University of Dundee

Deconvolution and Restoration of Optical Endomicroscopy Images

Karam Eldaly, Ahmed ; Altmann, Yoann; Perperidis, Antonios; Krstajic, Nikola; Choudhary, Tushar R. ; Dhaliwal, Kevin

[Link to publication in Discovery Research Portal]

Deconvolution and Restoration of Optical Endomicroscopy Images

Ahmed Karam Eldaly, *Student Member, IEEE,* Yoann Altmann, *Member, IEEE,* Antonios Perperidis, Nikola Krstajić, Tushar R. Choudhary, Kevin Dhaliwal, and Stephen McLaughlin, *Fellow, IEEE*

Abstract—Optical endomicroscopy (OEM) is an emerging technology platform with preclinical and clinical imaging applications. Pulmonary OEM via fibre bundles has the potential to provide *in vivo, in situ* molecular signatures of disease such as infection and inflammation. However, enhancing the quality of data acquired by this technique for better visualization and subsequent analysis remains a challenging problem. Cross coupling between fiber cores and sparse sampling by imaging fiber bundles are the main reasons for image degradation, and poor detection performance (i.e., inflammation, bacteria, etc.). In this work, we address the problem of deconvolution and restoration of OEM data. We propose a hierarchical Bayesian model to solve this problem and compare three estimation algorithms to exploit the resulting joint posterior distribution. The first method is based on Markov chain Monte Carlo (MCMC) methods, however, it exhibits a relatively long computational time. The second and third algorithms deal with this issue and are based on a variational Bayes (VB) approach and an alternating direction method of multipliers (ADMM) algorithm respectively. Results on both synthetic and real datasets illustrate the effectiveness of the proposed methods for restoration of OEM images.

Index Terms—Optical endomicroscopy, Deconvolution, Image restoration, Irregular sampling, Bayesian models.

I. INTRODUCTION

PNEUMONIA is a major cause of morbidity and mortality in mechanically ventilated patients in intensive care [1]. However, the accurate diagnosis and monitoring of suspected pneumonia remain challenging [2]. Current methodologies consist of culturing bronchoalveolar lavage fluid (BALF) retrieved from bronchoscopy, but this often takes 48 hours to yield a result which still has low specificity and sensitivity [3]. Structural imaging with X-ray or computed tomography (CT) scans are also often non-diagnostic.

Optical endomicroscopy (OEM) is an emerging, optical fibre-based medical imaging modality with utility in a range of clinical indications and organ systems, including gastro-intestinal, urological and respiratory tracts. The technology employs a proximal light source, laser scanning or Light Emitting Diode (LED) illumination, linked to a flexible fibre bundle, performing microscopic fluorescent imaging at its distal end. The diameter of the packaged fibre can be $< 500\,\mu\mathrm{m}$, enabling the real-time imaging of tissues that were previously inaccessible through conventional endoscopy. Probe-based confocal laser endomicroscopy, is currently the most widely used clinical OEM platform approved for clinical use. However, there have recently been a number of studies describing novel, flexible, versatile and low-cost OEM architectures [4]–[6], employing wide-field LED illumination sources, capable of imaging at multiple acquisition wavelengths [7]. Wide-field fiber optic imaging devices, such as the one being developed by our group provide sparse and usually irregularly-spaced intensity readings of the scene, due to the irregular packing of the fibre cores within the fibre bundle. Fibre bundles usually contain approximately 25,000 fibre cores that are transmitting and collecting the light simultaneously. Note that it is only the fibre cores which contain information while the cladding, (the space between the fibre cores), does not.

One of the main challenges of OEM images is enhancing the restoration of the signals at the receiver for better image visualization and/or subsequent analysis. Fiber core cross coupling is one of the main reasons for image degradation in this type of imaging [8], [9]. In confocal endomicroscopy, the detector pinhole can mask out light coupled to neighbouring cores before reaching the detector. Consequently, the effect of inter-core coupling in imaging capabilities is inherently of greater importance in wide-field endomicroscopy. Perperidis et al. [10] have quantified the average spread of inter-core coupled light, with approximately a third of the overall light coupling to neighbouring cores. Consequently, cross coupling causes severe blurring in the resulting images, whose restoration is formulated as an inverse problem. We will discuss in detail cross coupling effects in Section II. In this work, we consider a noisy observation vector $\mathbf{y}$, of an original intensity vector $\mathbf{x}$, that is modelled by the following linear forward model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \qquad (1)$$

where $\mathbf{A}$ is the matrix representing a linear operator which can model different degradation. Here, $\mathbf{A}$ models fiber core cross coupling and/or spatial blur. We specify the dimensions of the variables later in the text. In (1), the vector $\mathbf{w}$ stands for additive noise, modelling observation noise and model mismatch and is assumed to be a white Gaussian noise sequence. In wide-field OEM, the constant background fluorescence of the fiber bundle [7], [11], is significant (between 90% and

A. K. Eldaly, Y. Altmann, A. Perperidis and S. McLaughlin are with the Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK. (Emails: {AK577; Y.Altmann; A.Perperidis; S.Mclaughlin}@hw.ac.uk)

T. R. Choudhary is with the Institute of Biological Chemistry, Biophysics and Bioengineering, Heriot-Watt University, Edinburgh, United Kingdom (Email: T.Choudhary@hw.ac.uk)

N. Krstajić and K. Dhaliwal are with the EPSRC IRC Hub in Optical Molecular Sensing & Imaging, MRC Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK (Emails: {N.Krstajic; Kev.Dhaliwal}@ed.ac.uk)

60% of the total signal) providing a significant offset to all fluorescence measurements from tissue. Hence, the total noise level does not depend on the tissue signal level. Also, we consider applications where the photon flux is high (> 500 photoelectrons generated per pixel per typical exposure time 50 ms). Therefore, the Gaussian noise assumption holds [12]–[14].

The problem of estimating $\mathbf{x}$ from $\mathbf{y}$ is an ill-posed linear inverse problem (LIP); i.e., the matrix $\mathbf{A}$ is singular or very ill-conditioned. Consequently, this problem requires additional regularization (or prior information, in Bayesian inference terms) in order to reduce uncertainties and improve estimation performance. State-of-the-art algorithms for solving such problems can be split into either convex optimization or Bayesian methods.

In [15]–[18], the problem of estimating $\mathbf{x}$ given $\mathbf{y}$ is formulated as an unconstrained optimization problem as follows

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}) + i_{\mathbb{R}+}(\mathbf{x}), \qquad (2)$$

where $\phi(\cdot)$ is a regularization function, $\|.\|_2$ is the standard ℓ_2-norm, $\lambda \in \mathbb{R}_+$ is a regularization parameter, and $i_{\mathbb{R}+}(\mathbf{x})$ is the indicator function defined on the positive set of $\mathbf{x}$. For solving problems of the form (2), state-of-the-art algorithms potentially belonging to the iterative shrinkage/thresholding family [15]–[18] can be used. In [16], [19], the unconstrained problem in Eq.(2) is solved by an algorithm called split augmented Lagrangian shrinkage algorithm (SALSA) which is based on variable splitting [20], [21].

Alternatively, many studies have considered hierarchical Bayesian models to solve the deconvolution and restoration problem [22]–[31]. These models offer a flexible and consistent methodology to deal with uncertainty in inference when limited amount of data or information is available. Moreover, other unknown parameters can be jointly estimated within the algorithm such as noise variance(s) and regularization parameters. As such, they represent an attractive way to tackle ill-posed problems such as the one considered in this work. These methods rely on selecting an appropriate prior distribution for the unknown image and other unknown parameters. The full posterior distribution can then be derived from the Bayes' rule, and then exploited by optimization or simulation-based (Markov chain Monte Carlo) methods.

The main contributions of this work are fourfold:

1) We address the problem of deconvolution and restoration in OEM. To the best of our knowledge, it is the first time this problem is addressed in a statistical framework by using a hierarchical Bayesian model.

2) We develop algorithms dedicated to irregularly sampled images which do not rely on strong assumptions about the spatial structure of the sampling patterns. The developed methods can thus be applied to a wide range of imaging systems, and fiber bundle designs.

3) We derive three estimation algorithms associated with the proposed hierarchical Bayesian model and compare them using extensive simulations conducted using controlled and real data. The first algorithm generates samples distributed according to the posterior distribution using Markov chain Monte Carlo (MCMC) methods [32]. This approach also allows the estimation of the hyperparameters associated with the priors. However, as mentioned previously, the resulting MCMC-based algorithm presents a high computational complexity. The second and third algorithms deal with this limitation and approximate the joint posterior distribution. The second algorithm uses the variational Bayes (VB) methodology [33], [34] to approximate the joint posterior distribution by minimizing the KullbackLeibler (KL) divergence between the true posterior distribution and its approximation [35]. It can also estimate the hyperparameters associated with the prior distributions, and hence it is totally unsupervised, as is the MCMC-based method. The third algorithm is based on the alternating direction method of multipliers (ADMM). Although the low computation complexity of this algorithm, the hyperparameters associated with the priors need to be chosen carefully by the user, and hence it is considered as a semi-supervised method.

4) We use Gaussian Processes (GP) to interpolate the resulting samples to provide a meaningful image and quantify uncertainties at each interpolated sample.

The remaining sections of the paper are organized as follows. Section II discusses the cross coupling problem and formulates the problem of deconvolution and restoration of OEM data. The proposed hierarchical Bayesian model is then presented in Section III. Section IV introduces the three proposed estimation algorithms based on MCMC and optimization. Results of simulations conducted using synthetic and real datasets are discussed in Section VI and Section VII, respectively. Conclusions and future work are finally reported in Section VIII.

II. PROBLEM FORMULATION

Fig. 1 illustrates what happens in the fibre bundle when receiving fluorescent light from an object being imaged. The vectors $\mathbf{x}_o$, $\mathbf{x}$, and $\mathbf{g}$ represent light intensities at the object being imaged (tissue in this case), at the distal end of the fibre bundle, and at the image plane respectively. The transform $\mathbf{H}$ represents the cross coupling effect defined later in the text, $\mathbf{C}$ represents the spatial blur acting between the proximal end of the fibre bundle and the image plane, whereas $\mathbf{C}'$ is that between the distal end of the fibre bundle and the tissue being imaged. The two spatial blurs $\mathbf{C}$ and $\mathbf{C}'$ are spatially variant, $\mathbf{C}$ can be characterized as the distance d between the image plane and the proximal end of the fibre is known, whereas $\mathbf{C}'$ cannot be fully characterized as d' is unknown and the frames here are analyzed independently. Hence, to overcome this problem, we aim to recover the intensity vector $\mathbf{x}$ rather than $\mathbf{x}_o$.

Fig. 2 provides and illustrative example of cross coupling between fiber cores. If an individual fiber core is illuminated in $\mathbf{x}$, the neighbouring cores in $\mathbf{g}$ will be affected by a specific percentage of the incident light on the illuminated core. Experimental results in current fiber bundle (which might be different for other bundles) showed that around 61% of the light transmitted through a single core remains in that core, around 34% migrates to the immediate neighbouring cores,
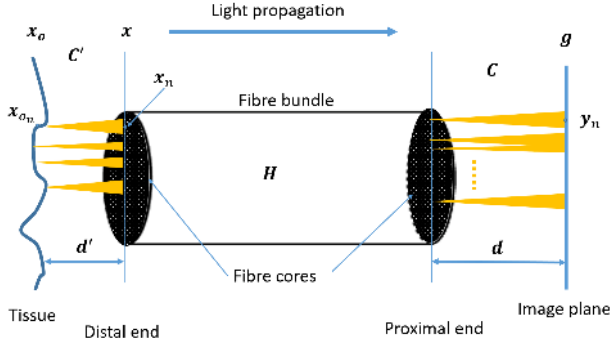
Fig. 1: Schematic diagram showing the forward model in OEM.

around 4% to the second order neighbours and less than 1% to the third, fourth, and fifth order neighbours [10].
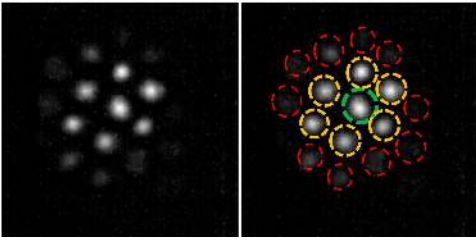
Fig. 2: Example of cross coupling between fiber cores, the green circle represents the central illuminated core and the yellow and red ones represent the immediate and further neighbours respectively.

Fig. 3 illustrates how we construct the forward observation model to mimic the same output as the endomicroscopy imaging system. The first image on the left-hand side of the figure represents the illumination of one fiber core. This results in cross coupling to the neighbouring cores (convolution with a first linear operator $\mathbf{H}$), then the spatial blurring effect around each fiber core (convolution with a second linear operator $\mathbf{C}$) and finally the fourth image of the figure shows the final system output after adding white Gaussian noise.
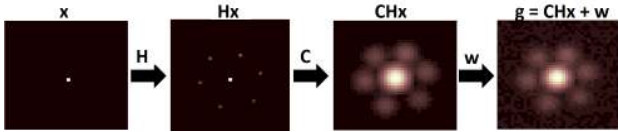
Fig. 3: Representation of the endomicroscopy system output images.

The linear model in (1) can now be written as

$$\mathbf{g} = \mathbf{CHx} + \mathbf{w}, \qquad (3)$$

where $\mathbf{A}$ in (1) is replaced by $\mathbf{CH}$ in (3), the vector $\mathbf{g}$ is the observed data matrix, and $\mathbf{x}$ is the image to be restored.

From preliminary results, we propose to model cross-coupling by an isotropic zero mean 2D generalized Gaussian kernel applied to the fiber intensities [10] as follows

$$[\mathbf{H}]_{i,j} = \exp\left(-\left(\frac{d_{i,j}}{\alpha_{\mathbf{H}}}\right)^{\beta_{\mathbf{H}}}\right), \qquad (4)$$

where $d_{i,j}$ denotes the euclidean distance between the cores (or spatial locations) i and j, which corresponds to approximately
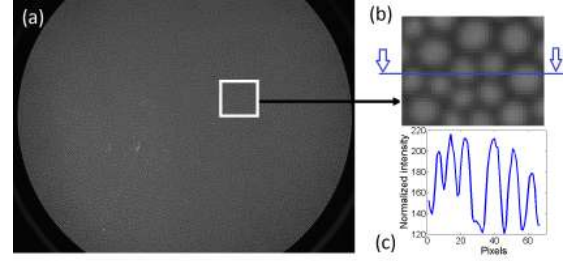
Fig. 4: (a) A background image, (b) a zoomed part of the image, and (c) the intensity profile across one line in the image.

3.3 pixels between neighbouring cores. From (4), it can be seen that neighbouring fiber cores will be more closely coupled than distant ones. The values of $\alpha_{\mathbf{H}}$ and $\beta_{\mathbf{H}}$, which control the amount of cross-coupling (the higher, the more coupling) and which are system dependent, are adjusted from preliminary measurements (calibration). Note that other cross-coupling models could also be considered instead of (4) depending on the imaging system used.

The spatial blur affecting each fiber core can be modelled by a Gaussian spatial filter, as illustrated in Fig. 4, which shows a background image i.e., an image from a sample presenting constant intensity, using an endomicroscopy imaging system, and a zoomed-in region of this image, bright and dark areas represent fiber cores and their cladding, respectively. The intensity profile across one line in this image is a series of Gaussian kernels. However, the variation of the shape and width of the kernels is due to the variation in core sizes.

Due to the variation in core sizes, the blurring kernel $\mathbf{C}$ varies accordingly, and hence the cores tend to overlap. So the complete model in (3) becomes more complex, and potentially computationally expensive for long image sequences (videos). Indeed, there is no structure in $\mathbf{C}$ which allows us to compute $\mathbf{CHx}$ rapidly. Hence we propose a simplification of this model and represent each core by a single intensity value. The mean intensities of fibre core pixels could be used, but the overlap between the cores makes its computation difficult. Since the variation of the width of this blur is not too significant, the maximum intensity of each core is considered instead ($\mathbf{y}_n$ in Fig. 1).

Following the above mentioned points, the model in (3) can be simplified to

$$\mathbf{y} = \mathbf{Hx} + \mathbf{w}. \qquad (5)$$

Assume that N is the total number of pixels in the image, and N_1 representing number of fibre cores in the image, the input $\mathbf{y} \approx \mathbf{C}^+\mathbf{g} \in \mathbb{R}^{N_1}$, where $\mathbf{C}^+$ is the pseudo-inverse of $\mathbf{C}$, and the output $\mathbf{x} \in \mathbb{R}^{N_1}$ are two vectors representing central core intensities, where, $N_1 << N$, and $\mathbf{H} \in \mathbb{R}^{N_1 \times N_1}$. The noise $\mathbf{w} \in \mathbb{R}^{N_1}$ is assumed to be additive white noise which is independent and identically distributed (i.i.d) zero mean Gaussian noise with variance σ^2, denoted as $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, where $\sim$ means "is distributed according to" and $\mathbf{I}$ is the identity matrix.

The problem investigated in this paper is to estimate the actual intensity values $\mathbf{x}$, and the noise variance σ^2 from the observation vector $\mathbf{y}$. As mentioned previously, to solve this

problem, we propose a hierarchical Bayesian model and a set of different estimation methods to estimate the unknown parameters.

III. HIERARCHICAL BAYESIAN MODEL

This section introduces a hierarchical Bayesian model proposed to estimate the unknown parameter vector $\mathbf{x}$ and σ^2. This model is based on the likelihood function of the observations and on prior distributions assigned to the unknown parameters.

A. Likelihood

Eq. (5) yields that $\mathbf{y}|(\mathbf{x}, \sigma^2) \sim \mathcal{N}(\mathbf{Hx}, \sigma^2\mathbf{I})$. Consequently, the likelihood can be expressed as

$$f(\mathbf{y}|\mathbf{x}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{N_1/2} \exp\left(-\frac{\|\mathbf{y} - \mathbf{Hx}\|_2^2}{2\sigma^2}\right). \quad (6)$$

B. Parameter Priors

1) Prior for the underlying intensity field $\mathbf{x}$: A truncated multivariate Gaussian distribution (MVG) is assigned to the intensity field $\mathbf{x}$.

$$f(\mathbf{x}|\gamma^2) \propto \left(\gamma^2\right)^{-d/2} \exp\left(-\frac{\mathbf{x}^T\boldsymbol{\Delta}^{-1}\mathbf{x}}{2\gamma^2}\right) 1_{\mathbb{R}^+}(\mathbf{x}), \quad (7)$$

where $1_{\mathbb{R}^+}(\mathbf{x})$ is the indicator function defined on the positive set of $\mathbf{x}$, γ^2 controls the global correlation between intensities, and the covariance matrix $\boldsymbol{\Delta}$ which defines the spatial correlation between the cores is defined by

$$[\boldsymbol{\Delta}]_{n,n'} = \exp\left(-\left(\frac{d_{n,n'}}{\ell}\right)^\kappa\right), \quad (8)$$

where $d_{n,n'}$ denotes the distance between the spatial locations n and n', and $d = N_1$. Equations (7) and (8) promote smooth intensity variations between neighbours while ensuring that the prior dependence between neighbouring cores decrease as $d_{n,n'}$ increases. In this work $d_{n,n'}$ is the standard euclidean distance. The parameters ℓ, κ were learned from the irregular sampling pattern of the OEM system. Precisely, we used known images and selected (ℓ, κ) by maximum likelihood estimation, which occurs when $p(\ell, \kappa|\mathbf{x})$ is at its greatest, which corresponds to maximizing $\log p(\ell, \kappa|\mathbf{x})$. While γ^2 is left unknown for each image, (ℓ, κ) are fixed in the rest of the simulations as the average values obtained with the training images.

Considering such a prior is equivalent to assuming a Gaussian process on $\mathbf{x}$, this allows us to interpolate the resulting deconvolved intensities using Gaussian processes [36] as we will see in section V.

2) Prior for the noise variance σ^2: A conjugate inverse-Gamma $\mathcal{IG}$ prior is assigned to the noise variance σ^2

$$f(\sigma^2|\alpha, \beta) \sim \mathcal{IG}(\alpha, \beta), \quad (9)$$

where $\alpha = 10$ is fixed arbitrarily, while the hyperparameter β is estimated within the algorithm.

3) Prior for the hyperparameter β: The hyperparameter associated with the parameter prior defined above is assigned to a conjugate Gamma distribution:

$$\beta \sim \mathcal{G}(\alpha_o, \beta_o), \quad (10)$$

where α_o and β_o are fixed and user-defined parameters which might depend on the quality of the data to be recovered. In this work, we fixed $(\alpha_o, \beta_o) = (10, 0.1)$ arbitrarily.

4) Prior for the hyperparameter γ^2: To reflect the lack of prior knowledge about the regularization parameter γ^2 in (7), the following weakly informative conjugate inverse-Gamma prior is assigned to it.

$$\gamma^2 \sim \mathcal{IG}(\eta, \nu), \quad (11)$$

where (η, ν) are fixed to $(\eta, \nu) = (10^{-3}, 10^{-3})$. Note that we did not observe significance change in the results when changing these hyperparameters.

The next section derives the joint posterior distribution of the unknown parameters associated with the proposed Bayesian model.

C. Joint posterior distribution

Assuming the parameters $\mathbf{x}$ and σ^2 are *a priori* independent, the joint posterior distribution of the parameter vector $\boldsymbol{\Omega} = \{\mathbf{x}, \sigma^2\}$ and hyperparameters $\boldsymbol{\phi} = \{\beta, \gamma^2\}$ can be expressed as

$$f(\boldsymbol{\Omega}, \boldsymbol{\phi}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\Omega})f(\boldsymbol{\Omega}|\boldsymbol{\phi})f(\boldsymbol{\phi}), \quad (12)$$

where

$$f(\boldsymbol{\Omega}|\boldsymbol{\phi}) = f(\mathbf{x}|\gamma^2)f(\sigma^2|\beta), \text{ and } f(\boldsymbol{\phi}) = f(\gamma^2)f(\beta). \quad (13)$$

The directed acyclic graph (DAG) summarizing the structure of proposed Bayesian model is depicted in Fig. 5. This posterior distribution will be used to evaluate Bayesian estimators of $\boldsymbol{\Theta} = \{\boldsymbol{\Omega}, \boldsymbol{\phi}\}$. For this purpose, we propose three algorithms: an MCMC-based approach and two optimization-based approaches, in which VB and ADMM are considered. The first approach uses an MCMC method to evaluate the minimum-mean-square-error (MMSE) estimator of $\boldsymbol{\Theta}$ by generating samples according to the joint posterior distribution. Moreover, it allows the estimation of the hyperparameter vector $\boldsymbol{\phi}$ along with the noise variance σ^2. However, it exhibits a relatively long computational time. The second and third algorithms which deal with this issue and provide fast MMSE estimate for the VB approach and MAP estimate for the ADMM approach. The VB approach approximates the joint posterior distribution in (12) by minimizing the Kullback-Leibler (KL) divergence between the true posterior distribution and its approximation [35]. The ADMM approach is achieved by maximizing the posterior distribution (12) with respect to (w.r.t.) $\boldsymbol{\Theta}$. Note however, that the hyperparameters $\boldsymbol{\phi}$ as well as σ^2 are fixed for this approach. The three estimation algorithms are described in the next section.
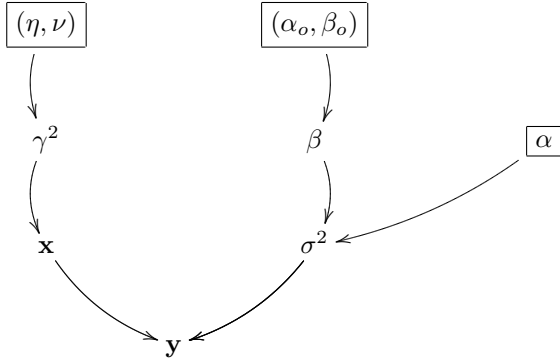
Fig. 5: Graphical model for the proposed hierarchical Bayesian model (fixed quantities appear in boxes).

IV. BAYESIAN INFERENCE

A. MCMC algorithm

To overcome the challenging derivation of Bayesian estimators associated with $f(\boldsymbol{\Theta}|\mathbf{y})$, we propose to use an efficient MCMC method to generate samples asymptotically distributed according to the posterior presented in (12). More precisely, we consider a Gibbs sampler described next. The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest [[32], Chap. 10]. In this work, we propose to sample sequentially the elements of $\boldsymbol{\Theta}$ using updates that are detailed below.

1) Sampling the intensity field $\mathbf{x}$*:* From (12), since the prior (7) is conjugate to the Gaussian distribution, the full conditional distribution of $\mathbf{x}$ is given by

$$f(\mathbf{x}|\mathbf{y}, \sigma^2) \sim \mathcal{N}_{\mathbb{R}^+}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad (14)$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \sigma^{-2}\boldsymbol{\Sigma}^T\mathbf{H}^T\mathbf{y}, \\ \boldsymbol{\Sigma} &= \left(\sigma^{-2}\mathbf{H}^T\mathbf{H} + \gamma^{-2}\boldsymbol{\Delta}^{-1}\right)^{-1}. \end{aligned} \qquad (15)$$

Sampling from (14) can be achieved efficiently by using the Hamiltonian method proposed in [37].

2) Sampling the noise variance σ^2*:* By cancelling out the terms that don't depend on σ^2 from the posterior distribution in (12), its conditional distribution can be written as

$$f(\sigma^2|\mathbf{y}, \mathbf{x}) \sim \mathcal{IG}\left(\alpha + \frac{N_1}{2}, \beta + \frac{\|\mathbf{y} - \mathbf{Hx}\|_2^2}{2}\right), \qquad (16)$$

which is easy to sample from.

3) Sampling the hyperparameters β *and* γ^2*:* It can be easily shown that β can be sampled from the following Gamma distribution

$$f(\beta|\sigma^2) \sim \mathcal{G}\left(\alpha + \alpha_o, \frac{\sigma^2\beta_o}{\sigma^2 + \beta_o}\right). \qquad (17)$$

In a similar fashion to the noise variance, γ^2 can be sampled from the following inverse-Gamma distribution

$$f(\gamma^2|\mathbf{x}) \sim \mathcal{IG}\left(\eta + \frac{N_1}{2}, \nu + \frac{\mathbf{x}^T\boldsymbol{\Delta}^{-1}\mathbf{x}}{2}\right). \qquad (18)$$

The algorithm for generating samples asymptotically distributed according to the posterior distribution using Gibbs sampler is shown in Algorithm 1.

Algorithm 1 Deconvolution via MCMC: Gibbs Sampling Algorithm

1: **Fixed input parameters**: Number of burn-in iterations N_{bi}, total number of iterations N_{MC}
2: **Initializations** ($k = 0$)
  • Set $\mathbf{x}^{(0)}, \sigma^{2(0)}, \beta^{(0)}, \gamma^{2(0)}$
3: **Repeat** ($1 \leq k \leq N_{\mathbf{MC}}$)
  • Sample $\mathbf{x}^{(k)}$ from (14)
  • Sample $\sigma^{2(k)}$ from (16)
  • Sample $\beta^{(k)}$ from (17)
  • Sample $\gamma^{2(k)}$ from (18)
4: **Set** $k = k + 1$.

The posterior distribution mean or minimum mean square error (MMSE) estimator of $\mathbf{x}$ can be approximated by

$$\hat{\mathbf{x}} = \frac{1}{N_{\mathrm{MC}} - N_{\mathrm{bi}}} \sum_{t=N_{\mathrm{bi}+1}}^{N_{\mathrm{MC}}} \mathbf{x}^{(t)}, \qquad (19)$$

where the samples from the first N_{bi} iterations (corresponding to the transient regime or burn-in period, which is determined visually from preliminary runs) of the sampler are discarded.

B. Variational Bayes algorithm

For this approach, we consider an approximation of $p(\boldsymbol{\Theta}|\mathbf{y})$ by a simpler tractable distribution $q(\boldsymbol{\Theta})$ following the variational methodology [34], moreover, here, we relax the positivity constraints about the intensity field vector $\mathbf{x}$. Note, however that the positivity constraints can be incorporated but the covariance matrix of the intensity field $\mathbf{x}$ would become more complicated [38], chap. 5. As will be shown in Sections VI and VII, this constraint relaxation yields a fast estimation procedure providing estimation results which compete with the methods incorporating this constraint. The distribution $q(\boldsymbol{\Theta})$ will be found by minimizing the Kullback-Leibler (KL) divergence, between the actual posterior distribution and its approximation, given by [35] [39]

$$\mathbf{D}_{\mathrm{KL}}\left(q(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\mathbf{y})\right) = \int q(\boldsymbol{\Theta}) \log\left(\frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\mathbf{y})}\right) d\boldsymbol{\Theta}, \quad (20)$$

which is always non-negative and equal to zero only when $q(\boldsymbol{\Theta}) = p(\boldsymbol{\Theta}|\mathbf{y})$. In order to obtain a tractable approximation, the family of distributions $q(\boldsymbol{\Theta})$ are restricted utilizing the mean field approximation [40] so that $q(\boldsymbol{\Theta}) = q(\phi)q(\mathbf{x})q(\sigma^2)$, where $q(\phi) = q(\gamma^2)q(\beta)$.

The lower bound of the KL divergence is given by

$$p(\boldsymbol{\Theta}, \mathbf{y}) \geq p(\mathbf{y}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\phi)p(\phi) = F(\boldsymbol{\Theta}, \mathbf{y}). \qquad (21)$$

For $\mathcal{H} \in \{x, \sigma^2, \gamma^2, \beta\}$, let us denote by $\boldsymbol{\Theta}_{\backslash\mathcal{H}}$, the subset of $\boldsymbol{\Theta}$ with $\mathcal{H}$ removed; for instance, if $\mathcal{H} = \mathbf{x}$, $\boldsymbol{\Theta}_{\backslash\mathbf{x}} = \{\sigma^2, \gamma^2, \beta\}$. Then utilizing the lower bound $\mathbf{F}(\boldsymbol{\Theta}, \mathbf{y})$ for the joint probability distribution in (20) we obtain an upper bound for the KL divergence as follows

$$\mathcal{M}(q(\mathbf{\Theta})) = \int q(\mathbf{\Theta}) \log \left(\frac{q(\mathbf{\Theta})}{p(\mathbf{\Theta}|\mathbf{y})} \right) d\mathbf{\Theta}$$
$$\leq \int q(\mathcal{H}) \left(\int q(\mathbf{\Theta}_{\backslash\mathcal{H}}) \log \left(\frac{q(\mathcal{H})q(\mathbf{\Theta}_{\backslash\mathcal{H}})}{F(\mathbf{\Theta},\mathbf{y})} \right) d\mathbf{\Theta}_{\backslash\mathcal{H}} \right) d\mathcal{H}$$
$$= \mathcal{M}(q(\mathcal{H})). \tag{22}$$

Therefore, we minimize this upper bound instead of minimizing the KL divergence in (20). Note that the form of the inequality in (22) suggests an alternating (cyclic) optimization strategy where the algorithm cycles through the unknown distributions and replaces each variable with a revised estimate given by the minimum of (22) with the other distributions held constant. Thus, given $q(\mathbf{\Theta}_{\backslash\mathcal{H}})$, the posterior distribution approximation $q(\mathcal{H})$ can be computed by solving

$$\hat{q}(\mathcal{H}) = \underset{q(\mathcal{H})}{\text{minimize}} \quad \mathbf{D}_{\text{KL}}\left(q(\mathbf{\Theta}_{\backslash\mathcal{H}})q(\mathcal{H})||F(\mathbf{\Theta},\mathbf{y}) \right). \tag{23}$$

In order to solve this equation, we note that differentiating the integral on the right hand side in (22) w.r.t. $q(\mathcal{H})$ results in (see [41], Eq. (2.28))

$$\hat{q}(\mathcal{H}) = const \times \exp\left(E_{q(\mathbf{\Theta}_{\backslash\mathcal{H}})}[\log F(\mathbf{\Theta},\mathbf{y})] \right), \tag{24}$$

where

$$Eq(\mathbf{\Theta}_{\backslash\mathcal{H}})[\log F(\mathbf{\Theta},\mathbf{y})] = \int \log F(\mathbf{\Theta},\mathbf{y})q(\mathbf{\Theta}_{\backslash\mathcal{H}})d\mathbf{\Theta}_{\backslash\mathcal{H}}. \tag{25}$$

We obtain the following iterative procedure to find $q(\mathbf{\Theta})$ by applying this minimization to each unknown in an alternating way

---

Algorithm 2 VB algorithm

1: **Set** $k = 1$, **choose** $q^1(\sigma^2), q^1(\beta)$ and $q^1(\gamma^2)$, initial estimates of the distributions $q(\sigma^2), q(\beta)$ and $q(\gamma^2)$,
2: **repeat** (k = k + 1)
3: $q^k(\mathbf{x}) = \underset{q(\mathbf{x})}{\text{minimize}} \int\int q^k(\mathbf{\Theta}_{\backslash\mathbf{x}})q(\mathbf{x}) \times$
$\log \left(\frac{q^k(\mathbf{\Theta}_{\backslash\mathbf{x}})q(\mathbf{x})}{F(\mathbf{\Theta}^k_{\backslash\mathbf{x}},\mathbf{x},\mathbf{y})} \right) d\mathbf{\Theta}_{\backslash\mathbf{x}}d\mathbf{x}$
4: $q^k(\sigma^2) = \underset{q(\sigma^2)}{\text{minimize}} \int\int q^k(\mathbf{\Theta}_{\backslash\sigma^2})q(\sigma^2) \times$
$\log \left(\frac{q^k(\mathbf{\Theta}_{\backslash\sigma^2})q(\sigma^2)}{F(\mathbf{\Theta}^k_{\backslash\sigma^2},\mathbf{x},\mathbf{y})} \right) d\mathbf{\Theta}_{\backslash\sigma^2}d\sigma^2$
5: $q^k(\gamma^2) = \underset{q(\gamma^2)}{\text{minimize}} \int\int q^k(\mathbf{\Theta}_{\backslash\gamma^2})q(\gamma^2) \times$
$\log \left(\frac{q^k(\mathbf{\Theta}_{\backslash\gamma^2})q(\gamma^2)}{F(\mathbf{\Theta}^k_{\backslash\gamma^2},\mathbf{x},\mathbf{y})} \right) d\mathbf{\Theta}_{\backslash\gamma^2}d\gamma^2$
6: $q^k(\beta) = \underset{q(\beta)}{\text{minimize}} \int\int q^k(\mathbf{\Theta}_{\backslash\beta})q(\beta) \times$
$\log \left(\frac{q^k(\mathbf{\Theta}_{\backslash\beta})q(\beta)}{F(\mathbf{\Theta}^k_{\backslash\beta},\beta,\mathbf{y})} \right) d\mathbf{\Theta}_{\backslash\beta}d\beta$
7: **until** some stopping criterion is satisfied.

Now we detail the solutions at each step of algorithm (2) explicitly.

1) Updating intensity field vector $\mathbf{x}$: From (24), it can be shown that $q^k(\mathbf{x})$ is an N_1-dimensional Gaussian distribution, rewritten as

$$q^k(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}; E_{q^k(\mathbf{x})}(\mathbf{x}), \mathbf{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x}) \right), \tag{26}$$

where the mean $E_{q^k(\mathbf{x})}(\mathbf{x})$ and covariance $\mathbf{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x})$ of this normal distribution can be calculated from step 3 in Algorithm 2 as

$$E_{q^k(\mathbf{x})}(\mathbf{x}) = \frac{(\mathbf{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x}))^T\mathbf{H}^T\mathbf{y}}{E_{q^k(\sigma^2)}(\sigma^2)}, \tag{27a}$$

$$\mathbf{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x}) = \left(\frac{\mathbf{H}^T\mathbf{H}}{E_{q^k(\sigma^2)}(\sigma^2)} + \frac{\mathbf{\Delta}^{-1}}{E_{q^k(\gamma^2)}(\gamma^2)} \right)^{-1}. \tag{27b}$$

2) Updating noise variance σ^2: It is easy to show from (24) that the noise variance follows an inverse-Gamma distribution given by

$$q^k(\sigma^2) = \mathcal{IG}\left(\sigma^2; \frac{N_1}{2} + \alpha, E_{q^k(\beta)}(\beta) + E_{q^k(\mathbf{x})}\left[\|\mathbf{y} - \mathbf{Hx}\|_2^2 \right] \right), \tag{28}$$

whose mean is given by

$$E_{q^k(\sigma^2)}(\sigma^2) = \frac{E_{q^k(\beta)}(\beta) + E_{q^k(\mathbf{x})}\left[\|\mathbf{y} - \mathbf{Hx}\|_2^2 \right]}{N_1/2 + \alpha - 1}, \tag{29}$$

where

$$E_{q^k(\mathbf{x})}\left[\|\mathbf{y} - \mathbf{Hx}\|_2^2 \right] = \|\mathbf{y} - \mathbf{H}E_{q^k(\mathbf{x})}(\mathbf{x})\|_2^2$$
$$+ \text{tr}\left(\mathbf{H}^T\mathbf{H}\mathbf{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x}) \right). \tag{30}$$

where tr(.) denotes the trace of the matrix.

3) Updating regularization parameter γ^2: In a similar fashion to noise variance, the regularization parameter γ^2 follows an inverse-Gamma distribution given by

$$q^k(\gamma^2) = \mathcal{IG}\left(\gamma^2; \frac{N_1}{2} + \eta, \nu + \frac{1}{2}E_{q^k(\mathbf{x})}\left[\mathbf{x}^T\mathbf{\Delta}^{-1}\mathbf{x} \right] \right), \tag{31}$$

whose mean is given by

$$E_{q^k(\gamma^2)}(\gamma^2) = \frac{\nu + \frac{1}{2}E_{q^k(\mathbf{x})}\left[\mathbf{x}^T\mathbf{\Delta}^{-1}\mathbf{x} \right]}{N_1/2 + \eta - 1} \tag{32}$$

where

$$E_{q^k(\mathbf{x})}\left[\mathbf{x}^T\mathbf{\Delta}^{-1}\mathbf{x} \right] = E_{q^k(\mathbf{x})}(\mathbf{x}^T)\mathbf{\Delta}^{-1}E_{q^k(\mathbf{x})}(\mathbf{x})$$
$$+ \text{tr}\left(\mathbf{\Delta}^{-1}\mathbf{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x}) \right). \tag{33}$$

4) Updating the hyperparameter β: The hyperparameter β follows a Gamma distribution given by

$$q^k(\beta) = \mathcal{G}\left(\beta; \alpha + \alpha_o, \frac{\beta_o E_{q^k(\sigma^2)}(\sigma^2)}{\beta_o + E_{q^k(\sigma^2)}(\sigma^2)} \right), \tag{34}$$

whose mean is given by

$$E_{q^k(\beta)}(\beta) = \frac{(\alpha + \alpha_o)\beta_o E_{q^k(\sigma^2)}(\sigma^2)}{\beta_o + E_{q^k(\sigma^2)}(\sigma^2)}. \tag{35}$$

In Algorithm 2, no assumptions were imposed on the posterior approximation of $q(\mathbf{x})$. We can, however, assume as [28]–[31], [42], that this distribution is degenerate, i.e., distribution which takes one value with probability one and the rest of the values with probability zero. We can obtain another algorithm under this assumption which is similar to algorithm 2.

Algorithm 3 Deconvolution via VB

1: **Set** $k = 1$,

2: **Initialize** $E_{q^1(\sigma^2)}(\sigma^2), E_{q^1(\gamma^2)}(\gamma^2)$ and $E_{q^1(\beta)}(\beta)$,

3: **repeat** (k = k + 1)

4: $E_{q^k(\mathbf{x})}(\mathbf{x}) = \left(\frac{\mathbf{H}^T\mathbf{H}}{E_{q^k(\sigma^2)}(\sigma^2)} + \frac{\mathbf{\Delta}^{-1}}{E_{q^k(\gamma^2)}(\gamma^2)} \right)^{-1} \frac{\mathbf{H}^T\mathbf{y}}{E_{q^k(\sigma^2)}(\sigma^2)}$

5: $E_{q^k(\sigma^2)}(\sigma^2) = \frac{E_{q^k(\beta)}(\beta) + \|\mathbf{y} - \mathbf{H}E_{q^k(\mathbf{x})}(\mathbf{x})\|_2^2}{N_1/2 + \alpha - 1}$

6: $E_{q^k(\gamma^2)}(\gamma^2) = \frac{\nu + \frac{1}{2}\left(E_{q^k(\mathbf{x})}(\mathbf{x})\right)^T \mathbf{\Delta}^{-1} E_{q^k(\mathbf{x})}(\mathbf{x})}{N_1/2 + \eta - 1}$

7: $E_{q^k(\beta)}(\beta) = \frac{(\alpha + \alpha_o)\beta_o E_{q^{k+1}(\sigma^2)}(\sigma^2)}{\beta_o + E_{q^{k+1}(\sigma^2)}(\sigma^2)}$

8: **until** some stopping criterion is satisfied.

9: **Set** $\hat{\mathbf{x}} = E_{q^k(\mathbf{x})}(\mathbf{x}), \hat{\sigma^2} = E_{q^k(\sigma^2)}(\sigma^2), \hat{\gamma^2} = E_{q^k(\gamma^2)}(\gamma^2)$, and $\hat{\beta} = E_{q^k(\beta)}(\beta)$

The stopping criterion we use is $\sum_{\mathcal{H} \in \{\mathbf{x}, \sigma^2, \beta, \gamma^2\}} \|\mathcal{H}^{(k)} - \mathcal{H}^{(k+1)}\|_F \leq \epsilon$, where $\epsilon = \sqrt{N_1} \times 10^{-5}$ [43].

It is clear that using degenerate distribution for $q(\mathbf{x})$ in Algorithm 3 removes the uncertainty terms of the intensity field estimate. It has been shown that this helps to improve the restoration performance [28]–[31], [42]. Moreover, it also reduces the computational complexity as there is no need to compute explicitly the covariance matrix $\mathbf{\Sigma}_{q^k(\mathbf{x})}(\mathbf{x})$ at each iteration. Finally, a few remarks are needed to obtain a fast algorithm. The inverse of the covariance matrix $\mathbf{\Delta}$ needs to be computed only once before the loop in Algorithm 3. We also considered the MATLAB operation $\left(\frac{\mathbf{H}^T\mathbf{H}}{E^k(\sigma^2)} + \frac{\mathbf{\Delta}^{-1}}{E^k(\gamma^2)} \right) \backslash (\mathbf{H}^T\mathbf{y})$ for the update of the intensity field vector $\mathbf{x}$, which is faster than computing the covariance matrix in (27b), then updating the mean in (27a). For very big images, diagonal approximation [29] or conjugate gradient [44] can be considered for the update of the intensity field vector $\mathbf{x}$.

C. ADMM algorithm

This section describes another alternative to the MCMC algorithm which is based on an optimization algorithm. The latter maximizes the joint posterior distribution (12) $f(\mathbf{\Omega}|\mathbf{y}, \boldsymbol{\phi})$ with respect to (w.r.t.) the parameters of interest, with fixing the hyperparameter vector $\boldsymbol{\phi}$, to approximate the MAP estimator of $\mathbf{\Theta}$, or equivalently, by minimizing the negative log-posterior distribution given by $\mathcal{F} = -\log[f(\mathbf{\Theta}|\mathbf{y})]$. The resulting optimization problem is tackled using ADMM that sequentially updates the different parameters, which is widely used in the literature for solving imaging inverse problems [19], [43], [45]. We rewrite the model as an optimization problem as follows

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}) + i_{\mathbb{R}+}(\mathbf{x}), \quad (36)$$

where the regularization function $\phi(\mathbf{x})$ is proportional to the negative logarithm of the intensity field prior considered in (7) up to an additive constant, i.e. $\phi(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{\Delta}^{-1}\mathbf{x}}{2}$, and $\lambda = \sigma^2/\gamma^2$ is the regularization parameter. Given this objective function, we write the constrained equivalent formulation as follows

$$\underset{\mathbf{u}, \mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}) + i_{\mathbb{R}+}(\mathbf{u}),$$
$$\text{subject to} \qquad \mathbf{u} = \mathbf{x}, \quad (37)$$

where $\mathbf{u}$ and $\mathbf{x}$ are the variables to minimize. In order to solve for $\mathbf{u}$ and $\mathbf{x}$, we construct the augmented Lagrangian corresponding to (37) as follows

$$\mathcal{L}(\mathbf{u}, \mathbf{x}, \mathbf{d}_1) = \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}) + i_{\mathbb{R}+}(\mathbf{u})$$
$$+ \frac{\mu}{2}\|\mathbf{x} - \mathbf{u} - \mathbf{d}_1\|_2^2, \quad (38)$$

where $\mu > 0$ is a positive parameter. The ADMM algorithm for solving (38) is shown in Algorithm (4). During each step of the iterative algorithm, $\mathcal{L}$ is optimized w.r.t. $\mathbf{u}$ (step 3) and $\mathbf{x}$ (step 4) and then the Lagrange multipliers are updated (step 6). The stopping criterion we use is $\|\mathbf{u}^{(k)} - \mathbf{x}^{(k)}\|_F \leq \epsilon$, where $\epsilon = \sqrt{N_1} \times 10^{-5}$ [43].

Algorithm 4 Deconvolution via ADMM

1: set $k = 0$, choose $\mu > 0, \mathbf{u}^{(0)}, \mathbf{x}^{(0)},$ and $\mathbf{d}_1^{(0)}$

2: **repeat** (k = k + 1)

3: $\mathbf{u}^{(k+1)} = \max \left(\mathbf{x}^{(k)} - \mathbf{d}_1^{(k)}, 0 \right)$

4: $\mathbf{x}^{(k+1)} = \left(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{\Delta}^{-1} + \mu\mathbf{I} \right)^{-1} \left[\mathbf{H}^T\mathbf{y} + \mu\left(\mathbf{u} + \mathbf{d}_1^{(k)}\right) \right]$

5: **Update Lagrange multipliers:**

6: $\qquad \mathbf{d}_1^{(k+1)} = \mathbf{d}_1^{(k)} - \left(\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)}\right)$

7: **Update iteration** $k \leftarrow k + 1$

8: **until** some stopping criterion is satisfied.

V. NON-LINEAR INTERPOLATION USING GAUSSIAN PROCESS REGRESSION

In order to visually view a meaningful image from the deconvolved intensities, we consider non-linear interpolation based on Gaussian processes (GP) [36], since it can provide confidence intervals for each interpolated pixel. A classic choice consists of considering a zero-mean GP with an arbitrary covariance matrix. Here, we choose this covariance matrix to be $\mathbf{\Delta}' = \mathbf{\Delta}/\gamma^2$. Precisely, we interpolate using the prior distribution previously defined in (8). If $d_{n,n'}$ is very small, then $\mathbf{\Delta}'(n, n')$ approaches its maximum $1/\gamma^2$. If n is distant from n', we have instead $\mathbf{\Delta}'(n, n') \approx 0$, i.e. the two points are considered to be *a priori* independent. So, for example, during interpolation at new n_* location, distant cores will have negligible effect. The amount of spatial correlation

depends on the parameters ℓ, and κ, which are estimated in the way we previously mentioned in section III-B1.

If we consider $\mathbf{\Delta}'(\mathbf{z}, \mathbf{z}) \in \mathbb{R}^{N_1 \times N_1}$, $\mathbf{z} = [z_1, \ldots, z_{N_1}]^T$ contains all the positions of all the observed cores (whose estimated intensities are gathered into $\mathbf{x}$), and a new spatial location z_* for which we want to predict the intensity x_*, the GP can be extended as follows

$$
\begin{bmatrix} \mathbf{x} \\ x_* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{\Delta}'(\mathbf{z}, \mathbf{z}) & \mathbf{\Delta}'(\mathbf{z}, z_*) \\ \mathbf{\Delta}'(z_*, \mathbf{z}) & 1/\gamma^2 \end{bmatrix} \right), \qquad (39)
$$

where $\mathbf{\Delta}'(\mathbf{z}, z_*) = \mathbf{\Delta}'(z_*, \mathbf{z})^T \in \mathbb{R}^{N_1}$. Eq. (39) shows that the conditional distribution of each predicted intensity given the previously estimated intensities, follows a Gaussian distribution $x_*|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ whose mean and variance are given by

$$
\begin{aligned}
\boldsymbol{\mu} &= \mathbf{\Delta}'(z_*, \mathbf{z})\mathbf{\Delta}'(\mathbf{z}, \mathbf{z})^{-1}\mathbf{x}, \\
\boldsymbol{\Sigma} &= 1/\gamma^2 - \mathbf{\Delta}'(z_*, \mathbf{z})\mathbf{\Delta}'(\mathbf{z}, \mathbf{z})^{-1}\mathbf{\Delta}'(\mathbf{z}, z_*).
\end{aligned} \qquad (40)
$$

By setting $\mathbf{x} = \hat{\mathbf{x}}$, the mean in (40) is finally used to estimate each interpolated intensity, while the variance is used to provide additional information (measure of uncertainty) about the interpolated intensity values.

VI. SIMULATIONS USING SYNTHETIC DATA
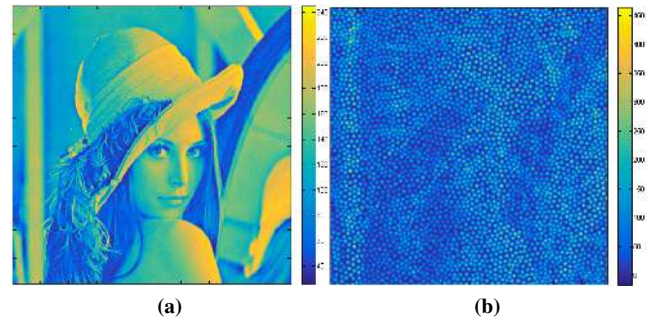
A. Data creation

The performance of the proposed methods is investigated by reconstructing a standard test image. A subsampled version of this image is obtained by considering the sampling pattern of an actual endomicroscopy system, as illustrated in Fig. 6. This figure provides an example of a homogeneous region imaged through Alveoflex (Mauna Kea Technologies, France) fiber bundle [46][47]. Such image is used for calibration and to identify the number and positions of the fiber cores. The build-in MATLAB function "vision.BlobAnalysis" was used to detect central fibre core pixels.
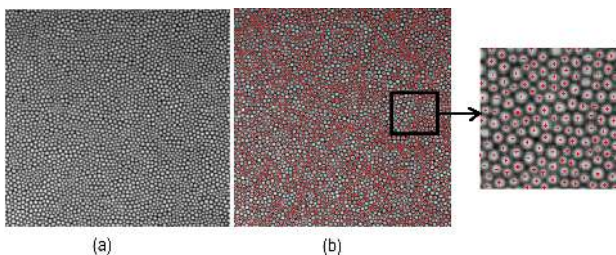
Fig. 6: (a) Example of 512×512 pixels image of the endomicroscopy system (b) Image with detected fiber core centres superimposed (red crosses).

Fig. 7 shows the original Lena image (left) and an example of system output (right) after applying the model in Eq. (3). This image is formed by creating a binary mask in which a value of 1 is assigned to pixels corresponding to the central pixels of each core in Fig. 6(b), and zero otherwise. This mask is then multiplied point by point by the Lena image in Fig. 7(a) in order to obtain the subsampled image. The model in Eq. (3) is then applied to obtain an image that simulates the system's output which is shown in Fig. 7(b). This image is

created using subsampled intensities corresponding to 1.29% of the original Lena image. For simulated data, we considered a Gaussian spatial blurring kernel with one size $\sigma_{\mathbf{C}}^2 = 2$ in all the simulations.

Fig. 7: Creation of the synthetic data: (a) Original image (b) example of final system output with $\sigma_{\mathbf{H}}^2 = 20$ and $\sigma_N^2 = 10$.

B. Performance analysis

The performance discriminator adopted in this work to measure the quality of the deconvolved fiber cores is the root mean square error (RMSE), which is computed using intensities at the core locations using

$$
\text{RMSE}(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\frac{\sum_{n=1}^{N_1} (\mathbf{x}(n) - \hat{\mathbf{x}}(n))^2}{N_1}}, \qquad (41)
$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are vectors of the subsampled reference Lena image and its deconvolved version respectively, and N_1 is the number of fibre cores.

For synthetic data, in order to check the performance of the algorithm with different cross coupling effects, different values of $\alpha_{\mathbf{H}}$ and $\beta_{\mathbf{H}}$ in (4) can be considered. However, this can be simplified by considering a 2D Gaussian kernel defined by (42)

$$
[\mathbf{H}]_{i,j} = \exp \left(\frac{-d_{i,j}^2}{2\sigma_{\mathbf{H}}^2} \right), \qquad (42)
$$

since it involves only one variable to change, namely $\sigma_{\mathbf{H}}^2$ (representing a squared distance, in pixels). This is equivalent to setting $\beta_{\mathbf{H}} = 2$ and $\alpha_{\mathbf{H}}^2 = \alpha_{\mathbf{H}}^2/2$. Note that this simplification is considered only for synthetic data in order to assess the influence of the kernel width. The generalized Gaussian cross coupling kernel $\mathbf{H}$ defined in (4) will be considered for real data.

The three methods showed similar results in terms of RMSE and interpolated images. The following shows the VB method's results. Fig. 8 shows examples of interpolated intensities after deconvolution using GP in the noise-free case ($\sigma_N^2 = 0$) and noisy case ($\sigma_N^2 = 10$) and different values of $\sigma_{\mathbf{H}}^2$, with the corresponding confidence interval images. we can observe that the structure of the Lena image can be recovered in the two cases. Moreover, in the confidence interval images, we can observe that as we go away from central cores, the confidence interval of the interpolated intensities decreases.

In order to measure the performance of the algorithms, we consider different noise variances (σ_N^2) as well as different

(a)   **(b)**

(c)   **(d)**
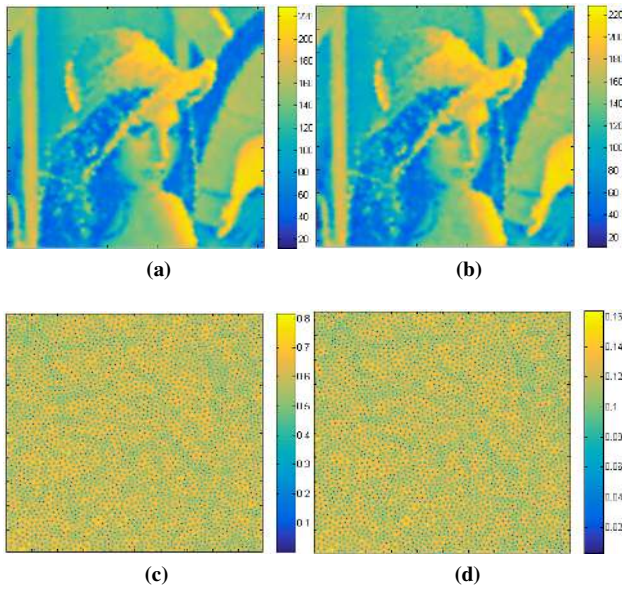
Fig. 8: Examples of interpolated samples by GP after deconvolution (a) $\sigma_N^2 = 0$ and $\sigma_H^2 = 1$, and (b) $\sigma_N^2 = 10$ and $\sigma_H^2 = 20$, and the corresponding confidence interval images.
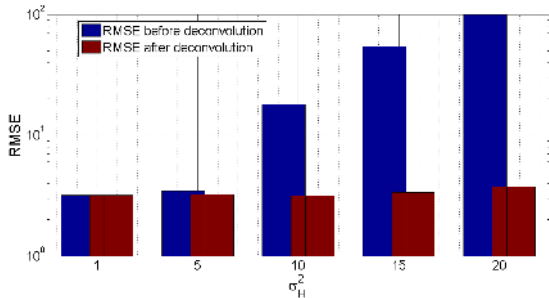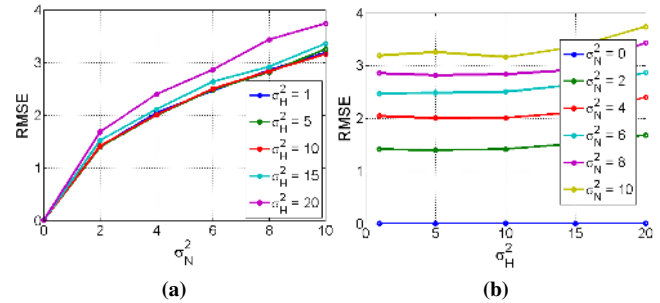

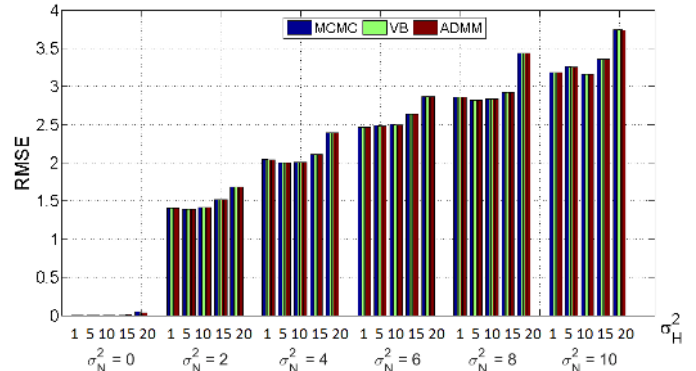**Fig. 9:** Plot of RMSEs before and after deconvolution (in-log scale) versus $\sigma_{\mathbf{H}}^2$ at $\sigma_N^2 = 10$.

cross coupling effects ($\sigma_{\mathbf{H}}^2$). Fig. 9 shows the RMSE (in log-scale) before and after deconvolution versus $\sigma_{\mathbf{H}}^2$ at $\sigma_N^2 = 10$. We can observe that all of the methods are very effective since the RMSE after deconvolution is always lower than that before deconvolution. Moreover, the gain increases with cross coupling.

In order to analyze the effect of noise variance and cross coupling separately, we fix one of them and change the other as shown in Fig. 10. In this figure, we show plots of RMSEs after deconvolution for different σ_N^2 at fixed $\sigma_{\mathbf{H}}^2$ and vice versa. In Fig. 10(a), we can observe that there is roughly a linear relationship between RMSE and σ_N^2 at fixed $\sigma_{\mathbf{H}}^2$. Moreover, the behaviour at $\sigma_{\mathbf{H}}^2 = 1, 5, 10$ and 15 is almost the same. In Fig. 10(b), we observe that RMSE is fairly constant as $\sigma_{\mathbf{H}}^2$ increases at constant σ_N^2. Furthermore, it starts to increase as σ_N^2 increases but still remains constant when changing $\sigma_{\mathbf{H}}^2$.

For the MCMC method, in all of the simulations in this paper including the real datasets, $N_{\mathrm{MC}} = 1500$, including $N_{\mathrm{bi}} = 500$, which were determined visually from preliminary runs, were used. For the ADMM method, different regularization parameter values are tested, we pick up the one corresponding to the lowest RMSE.

(a)   **(b)**

Fig. 10: Plot of RMSEs after deconvolution (a) versus σ_N^2 at fixed σ_H^2, and (b) versus σ_H^2 at fixed σ_N^2.

Fig. 11: Plot of RMSEs before and after deconvolution for the three methods versus σ_N^2 as well as σ_H^2.

C. Comparison

In this section, we compare the three proposed methods for deconvolution and restoration of OEM images. The comparison is conducted in terms of RMSE before and after deconvolution, as well as in terms of computation time.

Fig. 11 compares RMSEs after deconvolution versus different σ_N^2 as well as different $\sigma_{\mathbf{H}}^2$. We can observe that for all of the methods, as σ_N^2 increases at constant $\sigma_{\mathbf{H}}^2$, RMSE increases. On the other hand, at fixed σ_N^2, RMSE seems to be roughly constant for $\sigma_{\mathbf{H}}^2 = 1, 5$, and 10, then, it starts to increase as $\sigma_{\mathbf{H}}^2$ increases. It is clear that all the methods behave similarly in terms of RMSE.

Table I shows the average computation time (in seconds) of the three proposed methods. The experiments were conducted on ACER core-i3-2.0 GHz processor laptop with 8 GB RAM. It is clear that the MCMC method is the most computationally expensive method. The ADMM method is second, and the VB the least. Despite the relatively high computation time of the MCMC method, it is a parameter free method compared to the ADMM-based method in which the regularization parameter λ should be chosen carefully. The VB approach is considered to be the best compared to MCMC and ADMM, it can provide similar RMSE but with lower computation complexity, moreover, it is fully automatic in the sense that it can estimate the hyperparameters associated with the parameters as mentioned previously in section IV-B.

Although the MCMC and ADMM algorithms can estimate the noise variance and model hyperparameters, in practice these parameters are very difficult to estimate accurately,

TABLE I: The average computation time (in seconds) of the three proposed methods. In order to maintain a fair comparison between the three algorithms, the computational time of the ADMM algorithm corresponds to the duration of five runs (used to select the best regularization parameter among the five values).

Method	MCMC	ADMM	VB
Computation time (sec.)	3100	35.51	**5.12**

(specifically σ^2 and γ^2) due to the similarity between $\mathbf{H}^T\mathbf{H}$ and $\mathbf{\Delta}^{-1}$ in (15b) and (27b). Therefore, we have to make an informed choice about one of these parameters, specifically the choice of the hyperparameters α, α_0 and β_0 in (9) and (10). In Fig. 10(b), we observe that the RMSEs in practise are close to the true noise standard deviation, and hence the noise variance can be inferred.

## VII. SIMULATIONS USING REAL DATA

The performance of the proposed methods has been evaluated on two real datasets; the 1951 USAF resolution test chart and *ex vivo* human lung tissue. Both of them were collected using OEM system [7] with monochrome detection (Grasshopper3 camera GS3-U3-23S6M-C, Point Grey Research, Canada) and 470 nm LED illumination (M470L3, Thorlabs Ltd, UK) for lung autofluorescence excitation. Excised human lung tissue was placed in a well plate. Human tissue was used with regional ethics committee (REC: 13/ES/0126) approval and was retrieved from the periphery of specimens taken from lung cancer resections. In order to adjust the cross coupling kernel parameters $\alpha_{\mathbf{H}}$ and $\beta_{\mathbf{H}}$, a study was performed to measure, analyze and quantify intercore coupling within coherent fibre bundles [10]. This study showed how light is spread over the neighbouring cores, and gave statistical analysis on coupling percent in neighbouring cores. It showed that around 61% of transmitted light remains in the central core, around 34% in the first neighbouring cores, around 4% in the second neighbouring cores, and less than 1% in the third, fourth and fifth neighbouring cores. This leads to fixing $\alpha_{\mathbf{H}} = 4$ (in pixels) and $\beta_{\mathbf{H}} = 0.8$.

### A. 1951 USAF resolution test chart

The 1951 USAF chart is a resolution test pattern set by US Air Force in 1951. It is widely accepted to test the resolution of optical imaging systems such as microscopes, cameras and image scanners [48]. Fig. 12 (a) shows the original USAF resolution test chart used in the project. The resulting image obtained by fiber bundle is shown in Fig. 12 (b) with image size 760×760 and is composed of 7,776 fiber cores (1.34% of the image).

A non-linear interpolation based on GP of central core intensities of the image in Fig. 12(b) is presented in Fig.13(a), with the corresponding confidence intervals image in Fig.13(c). We can observe the blurring which is caused by the cross coupling effect as well as the sparsity of the data.

The outputs of the MCMC, VB, and ADMM algorithms are very similar. Thus, we show the results of the VB method. Fig. 13(b) shows an example of one of the output images with the corresponding confidence intervals in Fig. 13(d). The set of ticker strips (top left corner of the image) is now better
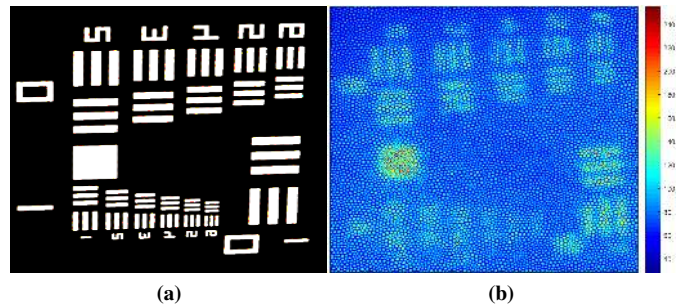
Fig. 12: (a) Scanned image of an USAF 1951 Resolution test chart. (b) The 1951 USAF resolution test chart imaged by the OEM system.
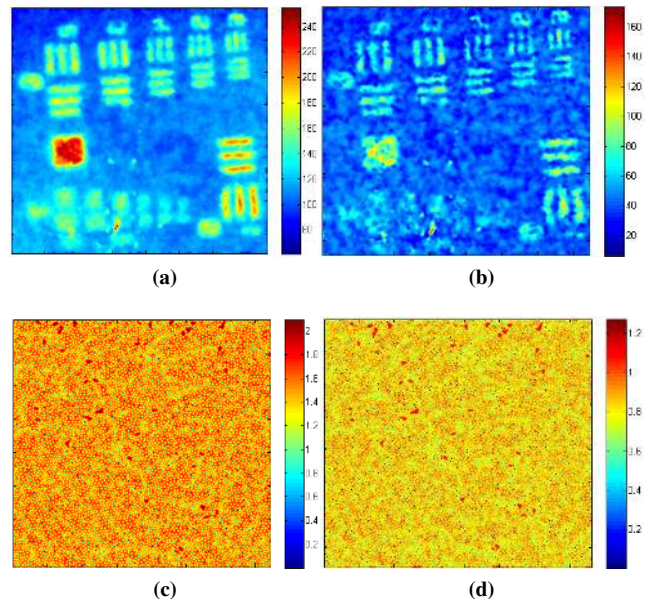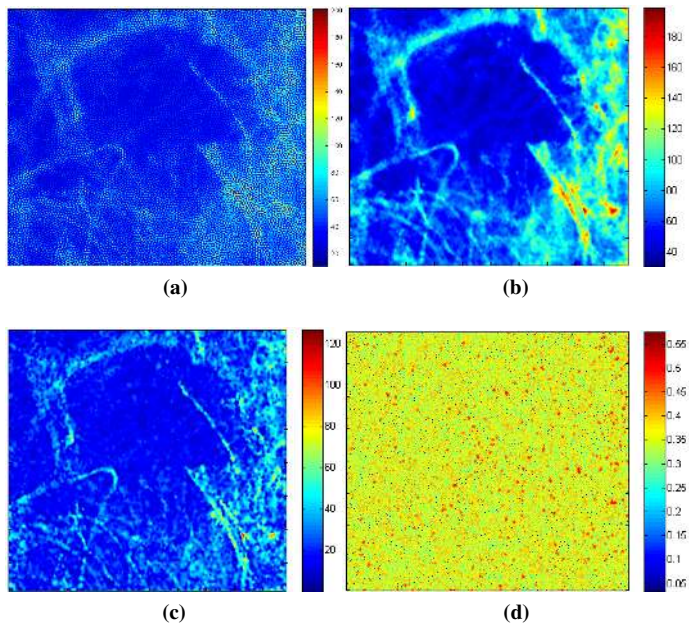

Fig. 13: Non-linear interpolation (a) before, and (b) after deconvolution, and their corresponding confidence intervals in (c), and (d) respectively.

resolved and the overlap between them is reduced. The small set of strips which is at the bottom could not be resolved, which gives an indication about the resolving resolution of this endomicroscopy system. Regions of high uncertainty (which appear as blobs in dark red) are where there may be no cores or they are dead, this in addition to the irregular core sampling are the reasons for some strips appear a bit fragmented.

B. Ex vivo human lung tissues

Fig. 14(a) shows the output image of the OEM system. Image size is 1000×800 and is composed of 13,343 fiber cores (1.66% of the image). Non-linear interpolation based on GP of central core intensities is presented in Fig.14(b). Similar to the USAF resolution test chart, we aim at reducing cross coupling effect as well as getting a more resolved image.

Similar to the USAF resolution test chart results, the outputs of the MCMC, VB, and ADMM algorithms are very similar. We only show the results of the VB method. Fig. 14(c) shows an example of interpolated deconvolved samples using GP. The lung structure is now better resolved and more sharper than before deconvolution. Moreover, confidence intervals are

Fig. 14: (a) *Ex vivo* lung tissue imaged by the endomicroscopy system [7]. Non-linear interpolation (b) before, and (c) after deconvolution, (d) the confidence intervals of the image in (c).

shown in Fig. 14(d). We can observe that as we move away from the central cores, the confidence of the interpolated intensities decreases and vice versa.

Table II provides the computation time of the 1951 USAF resolution test chart and the *ex vivo* lung tissue image. It is clear that the VB is still the fastest despite the change of the images size.

TABLE II: Computation time (in seconds) for the real data. In order to keep a fair comparison between the three algorithms, the computational times of the ADMM algorithm correspond to the duration of five runs (used to select the best regularization parameter among five values).

Dataset/Method	MCMC	ADMM	VB
USAF chart	1.12×10^5	250	**5.9**
Lung tissue	1.46×10^6	870	**16.05**

VIII. Conclusion and Future Work

This paper introduced a hierarchical Bayesian model and three estimation algorithms for the deconvolution of optical endomicroscopy images. The deconvolution accounts and compensates for fibre core cross coupling which causes major image degradation in this type of imaging. The resulting joint posterior distribution was used to approximate the Bayesian estimators. First, a Markov chain Monte Carlo procedure based on a Gibbs sampler algorithm was used to sample the posterior distribution of interest and to approximate the MMSE estimators of the unknown parameters using the generated samples. Second, a variational Bayes approach to approximate the joint posterior distribution by minimizing the Kullback-Leibler divergence was used. Third, an approach based on an alternating direction method of multipliers was used to approximate the maximum a posteriori estimators. The three algorithms showed similar estimation performance

while providing different characteristics, the MCMC and VB based approaches are fully automatic in the sense that they can jointly estimate the hyperparameters associated with the priors, however, the MCMC based approach showed high computational complexity which could be overcome by the VB and ADMM approaches. Although the ADMM approach has low computational complexity, it is semi-supervised in the sense that the hyperparameters associated with the priors need to be chosen carefully by the user. A non-linear interpolation approach based on Gaussian processes was considered to restore the full images from the samples to provide a meaningful image for interpretation. In the future, we will consider temporal information while deconvolving. Accounting for the different core sizes is also clearly an interesting route currently under investigation.

References

[1] J. Chastre and J.-Y. Fagon, "Ventilator-associated pneumonia," *American journal of respiratory and critical care medicine*, vol. 165, no. 7, pp. 867–903, 2002.

[2] P. Johnston, D. F. McAuley, and C. M. O'Kane, "Novel pulmonary biomarkers in the diagnosis of vap," *Thorax*, vol. 65, no. 3, pp. 190–192, 2010.

[3] V. S. Baselski and R. G. Wunderink, "Bronchoscopic diagnosis of pneumonia." *Clinical microbiology reviews*, vol. 7, no. 4, pp. 533–558, 1994.

[4] M. Pierce, D. Yu, and R. Richards-Kortum, "High-resolution fiber-optic microendoscopy for in situ cellular imaging," *Journal of visualized experiments: JoVE*, no. 47, 2011.

[5] D. Shin, M. C. Pierce, A. M. Gillenwater, M. D. Williams, and R. R. Richards-Kortum, "A fiber-optic fluorescence microscope using a consumer-grade digital camera for in vivo cellular imaging," *PLoS One*, vol. 5, no. 6, p. e11218, 2010.

[6] X. Hong, V. K. Nagarajan, D. H. Mugler, and B. Yu, "Smartphone microendoscopy for high resolution fluorescence imaging," *Journal of Innovative Optical Health Sciences*, vol. 9, no. 05, p. 1650046, 2016.

[7] N. Krstajić, A. R. Akram, T. R. Choudhary, N. McDonald, M. G. Tanner, E. Pedretti, P. A. Dalgarno, E. Scholefield, J. M. Girkin, A. Moore *et al.*, "Two-color widefield fluorescence microendoscopy enables multiplexed molecular imaging in the alveolar space of human lung tissue," *Journal of Biomedical Optics*, vol. 21, no. 4, pp. 046 009–046 009, 2016.

[8] H. A. Wood, K. Harrington, J. M. Stone, T. A. Birks, and J. C. Knight, "Quantitative characterization of endoscopic imaging fibers," *Optics Express*, vol. 25, no. 3, pp. 1985–1992, 2017.

[9] K. L. Reichenbach and C. Xu, "Numerical analysis of light propagation in image fibers or coherent fiber bundles," *Optics express*, vol. 15, no. 5, pp. 2151–2165, 2007.

[10] A. Perperidis, H. E. Parker, A. Karam-Eldaly, Y. Altmann, K. Dhaliwal, R. R. Thomson, M. G. Tanner, and S. McLaughlin, "Characterization and modelling of inter-core coupling in coherent fiber bundles," *Optics Express*, vol. 25, no. 10, pp. 11 932–11 953, 2017.

[11] J. A. Udovich, N. D. Kirkpatrick, A. Kano, A. Tanbakuchi, U. Utzinger, and A. F. Gmitro, "Spectral background and transmission characteristics of fiber optic imaging bundles," *Applied optics*, vol. 47, no. 25, pp. 4560–4568, 2008.

[12] P. Sarder and A. Nehorai, "Deconvolution methods for 3-d fluorescence microscopy images," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 32–45, 2006.

[13] M. B. Cannell, A. McMorland, and C. Soeller, "Image enhancement by deconvolution," *Handbook of biological confocal microscopy*, pp. 488–500, 2006.

[14] J. Qin, X. Yi, S. Weiss, and S. Osher, "Shearlet-tgv based fluorescence microscopy image deconvolution," *CAM Report. University of California, Los Angeles (UCLA)*, pp. 14–32, 2014.

[15] F. Šroubek and J. Flusser, "Multichannel blind iterative image restoration," *Image Processing, IEEE Transactions on*, vol. 12, no. 9, pp. 1094–1106, 2003.

[16] M. A. Figueiredo, J. M. Bioucas-Dias, and M. V. Afonso, "Fast frame-based image deconvolution using variable splitting and constrained optimization," in *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on*. IEEE, 2009, pp. 109–112.

[17] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Non-cyclic deconvolution using an augmented lagrangian method," in *EUROCON-International Conference on Computer as a Tool (EUROCON), 2011 IEEE*. IEEE, 2011, pp. 1–4.

[18] X. Zhou, R. Molina, F. Zhou, and A. K. Katsaggelos, "Fast iteratively reweighted least squares for l p regularized image deconvolution and reconstruction," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1783–1787.

[19] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[20] R. Courant *et al.*, "Variational methods for the solution of problems of equilibrium and vibrations," *Bull. Amer. Math. Soc*, vol. 49, no. 1, pp. 1–23, 1943.

[21] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[22] A. Wong, X. Y. Wang, and M. Gorbet, "Bayesian-based deconvolution fluorescence microscopy using dynamically updated nonstationary expectation estimates," *Scientific reports*, vol. 5, 2015.

[23] T. E. Bishop, R. Molina, and J. R. Hopgood, "Blind restoration of blurred photographs via ar modelling and mcmc," in *Image Processing, ICIP*. IEEE, 2008, pp. 669–672.

[24] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind image deconvolution: A review," *Digital Signal Processing*, vol. 47, pp. 116–127, 2015.

[25] F. Al-Awadhi, C. Jennison, and M. Hurn, "Statistical image analysis for a confocal microscopy two-dimensional section of cartilage growth," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 53, no. 1, pp. 31–49, 2004.

[26] F. Al-Awadhi, M. Hurn, and C. Jennison, "Three-dimensional bayesian image analysis and confocal microscopy," *Journal of Applied Statistics*, vol. 38, no. 1, pp. 29–46, 2011.

[27] P. Pankajakshan, B. Zhang, L. Blanc-Féraud, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, "Blind deconvolution for thin-layered confocal imaging," *Applied optics*, vol. 48, no. 22, pp. 4437–4448, 2009.

[28] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind deconvolution using a total variation prior," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 12–26, 2009.

[29] S. D. Babacan, R. Molina, and A. Katsaggelos, "Variational bayesian super resolution," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 984–999, 2011.

[30] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3715–3727, 2006.

[31] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Parameter estimation in tv image restoration using variational distribution approximation," *IEEE transactions on image processing*, vol. 17, no. 3, pp. 326–339, 2008.

[32] C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[33] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[34] M. J. Beal, *Variational algorithms for approximate Bayesian inference*. University of London United Kingdom, 2003.

[35] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[36] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, 2006, vol. 1.

[37] A. Pakman and L. Paninski, "Exact hamiltonian monte carlo for truncated multivariate gaussians," *Journal of Computational and Graphical Statistics*, vol. 23, no. 2, pp. 518–542, 2014.

[38] V. Šmídl and A. Quinn, *The variational Bayes method in signal processing*. Springer Science & Business Media, 2006.

[39] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[40] G. Parisi and R. Shankar, "Statistical field theory," 1988.

[41] J. W. Miskin, "Ensemble learning for independent component analysis," in *in Advances in Independent Component Analysis*. Citeseer, 2000.

[42] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, "Blind image deconvolution: problem formulation and existing approaches," *Blind image deconvolution: Theory and Applications*, pp. 1–41, 2007.

[43] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *Image Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 681–695, 2011.

[44] M. Luessi, S. D. Babacan, R. Molina, J. R. Booth, and A. K. Katsaggelos, "Variational bayesian causal connectivity analysis for fmri," *Frontiers in neuroinformatics*, vol. 8, 2014.

[45] M. A. Figueiredo and J. M. Bioucas-Dias, "Restoration of poissonian images using alternating direction optimization," *IEEE transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, 2010.

[46] G. Le Goualher, A. Perchant, M. Genet, C. Cavé, B. Viellerobe, F. Berier, B. Abrat, and N. Ayache, "Towards optical biopsies with an integrated fibered confocal fluorescence microscope," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004*. Springer, 2004, pp. 761–768.

[47] N. Ayache, T. Vercauteren, G. Malandain, F. Oberrietter, N. Savoire, and A. Perchant, "Processing and mosaicing of fibered confocal images," in *MICCAI Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB06)*, 2006.

[48] SilverFast, "Silverfast resolution target (usaf 1951) by lasersoft imaging," *available online http://www.silverfast.com/show/resolution-target/en.html*, 2012.