the same support. Unfortunately, temporally changing image sequences often violate the common sparsity assumption. Recently, [42] addressed this problem by locally coupling the images in no-change regions while decoupling them in change regions. This was done by first computing a diagonal change mask  $C^{(j-1,j)}$  directly from the consecutive data sets  $\mathbf{y}^{(j-1)}, \mathbf{y}^{(j)}$  and using this change mask to penalize any difference between  $\mathbf{x}^{(j-1)}$  and  $\mathbf{x}^{(j)}$  in no-change regions. Let  $[C^{(j-1,j)}]_{n,n} = 0$  in change regions and  $[C^{(j-1,j)}]_{n,n} = 1$  in no-change regions. The *joint  $\ell^1$ -regularized inverse problem* used in [42] is

$$\min_{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(J)}} \left\{ \sum_{j=1}^J \|F^{(j)}\mathbf{x}^{(j)} - \mathbf{y}^{(j)}\|_2^2 + \sum_{j=1}^J \lambda_j \|R\mathbf{x}^{(j)}\|_1 + \sum_{j=2}^J \mu_j \|C^{(j-1,j)}(\mathbf{x}^{(j-1)} - \mathbf{x}^{(j)})\|_2^2 \right\}, \tag{3}$$



**Fig. 1** Three images of a temporal magnetic resonance image sequence with a rotating left (green) and a down-moving right (yellow) ellipse. The first row shows the reference images, the second row the separately recovered images using (2), and the third row the jointly recovered images using (3). The data sets are noisy and miss different Fourier samples (Color figure online)

where the  $\mu_j \geq 0$  are fixed coupling parameters. Equation (3) balances data fidelity, intra-image regularization (sparsity of  $R\mathbf{x}^{(j)}$ ), and inter-image regularization (coupling in no-change regions). Roughly speaking, the last term in (3),

$$\left\| C^{(j-1,j)}(\mathbf{x}^{(j-1)} - \mathbf{x}^{(j)}) \right\|_2^2 = \sum_{n=1}^N \left[ C^{(j-1,j)} \right]_{n,n} \left| x_n^{(j-1)} - x_n^{(j)} \right|^2, \tag{4}$$

penalizes any change between the two subsequent images  $\mathbf{x}^{(j-1)}$  and  $\mathbf{x}^{(j)}$  in no-change regions, where  $[C^{(j-1,j)}]_{n,n} = 1$ , since the value of the objective function in (3) increases with  $|x_n^{(j-1)} - x_n^{(j)}|$  in this case. At the same time, in change regions, the difference  $|x_n^{(j-1)} - x_n^{(j)}|$  does not influence the value of the objective function since  $[C^{(j-1,j)}]_{n,n} = 0$  in this case. Figure 1 illustrates the advantage of jointly recovering a temporal sequence of magnetic resonance images from noisy and under-sampled Fourier data (see Sect. 4 for more details).

While the joint recovery of changing images using (3) can yield an improved accuracy, there remain issues with robustness and the range of application: (II) Selecting appropriate regularization and coupling parameters is a non-trivial task, and their choice can critically influence the quality of the recovered images. Although (re-)weighted [1, 14, 22]

$\ell^1$ -regularization can increase the robustness w.r.t. the intra-image regularization parameters  $\lambda_j$ , selecting suitable inter-image coupling parameters  $\mu_j$  remains an open problem. (I2) The method proposed in [42, Section 3] to pre-compute the change masks uses Fourier data and assumes that the sequential images only include objects with closed boundaries. This assumption prevents the application to problems with other types of data acquisition. Finally, one has to tune some problem-dependent free parameters by hand on a case-by-case basis.

## Our Contribution

We propose a *joint hierarchical Bayesian learning (JHBL)* method for sequential image recovery. The method is easy to implement, efficient, and simple to parallelize. Our method avoids issues (I1) and (I2) by reinterpreting all involved parameters—including the change mask—as random variables, which we then estimate together with the recovered images. In particular, for the random variables responsible for the inter-image coupling, we use weakly-informative gamma distributions that favor small pixel-wise differences (in no-change regions) but allow for occasional outliers (in change regions). Our approach does not rely on Fourier data or images showing objects with closed boundaries. We demonstrate that our JHBL method improves the accuracy of all individual images. Another advantage is that our method quantifies the uncertainty in the recovered images, which is often desirable. Some preliminary results for magnetic resonance imaging and road-traffic monitoring indicate the advantage of JHBL for sequential image recovery.

## Outline

In Sect. 2, we present the JHBL model that promotes intra-image sparsity and inter-image coupling. In Sect. 3, we propose an efficient method for Bayesian inference. In Sect. 4, we demonstrate the performance of the resulting JHBL method for test cases from sequential deblurring and magnetic resonance imaging. Section 5 offers some concluding thoughts.

## 2 The Joint Hierarchical Bayesian Model

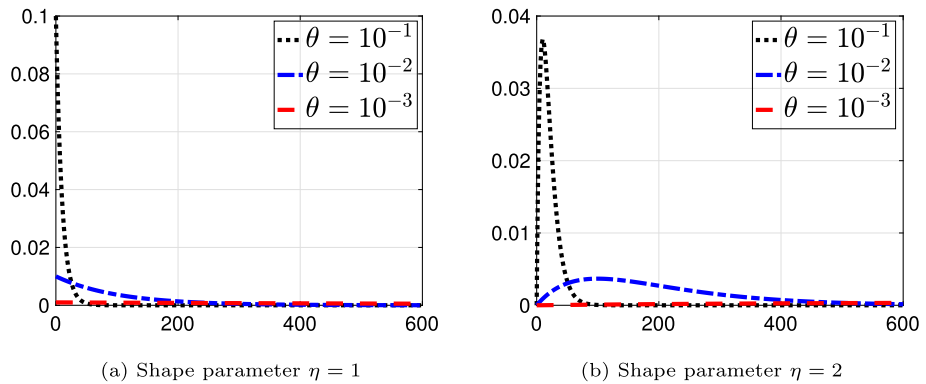
We now describe the hierarchical Bayesian model we use to develop our JHBL method.

### 2.1 The Likelihood

Consider the data model (1) and assume that  $\mathbf{y}^{(j)} \in \mathbb{R}^{M_j}$ ,  $F \in \mathbb{R}^{M_j \times N}$ ,  $\mathbf{x}^{(j)} \in \mathbb{R}^N$ , and that  $\mathbf{e}^{(j)} \in \mathbb{R}^{M_j}$  is independent and identically distributed (i.i.d.) zero-mean normal noise,  $e_m^{(j)} \sim \mathcal{N}(0, \alpha_j)$  for  $m = 1, \dots, M_j$ , with noise precision  $\alpha_j > 0$ .<sup>1</sup> The  $j$ th likelihood function, which is the conditional probability density of  $\mathbf{y}^{(j)}$  given  $\mathbf{x}^{(j)}$  and  $\alpha_j$ , is

$$p(\mathbf{y}^{(j)} | \mathbf{x}^{(j)}, \alpha_j) \propto \alpha_j^{M_j/2} \exp \left\{ -\frac{\alpha_j}{2} \left\| F^{(j)} \mathbf{x}^{(j)} - \mathbf{y}^{(j)} \right\|_2^2 \right\}, \quad (5)$$

<sup>1</sup> We can also use (1) for complex (Fourier) data.



**Fig. 2** Gamma density functions for different shape and rate parameters  $\eta, \theta$ . Recall that the mode, expected value, and variance of a gamma distribution  $\Gamma(\eta, \theta)$  are  $(\eta - 1)/\theta, \eta/\theta$ , and  $\eta/\theta^2$ , respectively. In particular, decreasing the rate  $\theta$  makes the density flatter, while  $\eta$  influences its peak (Color figure online)

where “ $\propto$ ” means that the two sides are equal up to a multiplicative constant. We also treat the noise precision  $\alpha_j$  as a random variable that is learned together with the images  $\mathbf{x}^{(j)}$ .<sup>2</sup> We assume that  $\alpha_j$  is gamma distributed,

$$p(\alpha_j) = \Gamma(\alpha_j | \eta_{\alpha_j}, \theta_{\alpha_j}) \propto \alpha_j^{\eta_{\alpha_j} - 1} \exp\{-\theta_{\alpha_j} \alpha_j\}, \quad j = 1, \dots, J. \tag{6}$$

For simplicity, we use the same shape and rate parameter for all noise precisions, denoted by  $\eta_\alpha$  and  $\theta_\alpha$ . We use the gamma distribution because it is conditionally conjugate to the normal distribution, which is convenient for Bayesian inference (see Sect. 3). Moreover, we can make the noise hyper-priors (6) flat and therefore uninformative by choosing  $\theta_\alpha \approx 0$ . Figure 2 illustrates the gamma density function for different shape and rate parameters. Remark 1 addresses the possibility of using informative hyper-priors.

If we denote the collection of all images by  $\mathbf{x} = [\mathbf{x}^{(1)}; \dots; \mathbf{x}^{(J)}]$ , of all data sets by  $\mathbf{y} = [\mathbf{y}^{(1)}; \dots; \mathbf{y}^{(J)}]$ , and of all noise precisions by  $\boldsymbol{\alpha} = [\alpha_1; \dots; \alpha_J]$ , then the *joint likelihood function* is

$$p(\mathbf{y} | \mathbf{x}, \boldsymbol{\alpha}) = \prod_{j=1}^J p(\mathbf{y}^{(j)} | \mathbf{x}^{(j)}, \alpha_j) \propto \prod_{j=1}^J \alpha_j^{M_j/2} \exp\left\{-\frac{\alpha_j}{2} \|F^{(j)} \mathbf{x}^{(j)} - \mathbf{y}^{(j)}\|_2^2\right\}, \tag{7}$$

assuming that the data sets are conditionally independent. A few remarks are in order.

**Remark 1** For simplicity we use the same hyper-prior  $\Gamma(\cdot | \eta_\alpha, \theta_\alpha)$  and parameters  $\eta_\alpha, \theta_\alpha$  for all components of  $\boldsymbol{\alpha}$ . We do this since we assume no prior knowledge about the noise variances of the different measurement vectors. However, if one has a reasonable a priori notion of the underlying noise variances, the choice for hyper-prior could be modified accordingly [5, 13].

**Remark 2** The data sets being conditionally independent means that if we know the images  $\mathbf{x}$  and the noise precisions  $\boldsymbol{\alpha}$ , then knowledge of one data set  $\mathbf{y}^{(j)}$  provides no additional information about the likelihood of another data set  $\mathbf{y}^{(i)}$  with  $i \neq j$ .

<sup>2</sup> Even when the exact noise precision is known, using it as a fixed value for  $\alpha_j$  can yield suboptimal reconstructions [44].

### 2.2 The Intra-image Prior

We assume that some linear transform of the images, say  $R\mathbf{x}^{(j)}$  with  $R \in \mathbb{R}^{K \times N}$ , is sparse. One can model sparsity by various priors, including TV [4, 31], mixture-of-Gaussian [19], Laplace [20], and hyper-Laplace [32, 34] priors. Here, we use conditionally Gaussian priors, which are particularly suited to promote sparsity and allow for efficient Bayesian inference [9, 13, 24]. The  $j$ th intra-image prior is

$$p(\mathbf{x}^{(j)}|\boldsymbol{\beta}^{(j)}) \propto \det \left( B^{(j)} \right)^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(j)})^T R^T B^{(j)} R \mathbf{x}^{(j)} \right\}, \tag{8}$$

where  $B^{(j)} = \text{diag}(\boldsymbol{\beta}^{(j)})$  and  $\boldsymbol{\beta}^{(j)} = [\beta_1^{(j)}, \dots, \beta_K^{(j)}]$  is treated as a random vector with i. i. d. gamma distributed components,

$$p(\beta_k^{(j)}) = \Gamma(\beta_k^{(j)}|\eta_{\beta_k^{(j)}}, \theta_{\beta_k^{(j)}}), \quad k = 1, \dots, K. \tag{9}$$

For simplicity, we use the same shape and rate parameter for all prior precisions, denoted by  $\eta_\beta$  and  $\theta_\beta$ . We choose  $\theta_\beta \approx 0$  to make the hyper-prior flat and thus uninformative. Again, if one has a reasonable a priori notion of the support of  $R\mathbf{x}^{(j)}$ , the choice for the hyper-prior (9) and the parameters  $\eta_{\beta_k^{(j)}}$ ,  $\theta_{\beta_k^{(j)}}$  could be modified correspondingly. If we denote the collection of all precision vectors by  $\boldsymbol{\beta} = [\boldsymbol{\beta}^{(1)}; \dots; \boldsymbol{\beta}^{(J)}]$ , then the *joint intra-image prior* is

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\beta}) &= \prod_{j=1}^J p(\mathbf{x}^{(j)}|\boldsymbol{\beta}^{(j)}) \\ &\propto \prod_{j=1}^J \det \left( B^{(j)} \right)^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(j)})^T R^T B^{(j)} R \mathbf{x}^{(j)} \right\}, \end{aligned} \tag{10}$$

assuming that the images are conditionally independent.

**Remark 3** The images being conditionally independent means that if we know the parameters  $\boldsymbol{\beta}$ , then knowledge of one image  $\mathbf{x}^{(j)}$  provides no additional information about the likelihood of another image  $\mathbf{x}^{(i)}$  with  $i \neq j$ . Although we assume that the images form a temporal sequence with only parts of the images changing, we treat this as qualitative information; We neither want to quantify the exact location nor the size of the (no-)change regions, making us assume that the images are conditionally independent.

### 2.3 The Inter-image Prior

Assume for the moment that we had a pre-computed diagonal change mask  $C^{(j-1,j)}$  with  $[C^{(j-1,j)}]_{n,n} = 0$  in change regions and  $[C^{(j-1,j)}]_{n,n} = 1$  in no-change regions<sup>3</sup> as in [42]. We could then translate the coupling term in (3) into the empirical conditionally Gaussian

<sup>3</sup> For instance, Fig. 9c and d in Sect. 4.1 illustrate the pre-computed binary change mask used to jointly recover the images in Fig. 1.

prior

$$\begin{aligned}
 p(\mathbf{x}|\mu_2, \dots, \mu_J) &\propto \prod_{j=2}^J \mu_j^{N/2} \exp \left\{ -\frac{\mu_j}{2} \left\| C^{(j-1,j)} \left( \mathbf{x}^{(j-1)} - \mathbf{x}^{(j)} \right) \right\|_2^2 \right\} \\
 &= \prod_{j=2}^J \mu_j^{N/2} \exp \left\{ -\frac{\mu_j}{2} \left( \mathbf{x}^{(j-1)} - \mathbf{x}^{(j)} \right)^T C^{(j-1,j)} \left( \mathbf{x}^{(j-1)} - \mathbf{x}^{(j)} \right) \right\},
 \end{aligned}
 \tag{11}$$

where we have used that  $(C^{(j-1,j)})^2 = C^{(j-1,j)}$ . As mentioned before, the method proposed in [42, Section 3] to pre-compute the change masks uses Fourier data and assumes that the sequential images only include objects with closed boundaries, however. We overcome these restrictions by replacing the pre-computed diagonal elements of the change mask with random variables, which we then estimate together with the other parameters and images. We propose to use the *inter-image prior*

$$p(\mathbf{x}|\boldsymbol{\gamma}) \propto \prod_{j=2}^J \det \left( C^{(j-1,j)} \right)^{1/2} \exp \left\{ -\frac{1}{2} \left( \mathbf{x}^{(j-1)} - \mathbf{x}^{(j)} \right)^T C^{(j-1,j)} \left( \mathbf{x}^{(j-1)} - \mathbf{x}^{(j)} \right) \right\}
 \tag{12}$$

with  $C^{(j-1,j)} = \text{diag}(\boldsymbol{\gamma}^{(j-1,j)})$  and  $\boldsymbol{\gamma}^{(j-1,j)} = [\gamma_1^{(j-1,j)}, \dots, \gamma_N^{(j-1,j)}]^T$ . Furthermore,  $\boldsymbol{\gamma}$  in (12) denotes the collection of all coupling vectors  $\boldsymbol{\gamma}^{(1,2)}, \dots, \boldsymbol{\gamma}^{(J-1,J)}$ . The random variables  $\gamma_n^{(j-1,j)}$  introduce an adaptive and spatially varying weighting in the change mask. Another advantage of (12) is that we stay within the class of conditionally Gaussian densities, which is convenient for Bayesian inference. This further motivates us to assume that the elements of  $\boldsymbol{\gamma}^{(j-1,j)}$  are i. i. d. gamma distributed,

$$p\left(\gamma_n^{(j-1,j)}\right) = \Gamma\left(\gamma_n^{(j-1,j)} \mid \eta_{\gamma_n^{(j-1,j)}}, \theta_{\gamma_n^{(j-1,j)}}\right), \quad n = 1, \dots, N.
 \tag{13}$$

For simplicity, we use the same shape and rate parameter for all elements, denoted by  $\eta_\gamma$  and  $\theta_\gamma$ . To make the hyper-prior flat and therefore uninformative again, we choose  $\theta_\gamma \approx 0$ . As will be discussed in greater detail in Sect. 4, the choice of  $\eta_\gamma$  is influenced by the magnitude of change we expect to occur between consecutive pairs of images. Moreover, if one has a reasonable a priori notion of the location or amount of change between subsequent images, the choice for the hyper-prior (13) and the parameters  $\eta_{\gamma_n^{(j-1,j)}}, \theta_{\gamma_n^{(j-1,j)}}$  could be modified correspondingly. We will investigate such informative hyper-priors in future works.

### 2.4 The Combined Prior

We now discuss how promoting intra-image sparsity as in Sect. 2.2 and inter-image coupling as in Sect. 2.3 can be combined in a single prior. To this end, we can re-write the joint intra-image prior (10) more compactly as

$$p(\mathbf{x}|\boldsymbol{\beta}) \propto \det(B)^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \tilde{R}^T B \tilde{R} \mathbf{x} \right\},
 \tag{14}$$

where  $B = \text{diag}(B^{(1)}, \dots, B^{(J)})$  and  $\tilde{R} = \text{diag}(R, \dots, R)$  are diagonal matrices, and  $\mathbf{x} = [\mathbf{x}^{(1)}; \dots; \mathbf{x}^{(J)}]$  again denotes the collection of all images. We can now note that the joint intra-image prior (14) encodes the assumption that

$$\tilde{R} \mathbf{x} \sim \mathcal{N}(\mathbf{0}, B^{-1}),
 \tag{15}$$

i.e.,  $\tilde{R}\mathbf{x}$  is zero-mean normal distributed with diagonal precision matrix  $B$ . At the same time, we can re-write the inter-image prior (12) more compactly as

$$p(\mathbf{x}|\boldsymbol{\gamma}) \propto \det(C)^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T S^T C S \mathbf{x} \right\}, \tag{16}$$

where  $C = \text{diag}(C^{(1,2)}, \dots, C^{(J-1,J)})$  is a diagonal matrix and

$$S = \begin{bmatrix} I & -I & & & \\ & \ddots & \ddots & & \\ & & & \ddots & \\ & & & & I & -I \end{bmatrix} \tag{17}$$

with  $I$  denoting the  $N \times N$  identity matrix. Basically,  $S$  transforms the collection of images  $\mathbf{x} = [\mathbf{x}^{(1)}; \dots; \mathbf{x}^{(J)}]$  into the collection of differences of sequential images  $[\mathbf{x}^{(1)} - \mathbf{x}^{(2)}; \dots; \mathbf{x}^{(J-1)} - \mathbf{x}^{(J)}]$ . Similar to before, we can therefore note that the inter-image prior (16) encodes the assumption that

$$S\mathbf{x} \sim \mathcal{N}(\mathbf{0}, C^{-1}), \tag{18}$$

Having (15) and (18) at hand is convenient since it positions us to combine these two assumptions in a single prior. To this end, note that (15) and (18) holding simultaneously is equivalent to

$$\begin{bmatrix} \tilde{R} \\ S \end{bmatrix} \mathbf{x} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}^{-1} \right), \tag{19}$$

which corresponds to the *combined prior*

$$p(\mathbf{x}|\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \det(\tilde{R}^T B \tilde{R} + S^T C S)^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T (\tilde{R}^T B \tilde{R} + S^T C S) \mathbf{x} \right\}. \tag{20}$$

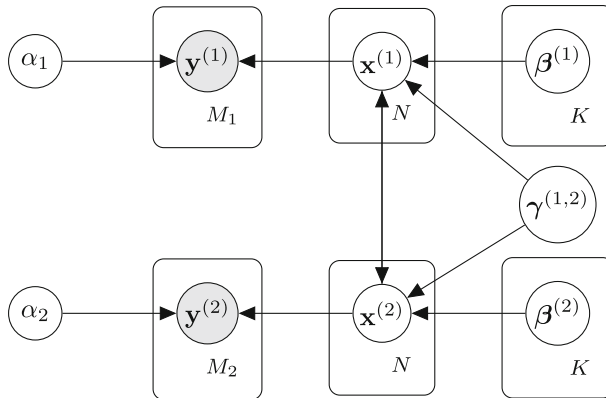
The combined prior (20) is a desirable model, but it can have issues when the covariance matrix  $\tilde{R}^T B \tilde{R} + S^T C S$  is non-invertible<sup>4</sup>, rendering the normalizing constant  $\det(\tilde{R}^T B \tilde{R} + S^T C S)^{1/2}$  zero. In this case, one may consider making the covariance matrix invertible by adding a small multiple of the identity to it. This approximation would allow some variability in the directions that collapse without changing in any substantial manner the variance along to dominant directions. Still, even when the covariance matrix is invertible, Bayesian inference can become computationally intractable due to the complicated relationship between the eigenvalues of  $\tilde{R}^T B \tilde{R} + S^T C S$  and the hyper-parameters  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$ . To overcome these challenges, we propose the following modification to the combined prior (20):

$$p(\mathbf{x}|\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \det(B)^{1/2} \det(C)^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T (\tilde{R}^T B \tilde{R} + S^T C S) \mathbf{x} \right\}. \tag{21}$$

The modified combined prior (21) offers computational benefits and produces satisfactory numerical results. We chose this particular form as the resulting fully conditional distributions for the hyper-parameters  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  are gamma distributions, making Bayesian maximum a posteriori estimation computationally efficient. Further details on the proposed model’s inference are provided in Sect. 3. We acknowledge that this modification is ad-hoc, and future

<sup>4</sup> For example, consider the discrete gradient operator  $R$  with  $[R\mathbf{x}^{(j)}]_k = x^{(j)}_k + 1 - x^{(j)}_k$ . If  $\mathbf{x}^{(1)} = \dots = \mathbf{x}^{(J)}$  is a constant sequence of constant images, where  $\mathbf{x}^{(j)} = C \cdot [1, \dots, 1]^T$  and  $C \in \mathbb{R}$ , then  $\tilde{R}^T B \tilde{R} + S^T C S$  maps it to zero. This illustrates that the kernel of  $\tilde{R}^T B \tilde{R} + S^T C S$  is not trivial and, as a result,  $\tilde{R}^T B \tilde{R} + S^T C S$  is non-invertible.





**Fig. 3** Graphical representation of the JHBL model for two images. Shaded and plain circles represent observed and unobserved random variables, respectively. The arrows indicate how the random variables influence each other. More specifically, the noise precisions  $\alpha_1, \alpha_2$  and images  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$  are connected to the data vectors  $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}$  via the likelihood (5); The hyper-parameters  $\beta^{(1)}, \beta^{(2)}$  are connected to the images  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$  via the intra-image prior (10); The images  $\mathbf{x}^{(1)}$  and  $\mathbf{x}^{(2)}$  are connected to each and the hyper-parameter  $\gamma^{(1,2)}$  via the inter-image prior (12), where  $\gamma^{(1,2)}$  encodes the (change and no-change) regions and the amount of coupling

research could explore other potential modifications. Overall, the modified combined prior strikes a balance between modeling considerations and computational efficiency. Figure 3 provides a graphical summary of the JHBL model described in this section.

**Remark 4** The joint intra-image and inter-image priors, (14) and (16), are not necessarily consistent. That is, the marginal distributions of  $\mathbf{x}$  derived from the conditional distributions of  $\mathbf{x}|\beta$  and  $\mathbf{x}|\gamma$ , respectively, can differ. For this reason, we advise against performing Bayesian inference using the marginal prior. Instead, our Bayesian inference algorithm outlined in Sect. 3 will utilize the hierarchical structure of the Bayesian model.

### 3 Bayesian Inference

Assume that we are interested in the most probable value (mode) of the posterior  $p(\mathbf{x}, \alpha, \beta, \gamma|\mathbf{y})$ , i. e., a combination of images and parameters for which the given data sets are most likely. To this end, we adopt the Bayesian coordinate descent (BCD) algorithm [24], which efficiently approximates the posterior mode by alternatingly updating the images and parameters based on the mode of the fully conditional distributions.

#### 3.1 The Fully Conditional Distributions

We chose conditionally Gaussian priors and gamma hyper-priors because they are conditionally conjugate. This allows us to derive analytical expressions for the fully conditional densities, which is convenient for Bayesian inference. Bayes’ theorem states that

$$\begin{aligned}
 p(\mathbf{x}, \alpha, \beta, \gamma|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{x}, \alpha, \beta, \gamma)p(\mathbf{x}, \alpha, \beta, \gamma) \\
 &= p(\mathbf{y}|\mathbf{x}, \alpha)p(\mathbf{x}|\beta, \gamma)p(\alpha)p(\beta)p(\gamma),
 \end{aligned}
 \tag{22}$$

where  $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\alpha})$  is the likelihood,  $p(\mathbf{x}|\boldsymbol{\beta}, \boldsymbol{\gamma})$  is the prior, and  $p(\boldsymbol{\alpha}), p(\boldsymbol{\beta}), p(\boldsymbol{\gamma})$  are the hyper-priors. If we substitute the joint likelihood (7) and combined prior (21) together with the hyper-priors (6), (9), (13) into (22), we get

$$\begin{aligned}
 & p(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}|\mathbf{y}) \\
 & \propto \left( \prod_{j=1}^J \alpha_j^{M_j/2} \exp \left\{ -\frac{\alpha_j}{2} \left\| F^{(j)} \mathbf{x}^{(j)} - \mathbf{y}^{(j)} \right\|_2^2 \right\} \right) \\
 & \left( \prod_{j=1}^J \det(B^{(j)})^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(j)})^T R^T B^{(j)} R \mathbf{x}^{(j)} \right\} \right) \\
 & \left( \prod_{j=2}^J \det(C^{(j-1,j)})^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(j-1)} - \mathbf{x}^{(j)})^T C^{(j-1,j)} (\mathbf{x}^{(j-1)} - \mathbf{x}^{(j)}) \right\} \right) \\
 & \left( \prod_{j=1}^J \alpha_j^{\eta_\alpha - 1} \exp\{-\theta_\alpha \alpha_j\} \right) \left( \prod_{j=1}^J \beta_j^{\eta_\beta - 1} \exp\{-\theta_\beta \beta_j\} \right) \left( \prod_{j=2}^J \gamma_j^{\eta_\gamma - 1} \exp\{-\theta_\gamma \gamma_j\} \right). \tag{23}
 \end{aligned}$$

We can note from (23) that

$$\begin{aligned}
 p(\alpha_j | \mathbf{x}^{(j)}, \mathbf{y}^{(j)}) & \propto \alpha_j^{M_j/2 + \eta_\alpha - 1} \exp \left\{ -\left( \left\| F^{(j)} \mathbf{x}^{(j)} - \mathbf{y}^{(j)} \right\|_2^2 / 2 + \theta_\alpha \right) \alpha_j \right\}, \\
 p(\beta_k^{(j)} | \mathbf{x}^{(j)}) & \propto (\beta_k^{(j)})^{1/2 + \eta_\beta - 1} \exp \left\{ -\left( [R \mathbf{x}^{(j)}]_k^2 / 2 + \theta_\beta \right) \beta_k^{(j)} \right\}, \\
 p(\gamma_n^{(j-1,j)} | \mathbf{x}^{(j-1)}, \mathbf{x}^{(j)}) & \propto (\gamma_n^{(j-1,j)})^{1/2 + \eta_\gamma - 1} \exp \left\{ -\left( [\mathbf{x}^{(j-1)} - \mathbf{x}^{(j)}]_n^2 / 2 + \theta_\gamma \right) \gamma_n^{(j-1,j)} \right\},
 \end{aligned} \tag{24}$$

and thus

$$p(\alpha_j | \mathbf{x}^{(j)}, \mathbf{y}^{(j)}) \propto \Gamma(\alpha_j | \eta_\alpha + M_j/2, \theta_\alpha + \left\| F^{(j)} \mathbf{x}^{(j)} - \mathbf{y}^{(j)} \right\|_2^2 / 2), \tag{25}$$

$$p(\beta_k^{(j)} | \mathbf{x}^{(j)}) \propto \Gamma(\beta_k^{(j)} | \eta_\beta + 1/2, \theta_\beta + [R \mathbf{x}^{(j)}]_k^2 / 2), \tag{26}$$

$$p(\gamma_n^{(j-1,j)} | \mathbf{x}^{(j-1)}, \mathbf{x}^{(j)}) \propto \Gamma(\gamma_n^{(j-1,j)} | \eta_\gamma + 1/2, \theta_\gamma + [\mathbf{x}^{(j-1)} - \mathbf{x}^{(j)}]_n^2 / 2), \tag{27}$$

where (25), (26) hold for  $j = 1, \dots, J$  and  $k = 1, \dots, K$ , while (27) holds for  $j = 2, \dots, J$  and  $n = 1, \dots, N$ . Similarly, for  $j = 2, \dots, J - 1$ , we have

$$\begin{aligned}
 & p(\mathbf{x}^{(1)} | \mathbf{x}^{(2)}, \mathbf{y}^{(1)}, \alpha_1, \boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1,2)}) \propto \mathcal{N}(\mathbf{x}^{(1)} | \boldsymbol{\mu}^{(1)}, \Sigma^{(1)}), \\
 & p(\mathbf{x}^{(j)} | \mathbf{x}^{(j-1)}, \mathbf{x}^{(j+1)}, \mathbf{y}^{(j)}, \alpha_j, \boldsymbol{\beta}^{(j)}, \boldsymbol{\gamma}^{(j-1,j)}, \boldsymbol{\gamma}^{(j,j+1)}) \propto \mathcal{N}(\mathbf{x}^{(j)} | \boldsymbol{\mu}^{(j)}, \Sigma^{(j)}), \\
 & p(\mathbf{x}^{(J)} | \mathbf{x}^{(J-1)}, \mathbf{y}^{(J)}, \alpha_J, \boldsymbol{\beta}^{(J)}, \boldsymbol{\gamma}^{(J-1,J)}) \propto \mathcal{N}(\mathbf{x}^{(J)} | \boldsymbol{\mu}^{(J)}, \Sigma^{(J)}),
 \end{aligned} \tag{28}$$

with covariance matrices

$$\begin{aligned} \Sigma^{(1)} &= \left( \alpha_1 \left( F^{(1)} \right)^T F^{(1)} + R^T B^{(1)} R + C^{(1,2)} \right)^{-1}, \\ \Sigma^{(j)} &= \left( \alpha_j \left( F^{(j)} \right)^T F^{(j)} + R^T B^{(j)} R + C^{(j-1,j)} + C^{(j,j+1)} \right)^{-1}, \\ \Sigma^{(J)} &= \left( \alpha_J \left( F^{(J)} \right)^T F^{(J)} + R^T B^{(J)} R + C^{(J-1,J)} \right)^{-1}, \end{aligned} \tag{29}$$

and means

$$\begin{aligned} \mu^{(1)} &= \Sigma^{(1)} \left( \alpha_1 \left( F^{(1)} \right)^T \mathbf{y}^{(1)} + C^{(1,2)} \mathbf{x}^{(2)} \right), \\ \mu^{(j)} &= \Sigma^{(j)} \left( \alpha_j \left( F^{(j)} \right)^T \mathbf{y}^{(j)} + C^{(j-1,j)} \mathbf{x}^{(j-1)} + C^{(j,j+1)} \mathbf{x}^{(j+1)} \right), \\ \mu^{(J)} &= \Sigma^{(J)} \left( \alpha_J \left( F^{(J)} \right)^T \mathbf{y}^{(J)} + C^{(J-1,J)} \mathbf{x}^{(J-1)} \right). \end{aligned} \tag{30}$$

Since  $\alpha_j$ ,  $\beta_k^{(j)}$ , and  $\gamma_n^{(j-1,j)}$  are gamma distributed and only take on positive values, the covariance matrices in (29) are symmetric and positive definite (SPD).

### 3.2 Proposed Method: Joint Hierarchical Bayesian Learning

We are now positioned to formulate our JHBL method by adapting the BCD algorithm [24] to our model in Sect. 2. The BCD algorithm is described in Algorithm 1 and approximates the mode (or mean) of the posterior  $p(\mathbf{x}, \alpha, \beta, \gamma | \mathbf{y})$  by alternately updating the images and parameters based on the mode (or mean) of the fully conditional distributions (25), (26), (27), (28).

---

#### Algorithm 1 BCD algorithm

---

- 1: Initialize  $\mathbf{x}^0$  and set  $l = 0$
  - 2: **repeat**
  - 3:   Update  $\alpha$  by setting  $\alpha^{l+1} = \arg \max_{\alpha} p(\alpha | \mathbf{x}^l)$
  - 4:   Update  $\beta$  by setting  $\beta^{l+1} = \arg \max_{\beta} p(\beta | \mathbf{x}^l)$
  - 5:   Update  $\gamma$  by setting  $\gamma^{l+1} = \arg \max_{\gamma} p(\gamma | \mathbf{x}^l)$
  - 6:   Update  $\mathbf{x}$  by setting  $\mathbf{x}^{l+1} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{x}^l, \alpha^{l+1}, \beta^{l+1}, \gamma^{l+1}, \mathbf{y})$
  - 7:   Increase  $l \rightarrow l + 1$
  - 8: **until** convergence or maximum number of iterations is reached
- 

Algorithm 1 is efficient and straightforward to implement because of the analytical expressions for the fully conditional distributions we derived in (25), (26), (27), (28). The mode of a gamma distribution  $\Gamma(\eta, \theta)$  with  $\eta, \theta > 0$  is  $\max\{0, (\eta - 1)/\theta\}$ . Thus, (25), (26), (27) imply that the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -updates in Algorithm 1 are equivalent to

$$\{\alpha_j\}^{l+1} = \frac{\eta_{\alpha} + M_j/2 - 1}{\theta_{\alpha} + \|F^{(j)}\{\mathbf{x}^{(j)}\}^l - \mathbf{y}^{(j)}\|_2^2/2}, \quad j = 1, \dots, J, \tag{31}$$

$$\{\beta_k^{(j)}\}^{l+1} = \frac{\eta_{\beta} - 1/2}{\theta_{\beta} + [R\{\mathbf{x}^{(j)}\}^l]_k^2/2}, \quad j = 1, \dots, J, \quad k = 1, \dots, K, \tag{32}$$

$$\left\{ \gamma_n^{(j-1,j)} \right\}^l = \frac{\eta_\gamma - 1/2}{\theta_\gamma + \left[ \{\mathbf{x}^{(j-1)}\}^l - \{\mathbf{x}^{(j)}\}^l \right]_n^2 / 2}, \quad j = 2, \dots, J, \quad n = 1, \dots, N, \quad (33)$$

assuming nonnegative numerators, where  $\{ \}^l$  denotes the  $l$ -th iteration of the term inside the curved brackets. Further, (28) implies that the  $\mathbf{x}$ -update in Algorithm 1 is equivalent to solving the linear systems

$$\left\{ G^{(j)} \right\}^{l+1} \left\{ \mathbf{x}^{(j)} \right\}^{l+1} = \left\{ \mathbf{b}^{(j)} \right\}^{l+1}, \quad j = 1, \dots, J, \quad (34)$$

with SPD coefficient matrices

$$\begin{aligned} \left\{ G^{(1)} \right\}^{l+1} &= \{\alpha_1\}^{l+1} (F^{(1)})^T F^{(1)} + R^T \{B^{(1)}\}^{l+1} R + \{C^{(1,2)}\}^{l+1}, \\ \left\{ G^{(j)} \right\}^{l+1} &= \{\alpha_j\}^{l+1} (F^{(j)})^T F^{(j)} + R^T \{B^{(j)}\}^{l+1} R + \{C^{(j-1,j)}\}^{l+1} \\ &\quad + \{C^{(j,j+1)}\}^{l+1}, \\ \left\{ G^{(J)} \right\}^{l+1} &= \{\alpha_J\}^{l+1} (F^{(J)})^T F^{(J)} + R^T \{B^{(J)}\}^{l+1} R + \{C^{(J-1,J)}\}^{l+1}, \end{aligned} \quad (35)$$

and right-hand sides

$$\begin{aligned} \left\{ \mathbf{b}^{(1)} \right\}^{l+1} &= \{\alpha_1\}^{l+1} (F^{(1)})^T \mathbf{y}^{(1)} + \{C^{(1,2)}\}^{l+1} \{\mathbf{x}^{(2)}\}^l, \\ \left\{ \mathbf{b}^{(j)} \right\}^{l+1} &= \{\alpha_j\}^{l+1} (F^{(j)})^T \mathbf{y}^{(j)} + \{C^{(j-1,j)}\}^{l+1} \{\mathbf{x}^{(j-1)}\}^l \\ &\quad + \{C^{(j,j+1)}\}^{l+1} \{\mathbf{x}^{(j+1)}\}^l, \\ \left\{ \mathbf{b}^{(J)} \right\}^{l+1} &= \{\alpha_J\}^{l+1} (F^{(J)})^T \mathbf{y}^{(J)} + \{C^{(J-1,J)}\}^{l+1} \{\mathbf{x}^{(J-1)}\}^l, \end{aligned} \quad (36)$$

for  $j = 2, \dots, J - 1$ . Notably, we compute the  $(l + 1)$ -th iteration of the  $\mathbf{x}^{(j)}$ 's using the  $l$ -th iteration of the neighboring images in (36). This allows us to parallelize the  $\mathbf{x}^{(j)}$ -updates (34), making our method efficient even for large image sequences. We can now summarize our JHBL method for sequential image recovery as in Algorithm 2.

---

**Algorithm 2** Joint hierarchical Bayesian learning (JHBL) algorithm

---

- 1: Initialize the images  $\{\mathbf{x}^{(1)}\}^0, \dots, \{\mathbf{x}^{(J)}\}^0$  and set  $l = 0$
  - 2: **repeat**
  - 3:   Update the noise parameters  $\alpha_1, \dots, \alpha_J$  according to (31)
  - 4:   Update the intra-image parameters  $\beta^{(1)}, \dots, \beta^{(J)}$  according to (32)
  - 5:   Update the inter-image parameters  $\gamma^{(1,2)}, \dots, \gamma^{(J-1,J)}$  according to (33)
  - 6:   Update the images  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(J)}$  according to (34)
  - 7:   Increase  $l \rightarrow l + 1$
  - 8: **until** convergence or maximum number of iterations is reached
- 

**3.3 Separate Recovery as a Special Case**

We can recover the generalized sparse Bayesian learning (GSBL) method [24] from our JHBL procedure as a special/limit case. If  $\eta_\gamma \leq 1/2$  or  $\theta_\gamma \rightarrow \infty$ , then the  $\gamma^{(j)}$ -update (33)

becomes

$$\{\mathbf{y}^{(j-1,j)}\}^l = 0, \quad j = 2, \dots, J, \quad (37)$$

and the  $\mathbf{x}^{(j)}$ -update (34) reduces to

$$\left(\{\alpha_j\}^{l+1}(F^{(j)})^T F^{(j)} + R^T \{B^{(j)}\}^{l+1} R\right) \{\mathbf{x}^{(j)}\}^{l+1} = \{\alpha_j\}^{l+1} (F^{(j)})^T \mathbf{y}^{(j)}, \quad (38)$$

for  $j = 1, \dots, J$ . In this case, Algorithm 2 corresponds to using the GSBL algorithm [24] to recover the images separately.

### 3.4 Efficient Implementation

A few remarks on Algorithm 2 are in order.

**Remark 5** [Initialization] We initialize the images  $\{\mathbf{x}^{(1)}\}^0, \dots, \{\mathbf{x}^{(J)}\}^0$  in Algorithm 2 as the separately recovered images using the GSBL algorithm [24], which we efficiently implemented in parallel.

**Remark 6** [Parallelization] The different  $\mathbf{x}^{j-}$ ,  $\alpha_j$ -,  $\beta^{(j)}$ -, and  $\mathbf{y}^{(j)}$ -updates can be easily parallelized. This makes Algorithm 2 efficient even for large image sequences.

**Remark 7** [Efficient  $\mathbf{x}^{(j)}$ -updates] The coefficient matrices in the  $\mathbf{x}^{(j)}$ -updates (34) can become prohibitively large in imaging applications. To avoid storage and efficiency issues, we identify the solutions of (34) with the unique minimizers of the quadratic functionals

$$L^{(j)}(\mathbf{x}) = \mathbf{x}^T G^{(j)} \mathbf{x} - 2\mathbf{x}^T \mathbf{b}^{(j)}, \quad j = 1, \dots, J, \quad (39)$$

which is possible since the  $G^{(j)}$ 's are (almost surely) SPD. We then efficiently solve for the minimizers of (39) using a gradient descent method [24]. In our implementation, we use five gradient descent steps for every iteration of the  $\mathbf{x}^{(j)}$ -updates.

**Remark 8** [Stopping criterion] We stop the iterations in Algorithm 2 if the average relative and absolute change between two subsequent image sequence iterations w. r. t. the  $\|\cdot\|_2$ -norm are less than  $10^{-3}$  or if a maximum number of  $10^3$  iterations is reached.

**Remark 9** [Convexity and convergence] A detailed analysis regarding the convexity of the cost function,  $-\log p(\mathbf{x}, \alpha, \beta, \mathbf{y}|\mathbf{y})$ , and the convergence of Algorithm 2 exceeds the scope of this paper and will be addressed in future works.

## 4 Numerical Tests

We consider two numerical tests to demonstrate the performance of our JHBL method.

### 4.1 Sequential Magnetic Resonance Imaging

Given is a temporal sequence of six  $128 \times 128$  phantom images obtained by a GE HTXT 1.5T clinical magnetic resonance imaging (MRI) scanner [33], from which we generate under-sampled and indirect Fourier data. Figure 4a, b and c show the first three reference images,

where the change consists of the rotating left (green) ellipse and the down-moving right (yellow) ellipse. The Fourier samples contained in the data sets are

$$y_{k,l}^{(j)} = \int_0^1 \int_0^1 x^{(j)}(s,t) e^{-i2\pi(ks+lt)} ds dt, \quad -\left\lceil \frac{N_1}{2} \right\rceil \leq k, l < \left\lceil \frac{N_1}{2} \right\rceil, \quad (40)$$

for  $j = 1, \dots, J$ , where  $N = N_1^2$  and  $x^{(j)}$  is a function describing the  $j$ th image. We use the discrete Fourier transform as a linear forward operator, thereby introducing model discrepancy and avoiding the inverse crime [30]. Further, each data set is missing Fourier samples for the symmetric bands

$$\mathcal{K}_j = [\pm(10j + 1), \pm(10j + 10)]^2, \quad j = 1, \dots, 6, \quad (41)$$

and the remaining Fourier samples contain additive i.i.d. zero-mean normal noise. The amount of noise is measured using the signal-to-noise ratio (SNR)

$$\text{SNR}^{(j)} = 10 \log_{10} \left( \alpha_j y_{0,0}^{(j)} \right), \quad j = 1, \dots, J, \quad (42)$$

where  $y_{0,0}^{(j)}$  is the average of the  $j$ th image. The SNR for all images in Fig. 4 is 2.

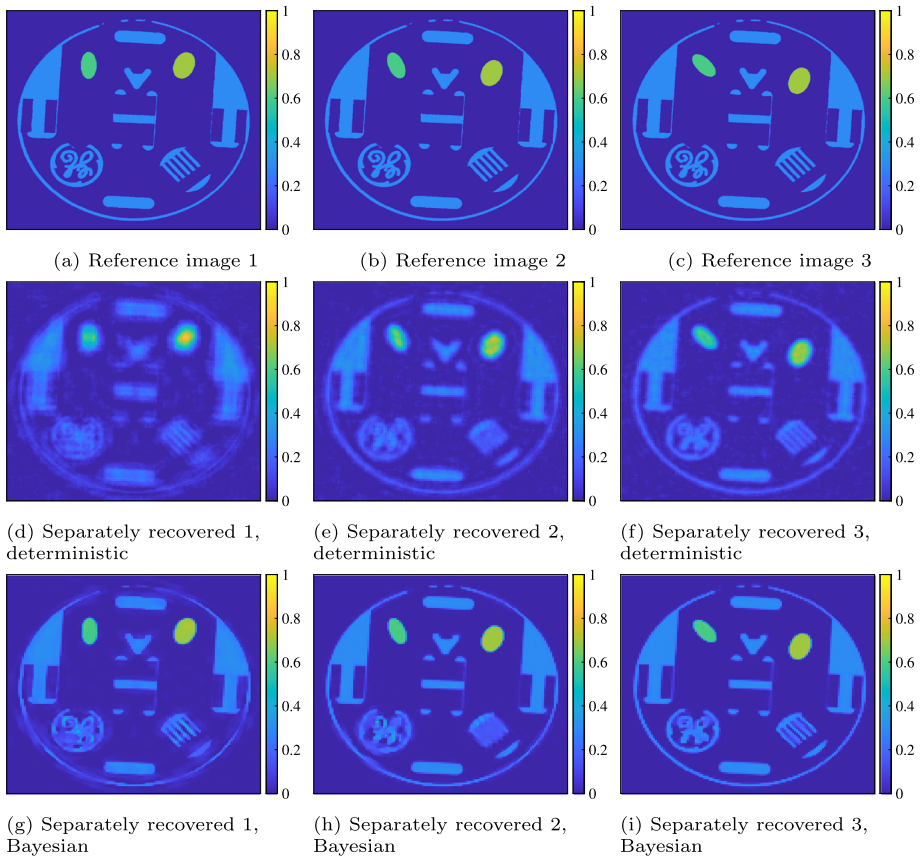
Figure 4d, e and f illustrate the separately recovered images from the noisy and under-sampled Fourier data by solving the (weighted)  $\ell^1$ -regularized inverse problems (2) using the alternating directions method of multipliers (ADMM) [7, 42]. Figure 4g, h and i visualizes the separately recovered images from the same data sets using the GSBL algorithm [24] with hyper-parameters  $\eta_\alpha = \eta_\beta = 1$  and  $\theta_\alpha = \theta_\beta = 10^{-3}$ . In both cases, we used an anisotropic first-order TV regularization operator

$$R = \begin{bmatrix} I \otimes D \\ D \otimes I \end{bmatrix} \quad \text{with} \quad D = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(N_1-1) \times N_1}, \quad (43)$$

to promote the images being piecewise constant. Figure 4 demonstrates that the Bayesian GSBL algorithm separately recovers the images more accurately than the deterministic algorithm, although we still observe some smeared features in Fig. 4g and h.

Figure 5 illustrates the corresponding jointly recovered images. Specifically, Fig. 5a, b and c visualizes the jointly recovered images using ADMM to solve the joint  $\ell^1$ -regularized inverse problem (3) as proposed in [42], which we use as a benchmark. Figure 5d, e and f show the jointly recovered images using our JHBL algorithm (Algorithm 2). Following the discussion in Sect. 2, we chose the hyper-parameters as  $\eta_\alpha = \eta_\beta = 1$ ,  $\eta_\gamma = 2$ , and  $\theta_\alpha = \theta_\beta = \theta_\gamma = 10^{-3}$ .

**Remark 10** Following [24],  $\eta_\alpha = \eta_\beta = 1$  and  $\theta_\alpha = \theta_\beta = 10^{-3}$  are usual choices for the GSBL algorithm. Initially, we also tried to use  $\eta_\gamma = 1$  and  $\theta_\gamma = 10^{-3}$  for the parameters of the inter-image hyper-prior, but observed that the inter-image coupling was sometimes too strong. We thus used  $\eta_\gamma = 2$  and  $\theta_\gamma = 10^{-3}$  instead. The heuristic behind increasing the shape parameter to  $\eta_\gamma$  is as follows: We assume that every consecutive pair of images contains more no-change regions than change regions. We thus expect  $\gamma_n^{(j-1,j)} \gg 0$  for most of the  $\gamma_n^{(j-1,j)}$ 's in (12), indicating no change, and  $\gamma_n^{(j-1,j)} \approx 0$  for only a few of the  $\gamma_n^{(j-1,j)}$ 's. We increasingly promote this behavior for the  $\gamma_n^{(j-1,j)}$ 's by increasing the shape parameter  $\eta_\gamma$  in (13). We further investigate the influence of the shape and rate parameter  $\eta_\gamma$  and  $\theta_\gamma$  on the inter-image coupling in Sect. 4.3.



**Fig. 4** Three temporal phantom images by a GE HTXT 1.5T clinical MRI scanner [33] (first row) and their separate reconstructions from noisy and under-sampled Fourier data solving the deterministic  $\ell^1$ -regularized inverse problems (2) (second row) and the Bayesian GSB algorithm (third row)

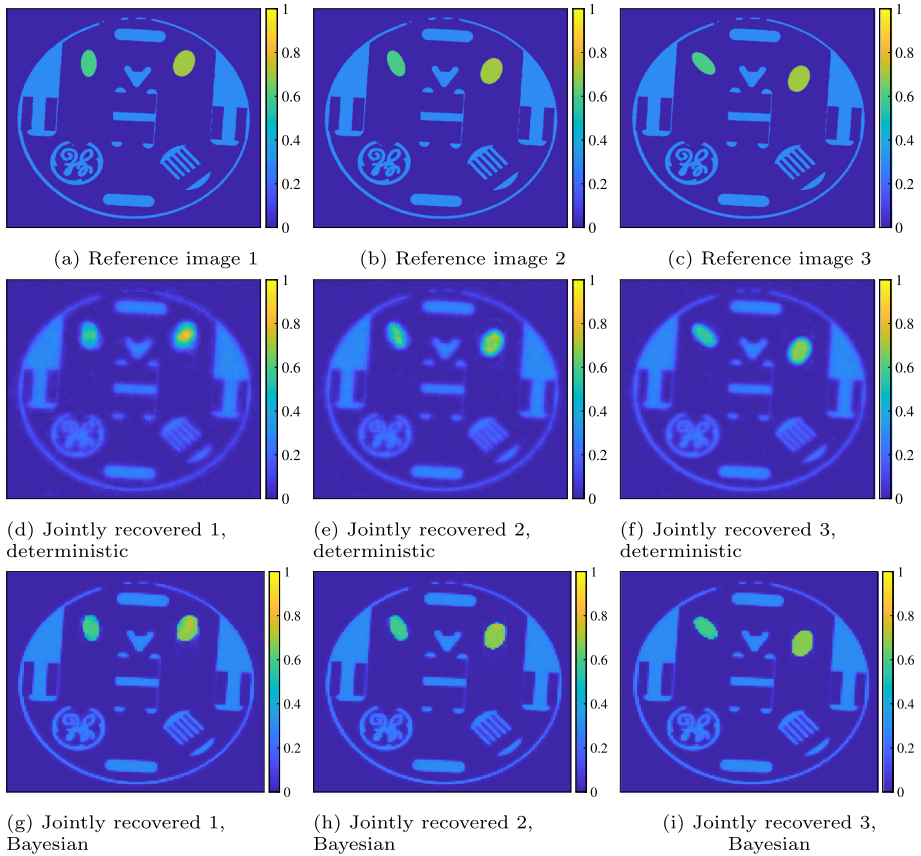
Both joint methods yield more accurate recovered images than the respective separate method. At the same time, our JHBL algorithm provides notably more accurate recovered images than the deterministic joint method proposed in [42].

We observed the proposed JHBL method to yield the most accurate recovered images for all combinations of Fourier samples and SNRs we considered. Figure 6a and b show the average relative log-error for different numbers of Fourier samples and SNRs. The relative log-error of the  $j$ th image is

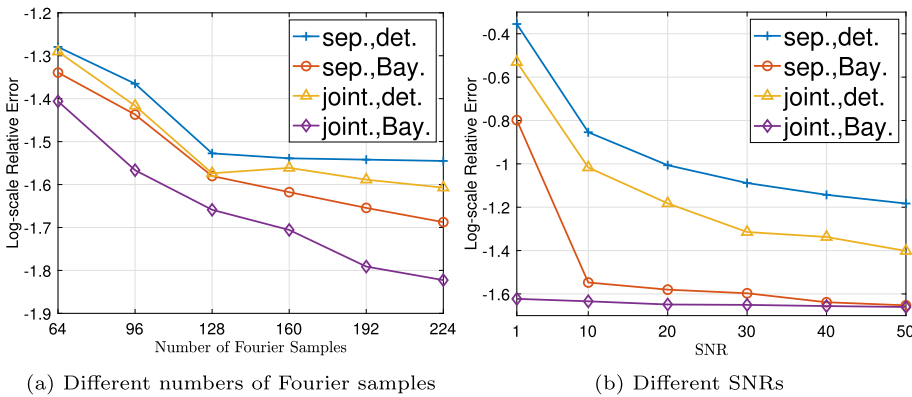
$$E_{\log}^{(j)} = \log_{10} \left( \frac{\|\mathbf{x}_{\text{ref}}^{(j)} - \mathbf{x}^{(j)}\|_2}{\|\mathbf{x}_{\text{ref}}^{(j)}\|_2} \right), \quad j = 1, \dots, J, \tag{44}$$

where  $\mathbf{x}_{\text{ref}}^{(j)}$  and  $\mathbf{x}^{(j)}$  are the reference and recovered image, respectively. The average relative log-error is  $(E_{\log}^{(1)} + \dots + E_{\log}^{(J)})/J$ .

Another advantage of our JHBL method is that it allows us to quantify uncertainty. Indeed, we recover a full Gaussian distribution,  $\mathcal{N}(\boldsymbol{\mu}^{(j)}, \boldsymbol{\Sigma}^{(j)})$ , for every individual image, which is

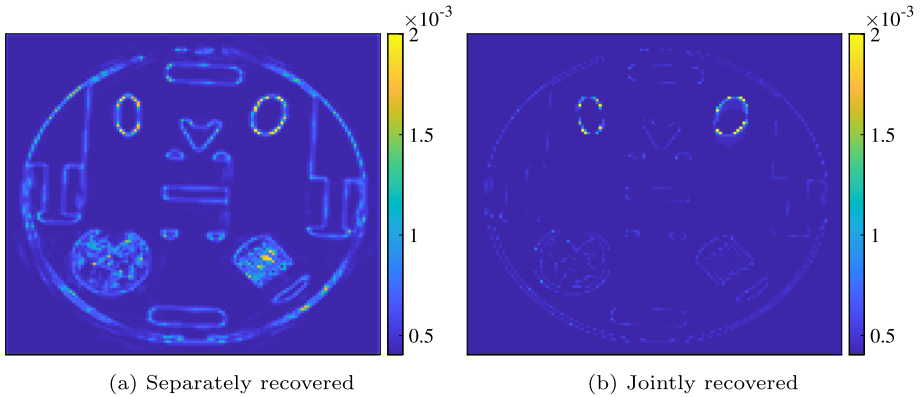


**Fig. 5** Three temporal phantom images by a GE HTXT 1.5T clinical MRI scanner [33] (first row) and their joint reconstructions from noisy and under-sampled Fourier data using the deterministic weighted  $\ell^1$ -regularized inverse problems (3) (second row) and the proposed JHBL algorithm (third row)

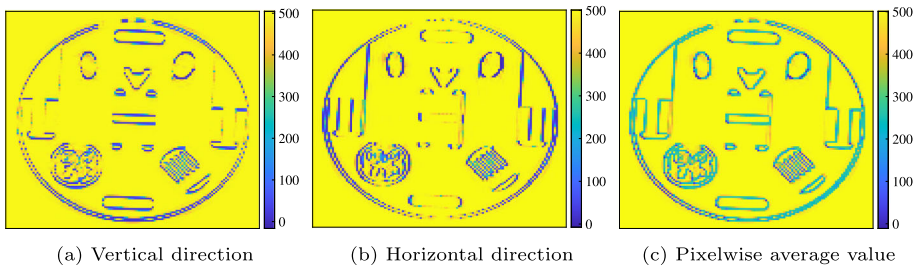


**Fig. 6** Average relative log-errors for the separate deterministic (blue strokes), the separate Bayesian (orange circles), the joint deterministic (yellow triangles), and our joint Bayesian (purple diamonds) algorithm. In Fig. 6a, the SNR is 2. In Fig. 6b, we started with  $128 \times 128$  Fourier samples and then removed the bands in (41) (Color figure online)





**Fig. 7** Pixelwise variances of the recovered first image. Jointly recovering the images by “borrowing” missing information from the other images reduces uncertainty in no-change regions

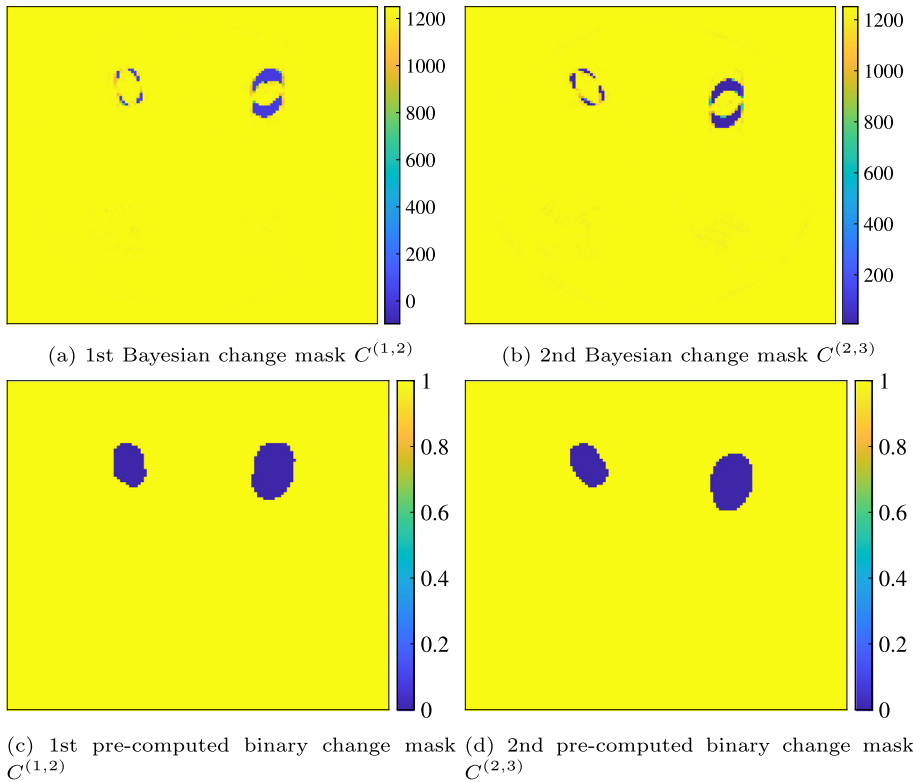


**Fig. 8** The final estimates for the intra-image regularization parameters in the vertical and horizontal direction for the first image in Fig. 5 and their pixelwise average

conditioned on the observed data sets and the other estimated parameters. To demonstrate this, Fig. 7a and b illustrate the pixelwise variance of the separately and jointly recovered first image using the GSBL and our JHBL algorithm. The pixelwise variance corresponds to the diagonal elements of the covariance matrices (29). Comparing Fig. 7a and b, we observe reduced uncertainty in the jointly recovered first image, except for the change regions around the two ellipses. The reduced uncertainty of the recovered first image by our JHBL method away from the change regions is due to the method “borrowing” information from the neighboring images.

Another convenient by-product of our JHBL method is that it allows for edge and change detection, which we illustrate in Figs. 8 and 9. Figure 8a and b respectively visualize the final estimate of the first and second half of the intra-image regularization parameter  $\beta^{(1)}$  of our JHBL method for the first recovered image. Since we used an anisotropic first-order TV operator (43), the intra-image regularization parameter values indicate edges in the vertical and the horizontal direction. We can combine them by considering the pixelwise average of the images in Fig. 8a and b to obtain the edge profile in Fig. 8c.

Figure 9a and b visualize the final estimate of the conditional change mask used by our JHBL method for the first and second pair of images, while Fig. 9c and d illustrates the corresponding pre-computed binary change masks used in [42]. The pre-computed binary change masks rely on the Fourier data sets and sequential images containing only objects



**Fig. 9** The final estimates for the Bayesian change masks (top row) and pre-computed binary change masks (bottom row) used by the Bayesian and deterministic method to jointly recover the temporal image sequence in Fig. 5

with closed boundaries. By contrast, the change masks used in our JHBL method neither rely on Fourier data nor the sequential images containing only objects with closed boundaries.

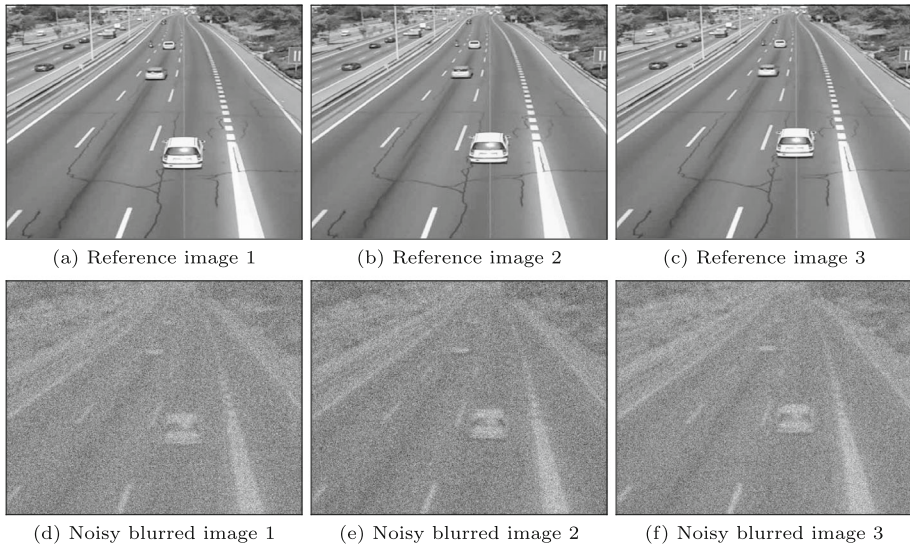
### 4.2 Sequential Image Deblurring

We next consider the deconvolution of a temporal sequence of six  $400 \times 400$  images from the GRAM road-traffic monitoring data set [26]. Figure 10 illustrates the first three reference images and their noisy blurred versions. The i.i.d. normal noise  $\mathbf{e}^{(j)}$  in the corresponding linear data model (1) has  $\text{SNR} = 2 + j, j = 1, \dots, J$ . The forward operator  $F$  is obtained by applying the tensor-product midpoint quadrature to the convolution equations

$$y^{(j)}(s, t) = \int_0^1 \int_0^1 k(s - s', t - t')x^{(j)}(s, t) ds' dt', \quad j = 1, \dots, J, \tag{45}$$

where  $x^{(j)}$  is a function describing the  $j$ th image. We assume a Gaussian convolution kernel

$$k(s, t) = \frac{1}{2\pi\gamma^2} \exp\left(-\frac{s^2 + t^2}{2\gamma^2}\right) \tag{46}$$



**Fig. 10** The first three reference images of a temporal image sequence coming from the GRAM road-traffic monitoring data set [26] (first row) and their noisy blurred versions (second row)

with blurring parameter  $\gamma = 5 \cdot 10^{-3}$ , which makes  $F$  highly ill-conditioned. We use an anisotropic second-order TV regularization operator

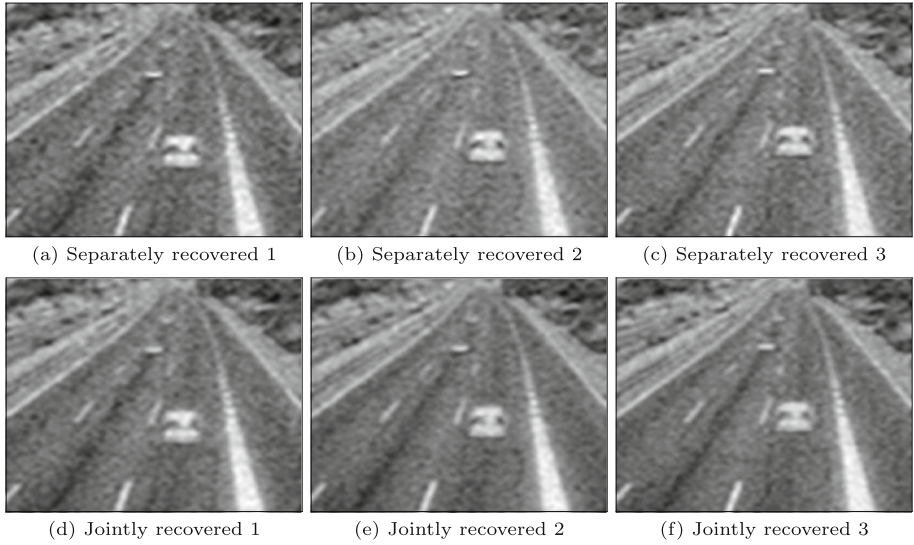
$$R = \begin{bmatrix} I \otimes D \\ D \otimes I \end{bmatrix} \quad \text{with} \quad D = \begin{bmatrix} -1 & 2 & -1 \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(N_1-2) \times N_1}, \quad (47)$$

where  $N_1 = 400$  is the number of pixels in each direction to promote the images being piecewise smooth—but not necessarily piecewise constant.

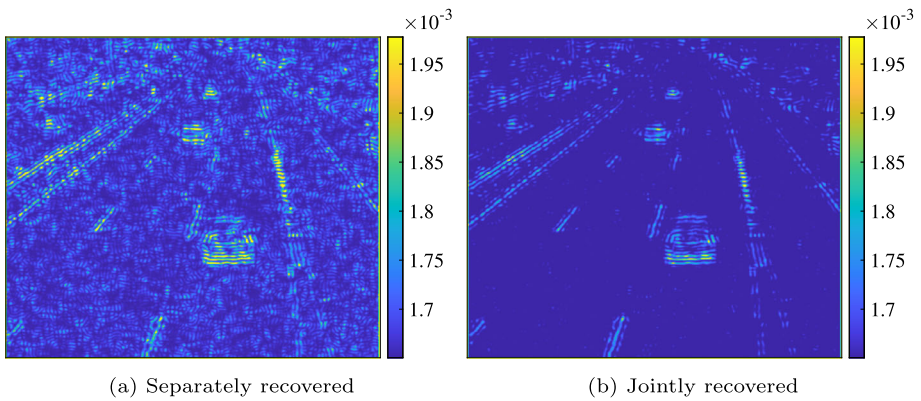
We can no longer use the deterministic method [42] to pre-compute binary change masks directly from the indirect data sets. However, we can still use our JHBL method to jointly recover the sequential images in a Bayesian setting. Figure 11 illustrates the separately (top row) and jointly (bottom row) recovered images from the noisy blurred images in Fig. 10d, e and f using the GSBL and our JHBL algorithm, respectively. The hyper-parameters are  $\eta_\alpha = \eta_\beta = 1, \eta_\gamma = 1, \theta_\alpha = \theta_\beta = 10^{-3}$ , and  $\theta_\gamma = 10^{-1}$ . The jointly recovered images using our JHBL algorithm are more accurate than the separately recovered images.

As mentioned, the proposed JHBL method can quantify uncertainty in the recovered images with increased reliability, which is often desirable in applications with no reference images. We demonstrate this in Fig. 12, which visualizes the pixelwise variances of the recovered first image in Fig. 11. We again see that jointly recovering the images by “borrowing” missing information from the other images reduces uncertainty.

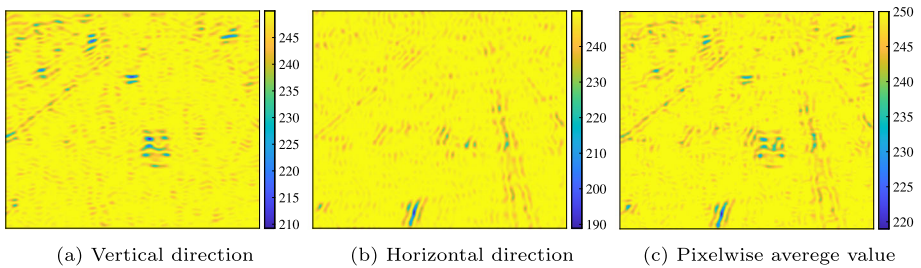
Moreover, Fig. 13a and b illustrate the final estimate of the first and second half of the intra-image regularization parameter  $\beta^{(1)}$  of our JHBL method for the first recovered image in Fig. 11d. Since we used an anisotropic second-order TV regularization operator (47), the intra-image regularization parameter values indicate edges in the vertical and horizontal direction. We can again combine them, e.g., by considering the pixelwise average of the images in Fig. 13a and b to obtain the edge profile in Fig. 13c.



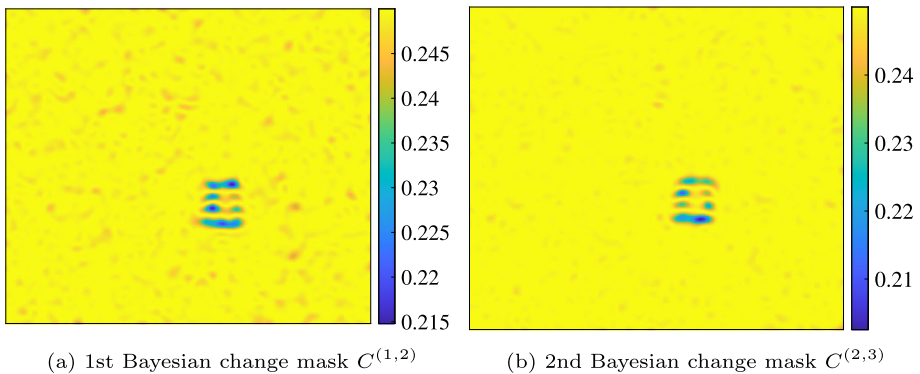
**Fig. 11** Separately recovered images using GSBL (top row) and the jointly recovered images using our JHBL algorithm (bottom row)



**Fig. 12** Pixelwise variances of the recovered first image in Fig. 11. Jointly recovering the images by “borrowing” missing information from the other images reduces uncertainty



**Fig. 13** The final estimates for the intra-image regularization parameters in the vertical and horizontal direction for the first recovered image in Fig. 11 using our JHBL algorithm and their pixelwise average



**Fig. 14** The final estimates for the Bayesian change masks for the first and second pair of images in Fig. 11 using our JHBL algorithm

Finally, Fig. 14 provides the final estimates for the Bayesian change mask for the first and second pair of images in Fig. 11 using our JHBL algorithm only.

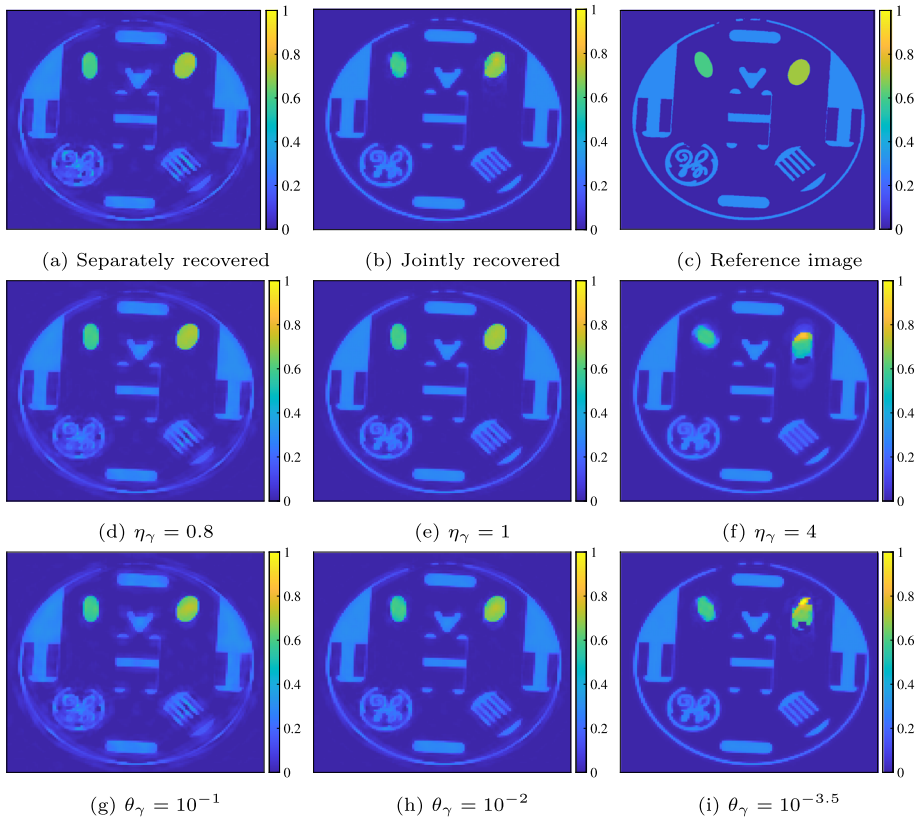
### 4.3 Investigating the Influence of the Shape and Rate Parameter

We end this section by briefly demonstrating how different choices for the shape and rate parameter,  $\eta_\gamma$  and  $\theta_\gamma$ , of the inter-image hyper-prior (see Sect. 2.3) influence the jointly recovered image. Recall (e.g., from Remark 10) that we expect the coupling between images to increase/decrease as smaller/larger values for the inter-image hyper-parameters  $\gamma_n^{(j-1,j)}$  become more likely. At the same time, the expected value and variance of the inter-image hyper-prior (13) are  $\eta_\gamma/\theta_\gamma$  and  $\eta_\gamma/\theta_\gamma^2$ , respectively. Hence, if  $\eta_\gamma$  is increased, we expect larger values for  $\gamma_n^{(j-1,j)}$  to become more likely and the inter-image coupling to decrease. On the other hand, if  $\theta_\gamma$  is increased, we expect smaller values for  $\gamma_n^{(j-1,j)}$  to become more likely and the inter-image coupling to increase.

Figure 15 illustrates the connection between  $\eta_\gamma, \theta_\gamma$  and the strength of the inter-image coupling for the first phantom image from the sequential MRI test case previously discussed in Sect. 4.1. Specifically, the second row of Fig. 15 shows that the jointly recovered image is visibly close to the separately recovered image for  $\eta_\gamma = 0.8$ . At the same time, the inter-image coupling becomes so strong that some spurious artifacts from the subsequent images are introduced in change regions for  $\eta_\gamma = 4$ . The third row of Fig. 15 demonstrates a similar behavior for fixed  $\eta_\gamma = 2$  and decreasing  $\theta_\gamma$ ; The jointly recovered image is visibly close to the separately recovered image for  $\theta_\gamma = 10^{-1}$ , while the inter-image coupling becomes so strong that some spurious artifacts are introduced in change regions for  $\theta_\gamma = 10^{-3.5}$ . Future work will optimize the shape and rate parameter selection of the inter-image hyper-prior to further increase the advantage of inter-image coupling between sequential images.

## 5 Summary

We presented a new method to jointly recover temporal image sequences by “borrowing” missing information in each image from the other images. Our JHBL method is simple to



**Fig. 15** Demonstrating the influence of the shape and rate parameter,  $\eta_\gamma$  and  $\theta_\gamma$ , on the inter-image coupling. First row: First phantom reference image, its joint reconstruction using the JHBL algorithm with  $\eta_\gamma = 2$  and  $\eta_\gamma = 10^{-3}$ , and its separate reconstruction using the GSB algorithm. Second row: Joint reconstructions for fixed  $\eta_\gamma = 10^{-3}$  and varying  $\eta_\gamma$ . Third row: Joint reconstructions for fixed  $\theta_\gamma = 2$  and varying  $\theta_\gamma$

implement, easily parallelized, and efficient. We found our method to yield more accurate recovered images than both, separately recovering the images using the Bayesian GSB algorithm and jointly recovering them using the deterministic method from [42]. In addition, our method avoids exhaustive parameter fine-tuning. Moreover, the deterministic method proposed in [42] is limited to Fourier data sets and images only containing objects with closed boundaries to pre-compute a binary change mask. By contrast, we treat the change mask that steers the coupling between neighboring images as a random variable, which we estimate together with the images and other parameters, making our method applicable to general modalities. We demonstrated this by considering a sequential image deblurring problem based on the GRAM road-traffic monitoring data set. Another distinct advantage of our method is that it allows us to quantify uncertainty, which is vital in applications without reference images. An additional valuable by-product is that our method allows for edge and change detection.

Future work will address the selection of hyper-parameters. Future efforts will also investigate the structural properties of the cost function and convergence of the proposed JHBL algorithm. Finally, a comparison or combination of the proposed BCD algorithm for Bayesian inference with other existing methods would be of interest.

**Acknowledgements** We thank Guohui Song and the anonymous reviewers for helpful feedback on an earlier version of this manuscript.

**Funding** 'Open Access funding provided by the MIT Libraries'

**Data Availability** All data sets are publicly available. MATLAB codes used for all numerical results are available from the authors upon request.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Adcock, B., Gelb, A., Song, G., Sui, Y.: Joint sparse recovery based on variances. *SIAM J. Sci. Comput.* **41**(1), A246–A268 (2019)
2. Archibald, R., Gelb, A., Platte, R.B.: Image reconstruction from undersampled Fourier data using the polynomial annihilation transform. *J. Sci. Comput.* **67**(2), 432–452 (2016)
3. Archibald, R., Gelb, A., Yoon, J.: Polynomial fitting for edge detection in irregularly sampled signals and images. *SIAM J. Numer. Anal.* **43**(1), 259–279 (2005)
4. Babacan, S.D., Molina, R., Katsaggelos, A.K.: Parameter estimation in TV image restoration using variational distribution approximation. *IEEE Trans. Image Process.* **17**(3), 326–339 (2008)
5. Bardsley, J.M.: MCMC-based image reconstruction with uncertainty quantification. *J. Sci. Comput.* **34**(3), A1316–A1332 (2012)
6. Benfold, B., Reid, I.: Stable multi-target tracking in real-time surveillance video. In: *CVPR 2011*, pp. 3457–3464. IEEE (2011)
7. Boyd, S., Parikh, N., Chu, E.: *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Now Publishers Inc (2011)
8. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**(3), 492–526 (2010)
9. Calvetti, D., Pragliola, M., Somersalo, E.: Sparsity promoting hybrid solvers for hierarchical Bayesian inverse problems. *J. Sci. Comput.* **42**(6), A3761–A3784 (2020)
10. Calvetti, D., Somersalo, E.: A Gaussian hypermodel to recover blocky objects. *Inverse Probl.* **23**(2), 733 (2007)
11. Calvetti, D., Somersalo, E.: *An Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, vol. 2. Springer Science & Business Media, London (2007)
12. Calvetti, D., Somersalo, E.: Hypermodels in the Bayesian imaging framework. *Inverse Probl.* **24**(3), 034013 (2008)
13. Calvetti, D., Somersalo, E., Strang, A.: Hierarchical Bayesian models and sparsity:  $\ell_2$ -magic. *Inverse Probl.* **35**(3), 035003 (2019)
14. Candes, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted  $\ell_1$  minimization. *J. Fourier Anal. Appl.* **14**(5), 877–905 (2008)
15. Cotter, S.F., Rao, B.D., Engan, K., Kreutz-Delgado, K.: Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Process.* **53**(7), 2477–2488 (2005)
16. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006)
17. Ehrhardt, M.J., Thielemans, K., Pizarro, L., Atkinson, D., Ourselin, S., Hutton, B.F., Arridge, S.R.: Joint reconstruction of PET-MRI by exploiting structural similarity. *Inverse Probl.* **31**(1), 015001 (2014)

18. Eldar, Y.C., Kutyniok, G.: *Compressed Sensing: Theory and Applications*. Cambridge University Press, Cambridge (2012)
19. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. In: *ACM SIGGRAPH 2006 Papers*, pp. 787–794. ACM (2006)
20. Figueiredo, M.A., Bioucas-Dias, J.M., Nowak, R.D.: Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. Image Process.* **16**(12), 2980–2991 (2007)
21. Foucart, S., Rauhut, H.: A mathematical introduction to compressive sensing. *Bull. Amer. Math. Soc.* **54**, 151–165 (2017)
22. Gelb, A., Scarnati, T.: Reducing effects of bad data using variance based joint sparsity recovery. *J. Sci. Comput.* **78**(1), 94–120 (2019)
23. Glaubitz, J., Gelb, A.: High order edge sensors with  $\ell^1$  regularization for enhanced discontinuous Galerkin methods. *SIAM J. Sci. Comput.* **41**(2), A1304–A1330 (2019)
24. Glaubitz, J., Gelb, A., Song, G.: Generalized sparse Bayesian learning and application to image reconstruction. *SIAM/ASA J. Uncertain. Quantif.* **11**(1), 262–284 (2023)
25. Groetsch, C.W.: *Inverse Problems in the Mathematical Sciences*, vol. 52. Springer, London (1993)
26. Guerrero-Gomez-Olmedo, R., Lopez-Sastre, R.J., Maldonado-Bascon, S., Fernandez-Caballero, A.: Vehicle tracking by simultaneous detection and viewpoint estimation. In: *IWINAC 2013, Part II, LNCS 7931*, pp. 306–316 (2013)
27. Hansen, P.C.: *Discrete Inverse Problems: Insight and Algorithms*. SIAM. University City Science Center Philadelphia, PA, USA (2010)
28. Hu, W., Xie, D., Fu, Z., Zeng, W., Maybank, S.: Semantic-based surveillance video retrieval. *IEEE Trans. Image Process.* **16**(4), 1168–1181 (2007)
29. Kaipio, J., Somersalo, E.: *Statistical and Computational Inverse Problems*, vol. 160. Springer Science & Business Media, London (2006)
30. Kaipio, J., Somersalo, E.: Statistical inverse problems: discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.* **198**(2), 493–504 (2007)
31. Kaipio, J.P., Kolehmainen, V., Somersalo, E., Vauhkonen, M.: Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Probl.* **16**(5), 1487 (2000)
32. Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-Laplacian priors. *Adv. Neural. Inf. Process. Syst.* **22**, 1033–1041 (2009)
33. Lalwani, G., Livingston Sundararaj, J., Schaefer, K., Button, T., Sitharaman, B.: Synthesis, characterization, in vitro phantom imaging, and cytotoxicity of a novel graphene-based multimodal magnetic resonance imaging - X-ray computed tomography contrast agent. *J. Mater. Chem. B* **2**(22), 3519–3530 (2015)
34. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* **26**(3), 70 (2007)
35. Parisotto, S., Lellmann, J., Masnou, S., Schönlieb, C.B.: Higher-order total directional variation: Imaging applications. *SIAM J. Imaging Sci.* **13**(4), 2063–2104 (2020)
36. Parisotto, S., Masnou, S., Schönlieb, C.B.: Higher-order total directional variation: Analysis. *SIAM J. Imaging Sci.* **13**(1), 474–496 (2020)
37. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D* **60**(1–4), 259–268 (1992)
38. Stuart, A.M.: Inverse problems: a Bayesian perspective. *Acta Numer.* **19**, 451–559 (2010)
39. Tipping, M.E.: Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1**, 211–244 (2001)
40. Vogel, C.R.: *Computational Methods for Inverse Problems*. SIAM. University City Science Center Philadelphia, PA, USA (2002)
41. Wipf, D.P., Rao, B.D.: An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* **55**(7), 3704–3716 (2007)
42. Xiao, Y., Glaubitz, J., Gelb, A., Song, G.: Sequential image recovery from noisy and under-sampled Fourier data. *J. Sci. Comput.* **91**(3), 1–29 (2022)
43. Yang, X., Shi, J., Zhou, Y., Wang, C., Hu, Y., Zhang, X., Wei, S.: Ground moving target tracking and refocusing using shadow in video-SAR. *Remote Sens.* **12**(18), 3083 (2020)
44. Zhang, Z., Rao, B.D.: Clarify some issues on the sparse Bayesian learning for sparse signal recovery. Technical Report. University of California at San Diego, La Jolla, CA, USA (2011)
45. Zhang, Z., Rao, B.D.: Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J. Sel. Top. Signal Process.* **5**(5), 912–926 (2011)