POLITECNICO DI TORINO
Repository ISTITUZIONALE

Nonconvex image reconstruction via expectation propagation

Original
Nonconvex image reconstruction via expectation propagation / Paola Muntoni, Anna; Diaz Hernandez Rojas, Rafael; Braunstein, Alfredo; Pagnani, Andrea; Perez Castillo, Isaac. - In: PHYSICAL REVIEW. E. - ISSN 2470-0053. - ELETTRONICO. - 100:(2019), p. 032134.

Availability:
This version is available at: 11583/2754133 since: 2019-09-23T17:24:31Z

Publisher:
American Physical Society

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

29 March 2024

Nonconvex image reconstruction via expectation propagation

Anna Paola Muntoni,[1,2,3] Rafael Díaz Hernández Rojas,[4] Alfredo Braunstein,[1,5,6,7,*]
Andrea Pagnani,[1,5,6,*] and Isaac Pérez Castillo[8,9,*]

[1]*Department of Applied Science and Technologies (DISAT), Politecnico di Torino, Corso Duca Degli Abruzzi 24, Torino, Italy*
[2]*Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005, Paris, France*
[3]*Sorbonne Université, CNRS, Institut de Biologie Paris-Seine, Laboratory of Computational and Quantitative Biology, F-75005, Paris, France*
[4]*Dipartimento di Fisica, Sapienza University of Rome, P.le Aldo Moro 5, I-00185 Rome, Italy*
[5]*Italian Institute for Genomic Medicine (form. HuGeF) SP142 km 3.95 - 10060 Candiolo, Italy*
[6]*INFN Sezione di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[7]*Collegio Carlo Alberto, Piazza Vincenzo Arbarello, 8 - 10122 Torino, Italy*
[8]*Departamento de Física Cuántica y Fótonica, Instituto de Física, UNAM, P. O. Box 20-364, 01000 Cd. Mx., México*
[9]*London Mathematical Laboratory, 8 Margravine Gardens, W6 8RH London, United Kingdom*

(Received 13 September 2018; published 23 September 2019)

The problem of efficiently reconstructing tomographic images can be mapped into a Bayesian inference problem over the space of pixels densities. Solutions to this problem are given by pixels assignments that are compatible with tomographic measurements and maximize a posterior probability density. This maximization can be performed with standard local optimization tools when the log-posterior is a convex function, but it is generally intractable when introducing realistic nonconcave priors that reflect typical images features such as smoothness or sharpness. We introduce a new method to reconstruct images obtained from Radon projections by using expectation propagation, which allows us to approximate the intractable posterior. We show, by means of extensive simulations, that, compared to state-of-the-art algorithms for this task, expectation propagation paired with very simple but non-log-concave priors is often able to reconstruct images up to a smaller error while using a lower amount of information per pixel. We provide estimates for the critical rate of information per pixel above which recovery is error-free by means of simulations on ensembles of phantom and real images.

DOI: 10.1103/PhysRevE.100.032134

I. INTRODUCTION

The classical problem in image reconstruction consists in recovering the density of an object's two-dimensional (2D) slice from a set of Radon transformations. These correspond to a set of projections on a plane, which can be usually identified with the absorption of radiation by the object along a given line of response. In the ideal case, and from a purely mathematical point of view, the original image can be reconstructed when enough projections are available by applying the inverse Radon transform, a method usually referred to as filtered back-projection [1]. However, in real scenarios, detectors and radiation sources have an actual finite size, data collection is restricted to a short time window and measurements are naturally noisy. With limited and/or noisy information, it is only possible, in principle, to reconstruct a finite-resolution discretization $\boldsymbol{x} = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ of the image from a finite set of Radon projections. In these more realistic scenarios, algebraic reconstruction techniques (ART) are normally used. Under reasonable working hypotheses, discretized images that are compatible with the measurements are the ones satisfying a set of linear relations:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{p}, \tag{1}$$

where $\boldsymbol{p} = (p_1, \ldots, p_M)^T \in \mathbb{R}^M$ is the M-dimensional measurements vector of projection data, and $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{M\times N}$ is the so-called *projection matrix*. Here variable x_j represents the density of the image at the position of pixel j, while the entries a_{ij} of the matrix $\boldsymbol{A}$ correspond to the length of the intersection of the ith projection ray with the jth pixel or, in other terms, to the contribution of the jth pixel to the total attenuation along the ith ray. In the field of image reconstruction, a pivotal role is played by algorithms capable of providing accurate reconstructions with the lowest possible number of measurements M. Often in medical imaging, one faces practical constraints posed by the acquisition system, and/or the need to mitigate the dangerous effects of ionizing radiation exposure. In the following we will mostly concentrate on this *limited data* regime that corresponds to an underdetermined system of equations in Eq. (1), i.e., $M < N$, where the system has infinitely many solutions in the noiseless regime and none in the noisy one. Using the ART algorithm [2,3], one can, however, obtain an approximate reconstruction of the image $\boldsymbol{x}$ by iteratively minimizing the ℓ_2 error, $||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{p}||_2$. Its performance can be drastically enhanced by combining it with the total variation (TV) method [4] which relies on an a

[*]These authors contributed equally to this work.

 ©2019 American Physical Society

priori knowledge that realistic images have intrinsic structure, in particular smoothness, that can be encoded by means of a ℓ_1 sparsity regularization on the (discrete) gradient of the image.

The reconstruction problem can be alternatively recast in the language of Bayesian inference by considering the posterior probability distribution of images for a given vector of measurements $\boldsymbol{p}$

$$P(\boldsymbol{x}|\boldsymbol{p}) = P(\boldsymbol{p})^{-1} P(\boldsymbol{p}|\boldsymbol{x}) P_0(\boldsymbol{x}). \tag{2}$$

The reconstructed image is often given by the *maximum a posteriori* (MAP) estimation $\boldsymbol{x}^\star = \arg\max_x P(\boldsymbol{x}|\boldsymbol{p})$.

The *likelihood* term $P(\boldsymbol{p}|\boldsymbol{x})$ corresponds to the discretized model in Eq. (1) and takes the form $P(\boldsymbol{p}|\boldsymbol{x}) = \delta(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{p})$ for the noiseless case [5]. If measurements are affected by independent additive Gaussian noise (although other noise models can be assumed), then the likelihood reads $P(\boldsymbol{p}|\boldsymbol{x}) \propto \exp[-\frac{\beta}{2}(\boldsymbol{p} - \boldsymbol{A}\boldsymbol{x})^2]$, where β is the inverse variance of the noise distribution.

The *prior* $P_0(\boldsymbol{x})$ plays a crucial role as it allows to include further information complementing the set of measurements, making the reconstruction possible in the underdetermined regime. Fairly intuitively, a smaller amount of image-specific information is needed to perform the reconstruction, provided we have access to a more informative prior on the class of images. Both ℓ_2 and ℓ_1 regularizations can be mapped in this framework as log-concave priors that admit the computation of the corresponding MAP estimates by means of standard convex optimization techniques. The mapping between Eq. (2) and ℓ_p regularization for $p > 0$ is straightforward by considering $P_0(\boldsymbol{x}) \propto \exp(-\lambda \|\boldsymbol{x}\|_p)$, as the measure then concentrates on the minima of ℓ_p for $\lambda \to \infty$.

A thorough analysis of actual tomographic and natural images reveals that in many cases the statistics of pixel intensities are ill fitted by trivial log-concave functions [6–9] but can, in principle, be well fitted by priors involving non-log-concave terms. In particular, as it can be seen in Fig. 1, the distribution of neighbor pixel differences in real tomographic images present a cusp at zero which seems much more compatible with a ℓ_p norm with $0 < p < 1$ than with $p = 1$ or $p = 2$. Unfortunately, the change from $p \geqslant 1$ to $p < 1$ renders in principle the MAP estimate a computationally formidable task as it leads to a nonconvex optimization problem. Furthermore, in this setting the mean value of $\boldsymbol{x}$ of the posterior distribution $P(\boldsymbol{x}|\boldsymbol{p})$ should be generally preferred to the MAP estimation as it is the vector that minimizes the mean squared error between the reconstructed and the original signal. To cope with these average values, the computation of the marginal probability densities plays a key role.

Recently, groundbreaking applications of statistical mechanics techniques to nonconvex computational problems have yielded very efficient algorithms and reliable methods to make the computation of marginals of complicated multivariate distributions computationally tractable. These techniques have been recently and successfully applied to image reconstruction of binary images for the case of discrete tomography [5]. Here it was noted that the belief propagation (BP) algorithm provides better reconstruction than TV in some cases, especially in the high-noise regime. Statistical techniques such as BP have the additional advantage of being able to deal more efficiently in the imperfect reconstruction regime than optimization methods, since the maximum probability point may be uninformative when the posterior distribution is not very concentrated. The BP algorithm, however, relies on the Bethe-Peierls approximation, which is inaccurate in many realistic scenarios. Although corrections to the Bethe-Peierls

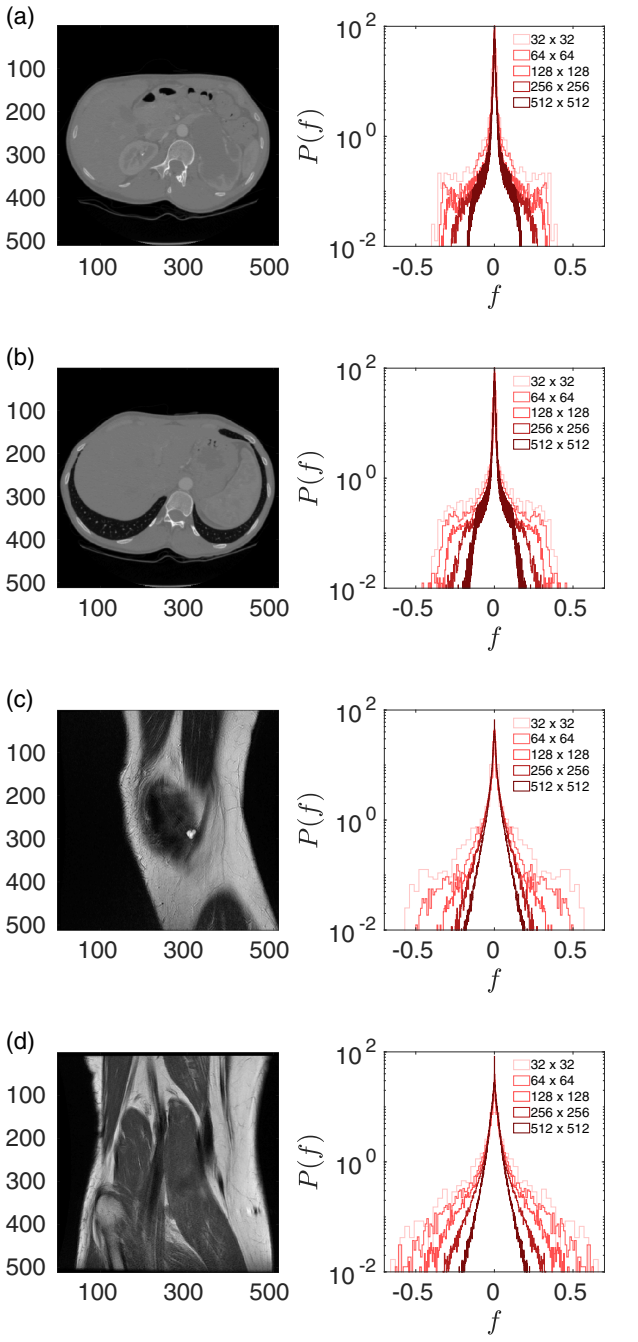


FIG. 1. Computed tomography scans (left panels) and their corresponding empirical distributions (right panels), going from full resolution images of 512×512 (corresponding to the narrower, dark red empirical distribution) to lower resolutions 256×256, 128×128, 64×64, and 32×32 (the latter corresponding to the wider, light red empirical distribution). Panels (a) and (b) show two different viewpoints of the abdomen while (c) and (d) two different views of a knee.

approximation abound in the literature [10–13], they become, more often than not, impractical for current applications.

Our main purpose is to introduce a family of priors with a corresponding family of algorithms based on expectation propagation (EP), whose reconstruction performance surpasses the ones obtained with standard log-concave priors and standard local optimization algorithms. Expectation propagation, originally introduced in Refs. [14–17], gives additionally a natural probabilistic framework to maximize the inference performance in the imperfect reconstruction regime. In particular, it allows us to compute an approximation of the posterior marginal distribution and the posterior average, allowing in principle for a more accurate reconstruction. It should be noted that ℓ_p regularization with $0 < p < 1$ for tomographic reconstruction has been considered at least in Ref. [18]. However, in this latter work, the proposed reconstruction algorithm is based on a local optimizer and it lacks the probabilistic framework and interpretation proposed here.

This work is organized as follows: In Sec. II we introduce the method of EP together with the different priors we have used when reconstructing images. Section III is dedicated to explain how one can estimate the various parameters of the method based solely on probabilistic arguments. Results of our approach for phantom and real images are presented and discussed in Sec. IV. We end up with some concluding remarks in Sec. V. Thorough mathematical derivations, together with details of previous reconstructing algorithms, can be found in the appendices.

II. IMPLEMENTATION OF PRIOR KNOWLEDGE AND THE METHOD OF EXPECTATION PROPAGATION

Before discussing the method of EP to approximate the posterior distribution $P(\boldsymbol{x}|\boldsymbol{p})$, it is crucial to have a prior distribution $P_0(\boldsymbol{x})$ that captures reasonably well some of the typical properties of the images we aim to reconstruct. To achieve this, we assume that the prior $P_0(\boldsymbol{x})$ can be written as a product $P_0(\boldsymbol{x}) \propto P_0^{(\mathrm{single})}(\boldsymbol{x}) P_0^{(\mathrm{pair})}(\boldsymbol{x})$, where the factor $P_0^{(\mathrm{single})}(\boldsymbol{x})$ imposes independent local constraints designed to catch the concrete nature and support of the pixels involved, whereas $P_0^{(\mathrm{pair})}(\boldsymbol{x})$ contains all priors that can be written as product of probability distribution over pairs of variables and it is supposed to model the highly correlated nature among pixels in real images.

For the factor $P_0^{(\mathrm{single})}(\boldsymbol{x})$, we will consider three different choices. In the first one, which we will call *interval* prior, we assume a uniform measure on a generic support $[x_i^{(m)}, x_i^{(M)}]$ of the pixels, that is,

$$P_{0,\mathrm{int}}^{(\mathrm{single})}(\boldsymbol{x}) = \prod_{i=1}^{N} \frac{\mathbb{I}_{x_i \in \left[x_i^{(m)}, x_i^{(M)}\right]}}{x_i^{(M)} - x_i^{(m)}} \equiv \prod_{i=1}^{N} \Lambda_i(x_i), \tag{3}$$

where $\mathbb{I}_A$ denotes the indicator function of condition A. A second viable choice, which it is usually called the *spike-and-slab* [19] or sparse prior, is particularly useful in the reconstruction of images with extensive monochromatic background:

$$P_{0,\mathrm{sparse}}^{(\mathrm{single})}(\boldsymbol{x}) = \prod_{i=1}^{N} [s\delta(x_i) + (1-s)\Lambda_i(x_i)]. \tag{4}$$

Here the weighting factor $s \in (0, 1)$ is the sparseness parameter of the image and is equal to the average fraction of background pixel within the image. Finally, in discrete binary tomography [20,21] one assumes that the two available colors are either black or white, corresponding to a region totally transparent or completely opaque, for which we will assign values $x_i = 0$ or $x_i = 1$, respectively. In this scenario, the single variable prior, that we will denote as the *binary* prior, takes the following simple form:

$$P_{0,\mathrm{bin}}^{(\mathrm{single})}(\boldsymbol{x}) = \prod_{i=1}^{N} [s\delta(x_i) + (1-s)\delta(x_i - 1)]. \tag{5}$$

The remaining factor $P_0^{(\mathrm{pair})}(\boldsymbol{x})$ in the prior probability is supposed to favor images with certain features, such as a smooth change in the intensities of neighboring pixels. This accounts for the fact that real images possess local structure. A standard choice for $P_0^{(\mathrm{pair})}$ is

$$P_{0,\mathrm{lap}}^{(\mathrm{pair})}(\boldsymbol{x}) \propto e^{-\frac{J}{2}\boldsymbol{x}^T \cdot \boldsymbol{L} \cdot \boldsymbol{x}} \propto e^{-\frac{J}{2}\sum_{i=1}^{N}\sum_{j\in\partial i}(x_i - x_j)^2}, \tag{6}$$

where $\boldsymbol{L}$ is the Laplacian matrix of the nearest-pixels adjacency graph, J a weight parameter, and ∂i denotes the set of neighbors of pixel i. We will denote the prior in Eq. (6) as ℓ_2 smoothness, as it favors small norms of the finite-differences gradient. Notice that this prior assumes a Gaussian profile for the probability density of the difference variables, which makes, in turn, the treatment more analytically amenable. Empirically, it turns out, at least for tomographic images, that the histogram of these auxiliary variables is far from being Gaussian distributed. Indeed, Fig. 1 shows the empirical frequency count of the gradient of the image, that is, $P(f) \propto \sum_{i=1}^{N}\sum_{j\in\partial i}\delta(f - x_i + x_j)$, for a series of real CT scans. Interestingly, the empirical profiles do not depend much on the type of organ analyzed but depend in a highly nontrivial way on the coarse-graining of the image. To capture these more realistic cases, we introduce the following *spike-and-slab* prior for neighboring pixel differences, which corresponds to finite-differences partial derivatives:

$$P_{0,\mathrm{diff}}^{(\mathrm{pair})}(x_i, x_j) \propto \rho\delta(x_i - x_j) + (1-\rho)e^{-\frac{\lambda}{2}(x_i - x_j)^2}, \tag{7}$$

with $\rho \in [0, 1]$ and $\lambda \geqslant 0$. For brevity, we will refer to this prior as ℓ_0 *smoothness*. As we will show below, this prior produces very accurate image reconstructions.

Once we have selected the prior distribution, the image is reconstructed by using the first moments of the marginals of the posterior distribution $P(\boldsymbol{x}|\boldsymbol{p})$. The main goal is to find an efficient way to extract the information of this posterior. Expectation propagation was introduced to approximate posterior distributions, along with their marginals [14–16], for a large class of intractable probabilistic models. In the image reconstruction problem, the posterior distribution of images, given the projections, takes the functional form of a multivariate Gaussian with positive definite covariance matrix $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}$ times a product of univariate distributions $\psi_i(x_i)$:

$$P(\boldsymbol{x}|\boldsymbol{p}) = \frac{1}{Z} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \prod_i \psi_i(x_i). \tag{8}$$

The multivariate Gaussian term takes into account the measurements, i.e., the likelihood term, times other interacting terms in the prior, if any. For instance, in the case of the prior given by Eq. (6), the expressions for $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are

$$\boldsymbol{\Sigma}^{-1} = \beta \boldsymbol{A}^T \boldsymbol{A} + J\boldsymbol{L}, \qquad \boldsymbol{\mu} = \beta \boldsymbol{\Sigma} \boldsymbol{A} \boldsymbol{p}, \tag{9}$$

respectively. The set of functions $\psi_i(x)$ account for non-Gaussian factors, such as density bounds between 0 and 1, ℓ_0 sparsity, binary constraints, or non-log-convex priors (explicit forms for other choices of priors can be found in Appendix B). To introduce the method of EP, we proceed as follows. Suppose that, to trade off accuracy for solvability, we approximate Eq. (8) by replacing each $\psi_i(x_i)$ term by a normal density,

$$\phi_i(x_i) = \frac{1}{\sqrt{2\pi b_i}} e^{-\frac{(x_i - a_i)^2}{2b_i}}. \tag{10}$$

Then the new posterior, denoted here as $Q_{\{\phi\}}(\boldsymbol{x})$, has the following expression:

$$Q_{\{\phi\}}(\boldsymbol{x}) = \frac{1}{Z_{Q_{\{\phi\}}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})} \prod_i \phi_i(x_i). \tag{11}$$

Note that $Q_{\{\phi\}}(\boldsymbol{x})$ is a multivariate Gaussian distribution for which it is easy to obtain the single variable marginals $Q_{\{\phi\}}(x_i)$, whence the value of each pixel can be inferred as the mean of its corresponding marginal.

We are now left with the problem of choosing the mean and variance vectors, $\boldsymbol{a} = (a_1, \ldots, a_N), \boldsymbol{b} = (b_1, \ldots, b_N)$, in order to best approximate the true posterior probability $P(\boldsymbol{x}|\boldsymbol{p})$ by using $Q_{\{\phi\}}(\boldsymbol{x})$. A seemingly reasonable form for the approximating factors ϕ_i would be the closest univariate Gaussians [in Kullback-Leibler (KL) distance] to ψ_i. This approach, more often than not, produces poor results in reconstruction. The EP algorithm improves strikingly the estimation of the marginal densities by approximating not the ψ_i measures themselves but their effect on the full distribution. More precisely, we introduce the so-called tilted distribution for pixel ith, $Q^{(i)}_{\{\phi\}}(\boldsymbol{x})$:

$$Q^{(i)}_{\{\phi\}}(\boldsymbol{x}) = \frac{1}{Z_{Q^{(i)}_{\{\phi\}}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})} \psi_i(x_i) \prod_{j \neq i} \phi_j(x_j). \tag{12}$$

One then chooses the parameters (a_i, b_i) of the Gaussian distribution $\phi_i(x_i)$ such that the KL distance D_{KL} between $Q_{\{\phi\}}$ and $Q^{(i)}_{\{\phi\}}$ is minimized, that is,

$$(a_i^\star, b_i^\star) = \arg\min_{(a_i, b_i)} D_{\mathrm{KL}}\left[Q^{(i)}_{\{\phi\}} || Q_{\{\phi\}}\right]. \tag{13}$$

It is straightforward to show [15–17] that the D_{KL} minimization is equivalent to the following moment-matching condition:

$$\langle x_i \rangle_{Q_{\{\phi\}}} = \langle x_i \rangle_{Q^{(i)}_{\{\phi\}}}, \quad \langle x_i^2 \rangle_{Q_{\{\phi\}}} = \langle x_i^2 \rangle_{Q^{(i)}_{\{\phi\}}}. \tag{14}$$

Here $\langle \cdots \rangle_{Q_{\{\phi\}}}$ and $\langle \cdots \rangle_{Q^{(i)}_{\{\phi\}}}$ correspond to performing the expectation values with respect to the distributions given by Eqs. (11) and (12), respectively. In this way, Eq. (13) can be used as an iterative procedure until convergence is reached for every pair of parameters (a_i, b_i). At convergence, the value of each reconstructed pixel is determined by the formula

$$x_i^* = \langle x_i \rangle_{Q^{(i)}_{\{\phi\}}}. \tag{15}$$

This choice is motivated by the fact that using the mean of a posterior minimizes the mean squared error with respect to the true value of a variable, with the added persuasive argument that, on convergence, we have morally obtained the best Gaussian approximation to the tilted distribution, which in turn includes the ψ_i measures present in the original posterior of Eq. (8).

The explicit form of the moment matching conditions (14) depends on the choice of priors. For instance, for the case of using the interval prior, the posterior distributions Q and $Q^{(i)}$ are given by:

$$\begin{aligned} Q(\boldsymbol{x}|\boldsymbol{p}) &= \frac{1}{Z_Q} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})^T \cdot \boldsymbol{\Sigma}^{-1}_{(i)} \cdot (\boldsymbol{x}-\boldsymbol{\mu}^{(i)})} \frac{e^{-\frac{(x_i - a_i)^2}{2b_i}}}{\sqrt{2\pi b_i}}, \\ Q^{(i)}(\boldsymbol{x}|\boldsymbol{p}) &= \frac{1}{Z_Q^{(i)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})^T \cdot \boldsymbol{\Sigma}^{-1}_{(i)} \cdot (\boldsymbol{x}-\boldsymbol{\mu}^{(i)})} \frac{\mathbb{I}_{x_i \in \left[x_i^{(m)}, x_i^{(M)}\right]}}{x_i^{(M)} - x_i^{(m)}}, \end{aligned} \tag{16}$$

where we have defined the following matrices and vectors:

$$\begin{aligned} \boldsymbol{\Sigma}^{-1}_{(i)} &= \beta \boldsymbol{A}^T \boldsymbol{A} + J\boldsymbol{L} + \boldsymbol{B}^{(i)}, \\ \boldsymbol{\mu}^{(i)} &= \boldsymbol{\Sigma}_{(i)} \cdot (\beta \boldsymbol{A}^T \cdot \boldsymbol{p} + \boldsymbol{B}^{(i)} \cdot \boldsymbol{a}), \\ \boldsymbol{B}^{(i)} &= \mathrm{diag}\left(b_1^{-1}, \ldots, b_{i-1}^{-1}, 0, b_{i+1}^{-1}, \ldots, b_N^{-1}\right). \end{aligned} \tag{17}$$

Solving the moment matching conditions of Eq. (14) for the parameters $\{(a_i, b_i)\}_{i=1}^N$ results in:

$$\begin{aligned} b_i &= \left[\frac{1}{\langle x_i^2 \rangle_{Q^{(i)}} - \langle x_i \rangle^2_{Q^{(i)}}} - \frac{1}{\Sigma_{ii}}\right]^{-1}, \\ a_i &= b_i \left[\langle x_i \rangle_{Q^{(i)}} \left(\frac{1}{b_i} + \frac{1}{\Sigma_{ii}}\right) - \frac{\mu_i}{\Sigma_{ii}}\right]. \end{aligned} \tag{18}$$

To shorten notation we have defined $\Sigma_{ii} \equiv (\boldsymbol{\Sigma}_{(i)})_{ii}$, whereas $\langle x_i \rangle_{Q^{(i)}}$ and $\sigma^2_{Q^{(i)}} \equiv \langle x_i^2 \rangle_{Q^{(i)}} - \langle x_i \rangle^2_{Q^{(i)}}$ have the following expressions:

$$\begin{aligned} \langle x_i \rangle_{Q^{(i)}} &= \mu_i + \frac{\mathcal{N}\left[\frac{x_i^{(m)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right] - \mathcal{N}\left[\frac{x_i^{(M)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right]}{\Phi\left[\frac{x_i^{(M)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right] - \Phi\left[\frac{x_i^{(m)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right]} \Sigma_{ii}, \\ \sigma^2_{Q^{(i)}} &= \Sigma_{ii} \left\{ 1 + \frac{\frac{x_i^{(m)} - \mu_i}{\Sigma_{ii}} \mathcal{N}\left[\frac{x_i^{(m)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right] - \frac{x_i^{(M)} - \mu_i}{\Sigma_{ii}} \mathcal{N}\left[\frac{x_i^{(M)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right]}{\Phi\left[\frac{x_i^{(M)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right] - \Phi\left[\frac{x_i^{(m)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right]} \right. \\ &\quad \left. - \left(\frac{\mathcal{N}\left[\frac{x_i^{(m)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right] - \mathcal{N}\left[\frac{x_i^{(M)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right]}{\Phi\left[\frac{x_i^{(M)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right] - \Phi\left[\frac{x_i^{(m)} - \mu_i}{\sqrt{\Sigma_{ii}}}\right]} \right)^2 \right\}, \end{aligned} \tag{19}$$

with definitions

$$\Phi(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \qquad \mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{20}$$

In this particular case, the moment matching condition given by the set of Eqs. (18) is solved by using a standard fixed-point iteration method, that is, starting with initial values of $\{(a_i, b_i)\}_{i=1}^N$ the set of Eqs. (18) is iterated until convergence. Even though a theoretical analysis of the convergence of the EP algorithm is beyond the scope of the present paper, we have generally found, when running it in phantom and real

tomographic images, that our method needs of the order of 200 to 400 iteration steps on the set of equations (18) to stop. Convergence is numerically reached when the error

$$\varepsilon^t = \max_i \left\{ \left| \langle x_i \rangle_{Q_t^{(i)}} - \langle x_i \rangle_{Q_{t-1}^{(i)}} \right|, \left| \langle x_i^2 \rangle_{Q_t^{(i)}} - \langle x_i^2 \rangle_{Q_{t-1}^{(i)}} \right| \right\} \quad (21)$$

is smaller than a tolerance of the order of 10^{-7}.

III. PARAMETERS ESTIMATION

Unlike other methods used in image reconstruction, those based in Bayesian inference allow us to estimate fairly naturally the set of parameters of the model to obtain an optimal reconstruction. In our particular case, the set of parameters to infer depends on the actual choice of the prior distribution. To fix ideas let us consider, for instance, a prior distribution consisting on the binary prior, together with the ℓ_2 smoothness prior, given by Eqs. (5) and (6), respectively. This choice then contains three parameters: the inverse variance noise distribution β, the weight of the Laplacian matrix J, and the sparseness parameter s. Our approach can be understood as an expectation maximization [22] technique applied at every time step of the main EP algorithm. In particular, at each update of the mean and variances of the approximation $\{(a_i, b_i)\}_{i=1}^N$, we temporarily fix the distribution of each pixel to its tilted distribution at time t and compute the values of the parameters s, J, β, which maximizes the likelihood of the data $P(\boldsymbol{p}|s, J, \beta)$. This computation leads to the following update equations for β and J:

$$\beta^{(t+1)} = \frac{M}{[\boldsymbol{A}\boldsymbol{x}_{\rm EP}^{(t)} - \boldsymbol{p}]^T [\boldsymbol{A}\boldsymbol{x}_{\rm EP}^{(t)} - \boldsymbol{p}]},$$
$$J^{(t+1)} = \frac{N}{\boldsymbol{x}_{\rm EP}^{(t)} \boldsymbol{L} \boldsymbol{x}_{\rm EP}^{(t)}}. \quad (22)$$

where $\boldsymbol{x}_{\rm EP}^{(t)}$ are the EP estimates at time t (see Appendix C for details). Unfortunately, there is no explicit formula for the sparsity parameter, but one can perform a gradient descent over $-\log[P(\boldsymbol{p}|s, J, \beta)]$. This function, in a statistical mechanics jargon, is equivalent to an intractable free energy that can be approximated using EP approximation. Thus the sparsity can be updated at each iteration according to

$$s^{(t+1)} = s^{(t)} - \eta \frac{\partial F_{\rm EP}^{(t)}}{\partial s}, \quad (23)$$

where $F_{\rm EP}^{(t)}$ is the EP free energy (see further details in Appendix C) at iteration step t evaluated using $\boldsymbol{x}_{\rm EP}^{(t)}$, while η is a relaxation parameter of the gradient descent algorithm. On convergence, this scheme yields an estimation of $s^\star$ which corresponds to a local minimum, thus assuming that the real image sparseness minimizes the EP free energy. This indeed seemed to be the case, as our binary reconstructions did show a low fraction of errors under many circumstances once $s^\star$ was inferred. The value of ρ and λ in the difference variables prior, appearing in Eq. (7), can be inferred using the same technique.

IV. RESULTS

To estimate the goodness of the prior distribution choice, we compare our performances against three reconstruction methods commonly used in the literature: TV (for ℓ_1 smoothness); quadratic programming (QP) (for ℓ_2 smoothness); and, for binary reconstruction only, the BP algorithm. Experiments consist in the reconstruction of *ensembles* of phantom images in different noise and measurements regimes as well as the reconstruction of real tomographic images. The noise distribution is considered to be known for QP, TV, and BP (since there is no clear strategy to estimate it within the algorithms), while, for the implementations of EP, we estimate β as described in Sec. III. A more detailed description of TV, QP, and BP is reported in Appendix D.

In the following we describe the measurement process we used and we comment our reconstruction performances for both synthetic and real images.

A. Acquisition process

The measurement matrices $\boldsymbol{A}$ are constructed mimicking a realistic tomographic acquisition process. When using the BP algorithm, i.e., when reconstructing binary images, the projection matrix is the result of the projections of several parallel rays along a single direction, and this is repeated for angles between 0° and 180° in regular steps. For all the other cases (nonbinary phantoms and real tomographic images), $\boldsymbol{A}$ was built using single ray projections along random directions. Importantly, for nonbinary or real images the entries of the projection matrix correspond to the length of the ray passing through that pixel, while for binary images its entries are 1 or 0, depending on whether a ray passed or not through the associated pixel. This latter constraint has been introduced to use fairly the implementation of the BP algorithm in Ref. [5].

To simulate the noisy regime we add a random variable with Gaussian distribution $\mathcal{N}(0, \sigma)$ to each component of the measurement vector $\boldsymbol{p}$, with $\sigma = \beta^{-1/2}$.

B. Results for phantom images

Synthetic phantoms represent light patches, or clusters, in a circular black background (as in the inset of Fig. 2) that are generated as follows: Starting from a black colored image of dimension $L \times L$, we color uniformly at random p^2 pixels of "white" that we will be used as centroids of a Gaussian filter of width $\sim 1/p$. For binary tomography only, we binarize the resulting images (further details in the generation of phantom images are provided in Appendix E). By tuning p we can control the structure and complexity of the phantoms since it has been found empirically that the number of pixels in the boundary between the light regions and the background, and thus the dimension of the patches, scales roughly linearly with p [5]. For each chosen value of p, we generate a set of images differing in the choice of the seeds.

Finally, to quantify the accuracy of the reconstructed image $\boldsymbol{x}^\star$ compared to the real one $\boldsymbol{x}$, we introduce two metrics: For binary images, we use the number of wrongly assigned pixel N_e, while for nonbinary images, we estimate the average ℓ_2 norm of the difference between the original image and the reconstructed one, that is, $E_2 = ||\boldsymbol{x} - \boldsymbol{x}^*||_2/N$, where N is the number of the pixels within the circular regions. We will say that we have achieved a perfect reconstruction for binary images whenever $N_e = 0$, while for nonbinary ones we have

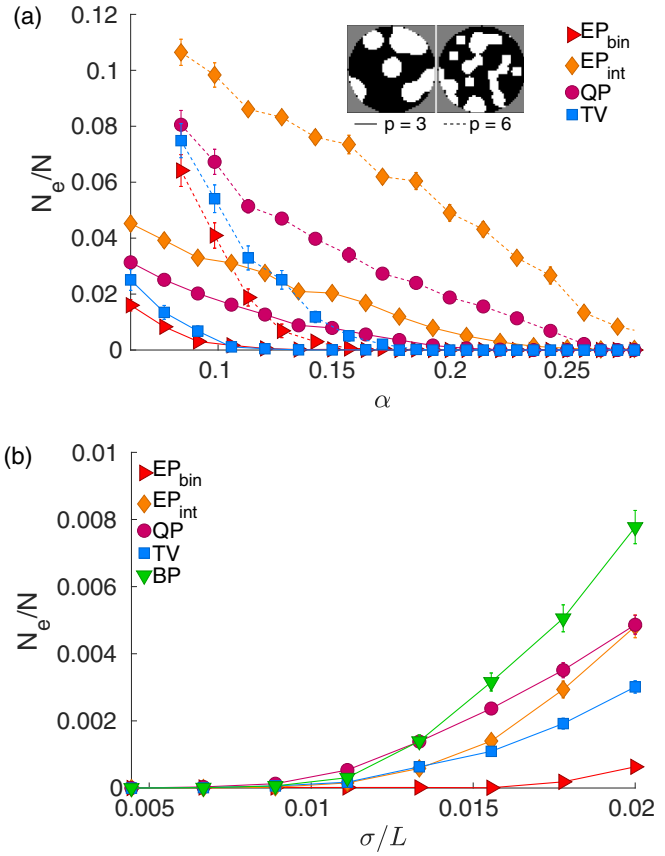


FIG. 2. (a) Fraction of wrong assigned pixels versus sampling rate α. Dashed (thick) lines are used for $p = 6$ ($p = 3$) phantoms while different symbols denote different reconstruction techniques. (b) Estimate of the error versus noise-to-signal ratio σ/L with σ the standard deviation for the Gaussian noise for binary images. In both panels, the points correspond to sample averages over 50 randomly generated phantoms, while the error bars are one standard deviation from the average.

found that errors E_2 significantly smaller than 10^{-4} do not improve the quality of the reconstructed image, in the sense they are indiscernible to the naked eye. All images used for these experiments have sizes $L = 50$ [23], while the values of p will be specified case by case.

Note that we have restricted our analysis to rather small images since EP running time is dominated by a matrix inversion per iteration that requires $\mathcal{O}(N^3)$ operations when using only priors over pixels (see Appendix A for details) and, only for EP with difference prior, $\mathcal{O}[(N+E)^3]$ operations when dealing with E first nearest-neighbors intensity differences. Two possible strategies for a direct application of these methods to real tomography where higher-resolution images are generally required can be suggested. First, an increasing-resolution iterative approach could be implemented, in which (cheaper) reconstruction at smaller resolutions is used as initial state of the iterative EP algorithm at larger ones (so that only a few iterations are needed at the largest resolution). Additionally, one may use GPUs to perform matrix inversion using parallel programming. Either way, the sizes considered here are large enough to draw strong conclusions of the EP algorithm when compared the the current state-of-the-art approaches.

The results for binary reconstruction obtained with EP, TV, and QP are shown in Fig. 2. The subscript `bin` (respectively, `int`) refers to the use of the binary (respectively, interval) prior in the posterior distribution. The top panel depicts the dependence of the fraction of errors N_e/N on the sampling rate $\alpha \equiv M/N$. The inset shows a typical realization of a synthetic random phantom generated using $p = 3$ and $p = 6$. The reconstruction error is non-negligible up to a certain value of α above which perfect reconstruction is reached. Recall that the smaller the value of α, the less number of measurements M we need to achieve a good performance. The lower panel shows the fraction of errors as a function of the noise-to-signal ratio σ/L for a fixed value of $\alpha = 0.255$ and $p = 6$. Notice that for this value of α the reconstruction error is zero for all the methods in the noiseless case (for σ very small) and then increases for non-negligible value of the noise. As we can see, the EP algorithm always achieves a lower error fraction when using a binary prior as it outperforms any other reconstruction method in the noiseless and noisy scenarios.

For nonbinary images, we first tested our method using the Shepp-Logan phantom [24], a well-known benchmark synthetic image representing a 2D section of a simplified human head. Experiments are performed in a noise-free setup and for $L = 80$. In the implementation of the EP algorithm applied to this image we have tried out three different priors, namely the interval, the sparse, and the difference priors, given by Eqs. (3), (4), and (7), respectively. The upper panel of Fig. 3 shows the reconstruction error of all the algorithms under study as a function of the sampling rate α. According to our findings, EP with the difference prior reaches this threshold error and thus achieves a perfect reconstruction, for α in the interval $(0.18, 0.20)$, considerably before the other algorithms. To further benchmark the versatility of EP algorithm, we have studied a noisy case in which we apply EP, QP, and TV to an ensemble of synthetic nonbinary images for $p = 6$. The lower panel of Fig. 3 depicts the behavior of E_2, averaged over 50 synthetic images of $N = 1959$ pixels, as a function of the noise-to-signal ratio. As we can see from this figure, EP with a difference variables prior outperforms the other algorithms for moderate values of noise (low values of σ) and then perform very similarly to TV (that computes the best reconstructed images) for larger value of the noise. A similar behavior is also found as a function of σ/L for smaller values of α. It is important to keep in mind, however, that when using TV and QP the noise distribution is assumed to be known, while in the EP approach it is an additional parameter to be inferred. This means that EP with the difference prior performs closely to TV even when less information about the measurement setup is available.

C. Results for real tomographic images

We report here the results of QP, TV, and the two implementations of EP algorithm for nonbinary pixels (with *interval* and ℓ_0 smoothness priors) on the reconstruction of four real computed tomography (CT) scans: a mouse's head and three images of a human head differing in the acquisition plane. The index i associated with each CT_head_i image

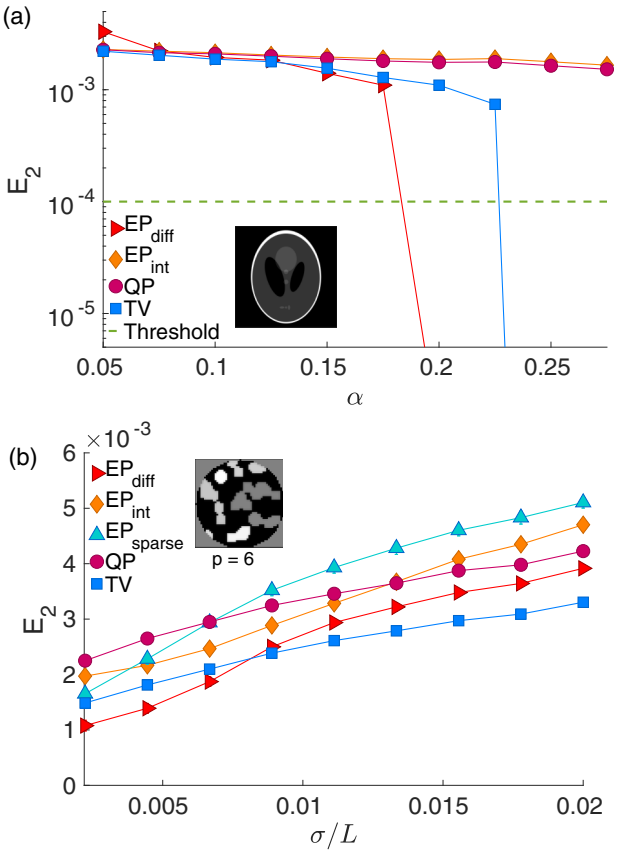


FIG. 3. (a) E_2 error for continuous images as a function of the sampling rate α for the Shepp-Logan phantom. The dashed line corresponds to error-free reconstruction threshold. (b) Reconstruction error versus noise-to-signal ratio σ/L using synthetic random phantoms for fixed $\alpha = 0.745$ and $p = 6$ (an example of phantom image is depicted in the inset).

refers the position of the scanner with respect to the neck of the patient (the smaller the index i the closer to the neck). For these experiments, the original high resolution images are rescaled to a smaller size of 100×100 pixels and the measuring process has been simulated by our acquisition algorithm for nonbinary images in the noiseless regime.

In Fig. 4 we plot the reconstruction error E_2 as a function of $\alpha \in [0.1, 0.6]$ for the following cases: (a) the mouse's skull, (b) CT_head_38, (c) CT_head_80, and (d) CT_head_100. The original images are reported in the same ordering in Fig. 5. Rather remarkably, contrary to what we observed in the Shepp-Logan phantom, here the E_2 error of all algorithms decreases rather smoothly as we include more measurements. However, the transition to a perfect reconstruction regime is reached only by TV and EP with difference prior. This clearly indicates that including pairwise interactions in the prior is certainly advantageous when dealing with tomographic images. It is worth noting that perfect reconstruction is reached by EP with the ℓ_0 smoothness prior at $\alpha = 0.42$ for $E_2 = 1.9 \times 10^{-4}$, while the TV algorithm needs more measurements to achieve the same result, namely $\alpha = 0.53$ for $E_2 = 2.1 \times 10^{-4}$. Hence, once again, EP surpasses the other algorithms. This is illustrated in Fig. 6, where we show several reconstructions of CT_head_100 as a function of α. The

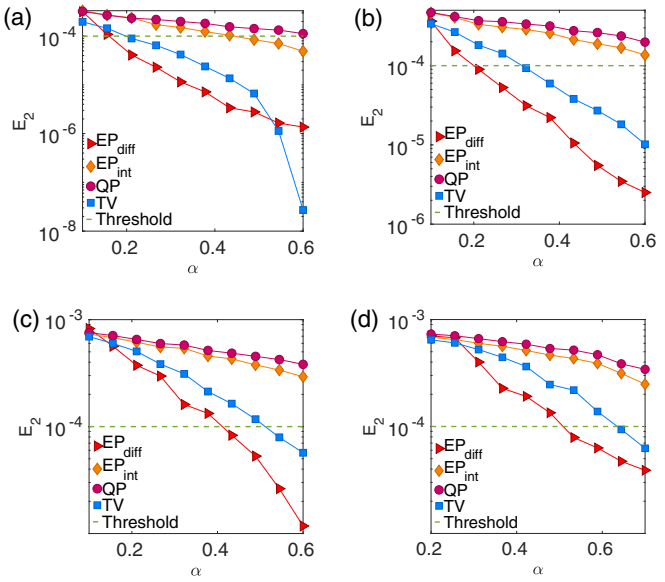


FIG. 4. E_2 error for real tomographic images as a function of the sampling rate α for (a) the mouse's head, (b) CT_head_38, (c) CT_head_80, and (d) CT_head_100.

red line marks the boundary between perfectly reconstructed images (right region) and less accurate reconstructions (left region). TV and EP with the ℓ_0 smoothness prior reach the perfect reconstruction at $\alpha = 0.53$ and $\alpha = 0.42$, respectively, while EP with interval prior and QP need more measurements (that is, larger values of α) to achieve an errorless reconstruction.

D. Phase-type diagram of perfect and imperfect reconstruction

To characterize the performances of all implementations of EP we show here the perfect and imperfect reconstruction diagrams in the (α, β) plane. The tested images are the ones

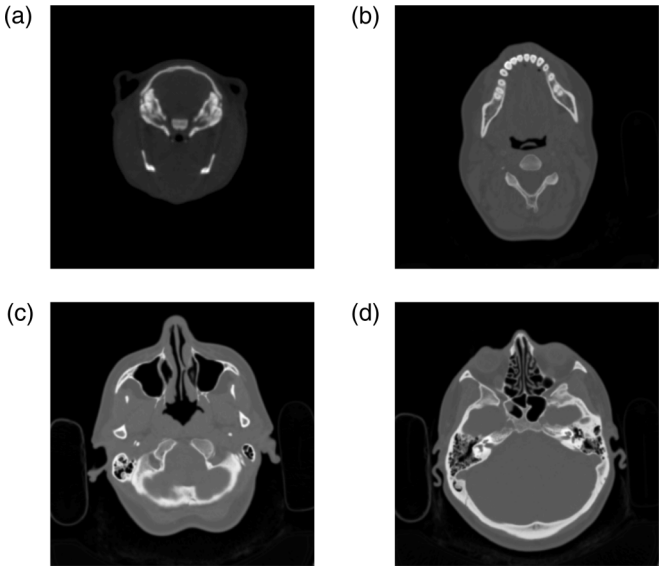


FIG. 5. Full resolution images for (a) the mouse's head, (b) CT_head_38, (c) CT_head_80, and (d) CT_head_100.

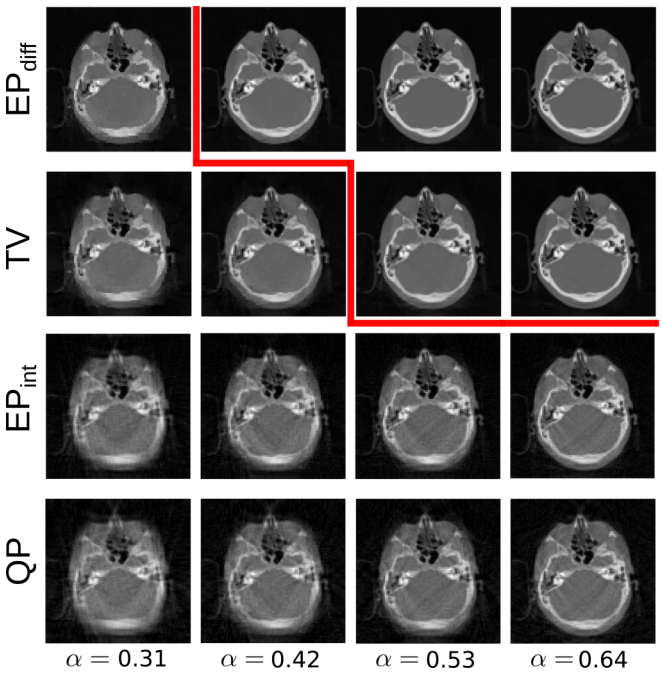


FIG. 6. Table containing the reconstructions of CT_head_100 using EP_{diff}, TV, EP_{int}, and QP as a function of the sampling rate α. Top right region, delimited by the red thick line, contains the perfectly reconstructed images.

used for the lower panels of Figures 2 and 3. Results are shown in Fig. 7 for both, binary (upper panel) and nonbinary (lower panel) reconstructions. As we can see from the plots, the implementations of EP that reach the best performances (lower reconstruction errors) for both binary and gray-scale images, are the ones with the binary prior and the difference prior respectively. Intuitively, this confirms that inference performance is strongly tied to the closeness of the prior distribution to the correct statistics of the target ensemble of images.

V. DISCUSSION

We have shown how to address the problem of reconstructing tomographic images by including nonstandard prior information about the image, normally resulting in non-log-concave prior weight functions. The reconstruction itself can be performed by the EP algorithm. EP is able to encode, within a Bayesian framework, empirical information about the statistics of the treated variables, using *ad-hoc* prior distributions that are rather difficult or even impossible to cope with standard tools. The results presented here employ prior knowledge both about single pixel and the differences of nearest-neighbor pixel intensities. For sake of simplicity, the prior distribution over these auxiliary variables does not take into account the spatial localization of the pixels but, in principle, EP can treat even this more specific case. For instance, one can exploit a collection of a certain class of tomographic images as a training set for the statistics of each difference variable, with the resulting histograms then encoded as prior distributions for future inferences. Notice that the treatment of difference variables is possible within the EP framework because it involves a linear transformation of the pixel intensities. From a more general perspective, one can think of extending the same formalism to any linear transformation of the pixel variables.

We have compared the performances of EP to the ones of standard convex optimization techniques and, only in the case of binary tomography, to the BP algorithm. Binary tomography results show that, using a binary prior, a perfect

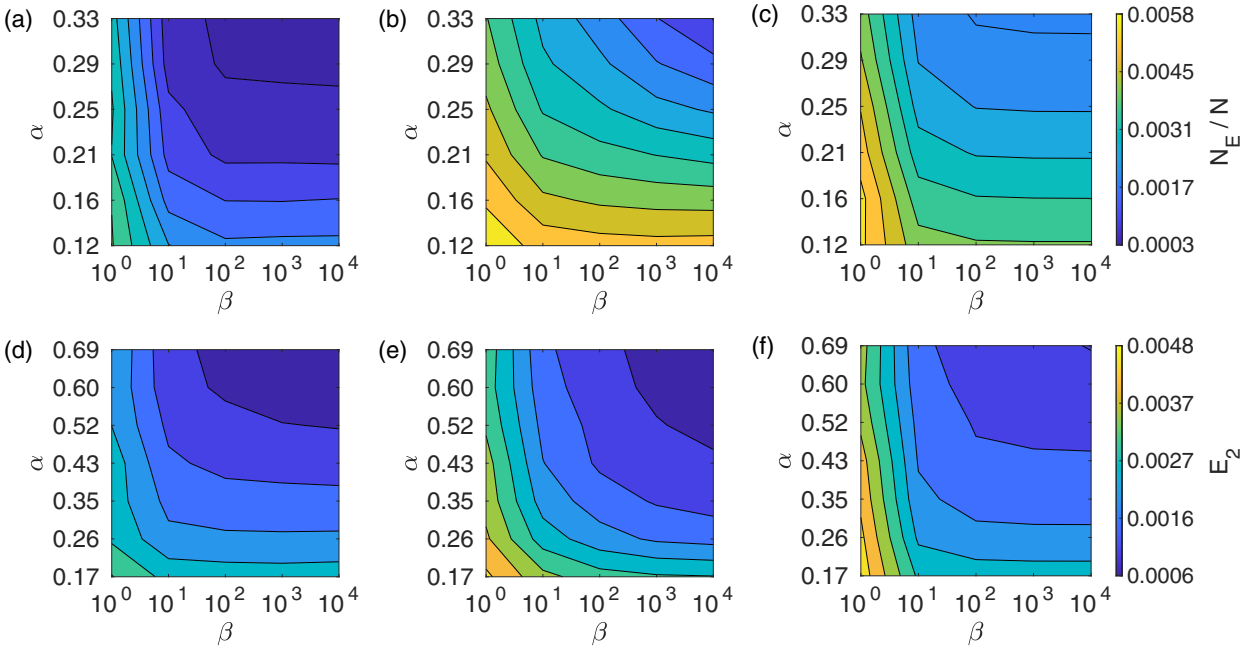


FIG. 7. Reconstruction error in the (α, β) plane for all our reconstruction techniques differing in the priors: (a) EP_{bin}, (b) EP_{sparse}, and (c) EP_{int} panels for binary images and (d) EP_{diff}, (e) EP_{sparse}, and (f) EP_{int} for continuous images.

reconstruction is possible for synthetic images even within the limited-data regime, outperforming any other algorithms adopted here. In the case of nonbinary images, EP performs remarkably well when using the difference prior, carrying a clear improvement in the inference when compared to EP implementations with interval and sparse prior. This suggests that when dealing with a more specific type of reconstruction problem, a drastic improvement can be attained by employing a prior that describes well the specific subclass of target images. With respect to other techniques, EP reconstructions present more accurate reconstruction with respect to TV ones in the case of the Shepp-Logan phantom, real tomographic images and for synthetic images affected by noise.

Not only is EP able to approximate well the posterior distributions of the models presented here but it also provides a powerful tool to estimate the parameters of the model and to have access to the properties of the noise distribution affecting the data that are, in real-case scenarios, unknown. We remark that, in contrast to EP, neither TV nor QP are able to infer the parameters of the optimization and thus several runs are needed before reaching the best reconstruction. Moreover, the model presented here deals with additive noise but the multiplicative noise regime can be faced by EP using a slightly different formulation of the likelihood.

ACKNOWLEDGMENTS

I.P.C. is grateful to Project No. DGAPA-UNAM-PAPIIT IA103417 and thanks HuGeF and Polito for their hospitality. A.P., A.B., and A.P.M. acknowledge funding from INFERNET, a European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant No. 734439. R.D.H.R. is grateful to CONACYT for financial support. I.P.C. and R.D.H.R. thank Arnulfo Martinez for interesting discussions and clarifications about image reconstruction on medical images. The medical images used in the present work were generated during the research carried out under Grants No. UNAM-DGAPA-PAPIIT IN108615 and No. IN110616, as well as CONACYT Problemas Nacionales 2015-01-612.

APPENDIX A: FAST COMPUTATION OF THE UPDATE EQUATIONS

The aforementioned moment matching conditions, which appear in any implementation of the EP algorithm, require inverting $\boldsymbol{\Sigma}_{(i)}^{-1}$, since a_i and b_i depend explicitly on the ith diagonal element of $\boldsymbol{\Sigma}$. On top of that, a direct implementation of the procedure described so far would involve performing this inversion for *each pixel*, thus resulting in an algorithm that scales as $\mathcal{O}(N^4)$ per iteration step. This is a consequence of the fact that inverting an $N \times N$ matrix scales as N^3. However, let us define

$$\begin{aligned} \overline{\boldsymbol{\Sigma}}^{-1} &= \beta \boldsymbol{A}^T \boldsymbol{A} + J\boldsymbol{L} + \boldsymbol{B}, \\ \overline{\boldsymbol{\mu}} &= \overline{\boldsymbol{\Sigma}}(\beta \boldsymbol{A}^T \boldsymbol{p} + \boldsymbol{B}\boldsymbol{a}), \end{aligned} \tag{A1}$$

with $\boldsymbol{B}$ a full diagonal matrix of elements $B_{ii} = b_i^{-1}$. Using these quantities, we can reduce the computational cost of the EP algorithm, since now we are able to compute the necessary elements for applying the moment matching conditions with a *single matrix inversion per iteration step*. After some basic algebra, we notice that we can express Σ_{ii} and μ_i as

$$\begin{aligned} \Sigma_{ii} &= \frac{\overline{\Sigma}_{ii}}{1 - \overline{\Sigma}_{ii}/b_i}, \\ \mu_i &= \frac{\overline{\mu}_i - \frac{a_i}{b_i}\overline{\Sigma}_{ii}}{1 - \overline{\Sigma}_{ii}/b_i}. \end{aligned} \tag{A2}$$

Even though this still requires $\mathcal{O}(N^4)$ operations, we have replaced N matrix inversions per iteration step for N arithmetic operations, as found in Eqs. (A2). As a side effect, we should also consider that this improvement in performance limits us to a *parallel* updating scheme for the values of $\boldsymbol{a}$ and $\boldsymbol{b}$, instead of a sequential one.

APPENDIX B: PRIOR ON DIFFERENCE VARIABLES

Let us introduce a set of difference variables $f_{ij} = x_i - x_j$ for $j \in \partial i$, along with the pixels variables $\boldsymbol{x}$, having a prior distribution as in Eq. (7) on the main text. The joint posterior probability of intensities and differences is written as

$$\begin{aligned} P(\boldsymbol{x}, \boldsymbol{f}|\boldsymbol{p}) = \frac{1}{Z} & e^{-\frac{\beta_1 (A\boldsymbol{x}-\boldsymbol{p})^T (A\boldsymbol{x}-\boldsymbol{p})}{2}} e^{-\frac{\beta_2}{2}\sum_{i\sim j}(x_i - x_j - f_{ij})^2} \\ & \times \prod_i \mathbb{I}_{x_i \in [x_i^{(\mathrm{m})}, x_i^{(\mathrm{M})}]} \prod_{i\sim j} \left[\rho\delta(f_{ij}) + (1-\rho)e^{-\frac{\lambda}{2}f_{ij}^2}\right], \end{aligned} \tag{B1}$$

where $i \sim j$ stands for summing over distinct pairs of neighboring pixels, and we expect to take the limit $\beta_2 \to \infty$. Let E be the number of difference variables and let us introduce the vector $\boldsymbol{t} = \binom{\boldsymbol{x}}{\boldsymbol{f}}$. Further, let us define the following $(M+E) \times (N+E)$ matrix $\boldsymbol{S}_{(M+E)\times(N+E)}$ written in block form:

$$\boldsymbol{S}_{(M+E)\times(N+E)} = \begin{pmatrix} \sqrt{\beta_1}\boldsymbol{A} & \boldsymbol{0}_{M\times E} \\ \sqrt{\beta_1}\boldsymbol{R}_{E\times N} & -\sqrt{\beta_2}\boldsymbol{I}_{E\times E} \end{pmatrix}. \tag{B2}$$

Here $\boldsymbol{R}$ is a matrix whose entries are given by $R_{i\sim j,i} = 1$ and $R_{i\sim j,j} = -1$. Then the posterior can be rewritten as:

$$\begin{aligned} P(\boldsymbol{t}|\boldsymbol{p}) \propto & e^{-\frac{1}{2}(\boldsymbol{S}\boldsymbol{t}-\tilde{\boldsymbol{p}})^T(\boldsymbol{S}\boldsymbol{t}-\tilde{\boldsymbol{p}})} \prod_{i=1}^{N} \mathbb{I}_{t_i \in \left[x_i^{(\mathrm{m})}, x_i^{(\mathrm{M})}\right]} \\ & \times \prod_{i=N+1}^{N+E} \left[\rho\delta(t_i) + (1-\rho)e^{-\frac{\lambda}{2}t_i^2}\right], \end{aligned} \tag{B3}$$

with $\tilde{\boldsymbol{p}} = \binom{\sqrt{\beta_1}\boldsymbol{y}}{\boldsymbol{0}}^T$. According to the EP approximation scheme, we approximate each single-variable non-Gaussian prior via $\mathcal{N}(a_i, b_i)$ whose parameters are determined through the set of Eqs. (18). Notice that it depends on the form of the tilted distribution that, for this choice of priors, reads

$$\begin{aligned} Q^{(i)}(\boldsymbol{t}|\boldsymbol{p}) = & \frac{1}{Z_Q^{(i)}} e^{-\frac{1}{2}(\boldsymbol{t}-\boldsymbol{\mu}^{(i)})^T \cdot \boldsymbol{\Sigma}_{(i)}^{-1} \cdot (\boldsymbol{t}-\boldsymbol{\mu}^{(i)})} \\ & \times \begin{cases} \mathbb{I}_{t_i \in [x_i^{(\mathrm{m})}, x_i^{(\mathrm{M})}]} & i \leqslant N \\ \rho\delta(t_i) + (1-\rho)e^{-\frac{\lambda}{2}t_i^2} & i > N \end{cases}, \end{aligned} \tag{B4}$$

with

$$\boldsymbol{\Sigma}_{(i)}^{-1} = \boldsymbol{S}^T\boldsymbol{S} + \boldsymbol{B}^{(i)}, \quad \boldsymbol{\mu}^{(i)} = \boldsymbol{\Sigma}_{(i)}\left(\boldsymbol{S}^T\tilde{\boldsymbol{p}} + \boldsymbol{B}^{(i)}\boldsymbol{a}\right). \tag{B5}$$

APPENDIX C: ESTIMATION OF THE PARAMETERS OF THE MODEL

As pointed out in Sec. III, Bayesian inference methods allow to naturally estimate the values of the parameters. We explain in detail the expectation maximization algorithm used for determining J and β. Given the probabilistic model described in the main text for solving the reconstruction problem, we define $P(\boldsymbol{p}|\beta, J)$ as the probability of observing the data $\boldsymbol{p}$. Using the actual measured values of $\boldsymbol{p}$, $P(\boldsymbol{p}|\beta, J)$ defines the likelihood of the parameters β and J,

$$P(\boldsymbol{p}|\beta, J) = \left(\frac{\beta}{2\pi}\right)^{\frac{M}{2}}\left(\frac{J}{2\pi}\right)^{\frac{N}{2}}\int d^N\boldsymbol{x}\, e^{-\frac{\beta}{2}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p})^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p})} - \frac{J}{2}\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x}\prod_{i=1}^{N}\psi_i(x_i), \tag{C1}$$

Hence, we would like to find the values of β and J such that the likelihood above is maximized. However, due to the functional dependence on the parameters, a direct maximization procedure is rather impractical. The EM algorithm provides an alternative by iteratively estimating the optimal value of β and J. Seeking indeed that $\frac{\partial P(\boldsymbol{p}|\beta,J)}{\partial\beta} = \frac{\partial P(\boldsymbol{p}|\beta,J)}{\partial J} = 0$, we obtain

$$\begin{aligned}\beta^\star &= \frac{M}{\left\langle(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p})^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p})\right\rangle_{P(\boldsymbol{p},\boldsymbol{x}|\beta,J)}},\\ J^\star &= \frac{N}{\left\langle\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x}\right\rangle_{P(\boldsymbol{p},\boldsymbol{x}|\beta,J)}},\end{aligned} \tag{C2}$$

with

$$\langle(\cdots)\rangle_{P(\boldsymbol{p},\boldsymbol{x}|\beta,J)} = \left(\frac{\beta}{2\pi}\right)^{\frac{M}{2}}\left(\frac{J}{2\pi}\right)^{\frac{N}{2}}\int d^N\boldsymbol{x}(\cdots) \times e^{-\frac{\beta^\star}{2}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p})^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p})-\frac{J^\star}{2}\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x}}\prod_{i=1}^{N}\psi_i(x_i). \tag{C3}$$

This provides a closed set of equations for the pair $(\beta^\star\ J^\star)$, which is solved by the fixed-point iteration method. Such a procedure will yield the same equations that a direct implementation of EM would [22]. On the other hand, the averages appearing in the formulas (C2) are rather slow to calculate (as they involve the computation of all covariances). We found out that an assumption of self-averageness $\boldsymbol{x} = \langle\boldsymbol{x}\rangle_{Q^{\rm EP}}$ is empirically reasonable (as the output distribution is generally well concentrated) and leads to excellent results, rendering the computation both trivial and fast. This finally results in:

$$\begin{aligned}\beta^\star &= \frac{M}{(\boldsymbol{A}\langle\boldsymbol{x}\rangle_{Q^{\rm EP}}-\boldsymbol{p})^T(\boldsymbol{A}\langle\boldsymbol{x}\rangle_{Q^{\rm EP}}-\boldsymbol{p})},\\ J^\star &= \frac{N}{\langle\boldsymbol{x}\rangle_{Q^{\rm EP}}^T\boldsymbol{L}\langle\boldsymbol{x}\rangle_{Q^{\rm EP}}}.\end{aligned} \tag{C4}$$

The direct approach explained here cannot be applied for parameters appearing in the priors, such as the sparsity parameter. In this case the maximization of the likelihood $P(\boldsymbol{p}|s)$ does not yield an explicit update equation for s^*, so we choose to optimize this parameter by gradient ascent. In this case, we can extremize the free energy $F = -\log P(\boldsymbol{p}|s)$, which is typically well approximated by the EP free energy [16]

$$F_{\rm EP} = (N-1)\log Z_Q - \sum_{i=1}^{N}\log Z_{Q^{(i)}}, \tag{C5}$$

where Z_Q is the partition function of the approximating distribution, $Q(\boldsymbol{x}|\boldsymbol{p})$, and $Z_{Q^{(i)}}$ the corresponding one for the tilted distribution $Q^{(i)}(\boldsymbol{x}|\boldsymbol{p})$. Finally, to find the optimal value of s we use the gradient-descent method,

$$s^{(t+1)} = s^{(t)} - \eta\frac{\partial F_{\rm EP}}{\partial s}, \tag{C6}$$

with $\eta > 0$ a relaxation parameter.

APPENDIX D: OTHER METHODS FOR RECONSTRUCTION

In this section we briefly review the other reconstruction algorithms we have compared our results to.

1. Quadratic programming

The reconstruction problem, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{p}$, with the prior information about $\boldsymbol{x}$ coming from the Laplacian matrix and in the noiseless scenario, can be recast as a constrained quadratic minimization problem:

$$\boldsymbol{x}^* = \underset{\substack{\boldsymbol{x}\,:\,\boldsymbol{A}\boldsymbol{x}=\boldsymbol{p}\\ \boldsymbol{x}_{\rm inf}\leqslant\boldsymbol{x}\leqslant\boldsymbol{x}_{\rm sup}}}{\arg\min}\ \boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x}. \tag{D1}$$

Here we have used $\boldsymbol{x}_{\rm inf}$ and $\boldsymbol{x}_{\rm sup}$ to denote the lower and upper limits for the pixel values, to mimic the constraint of the interval prior. When dealing with binary images, the pixels of the above solution, whose value are larger than 0.5 are set to 1 or to 0 otherwise. In the noisy regime, we instead assume to know $\sigma = \beta^{-1/2}$ and we minimize instead:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}_{\rm inf}\leqslant\boldsymbol{x}\leqslant\boldsymbol{x}_{\rm sup}}{\arg\min}\ J\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x} + \beta(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p})^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p}) \tag{D2}$$

for different values of the parameter J. For each trial J we compute the reconstruction error and we keep the smallest one.

2. Total variation

As explained in the main text, we can pose the reconstruction problem as an optimization one whose objective function is the ℓ_2 norm of the error: $||\boldsymbol{A}\boldsymbol{x}-\boldsymbol{p}||_2$. TV is an improvement on this approach by adding the requirement that the solution also minimizes the ℓ_1 norm of the image gradient, $||\nabla_{\rm img}\boldsymbol{x}||_1$, which is defined as

$$(\nabla_{\rm img}\boldsymbol{x})_i = (x_{i_x} - x_i, x_{i_y} - x_i), \tag{D3}$$

where i_x and i_y denote the neighboring pixel to the right and below i, respectively. Hence, the TV optimization problem

reads:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \; ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{p}||_2 + \lambda||\nabla_{\text{img}}\boldsymbol{x}||_1. \qquad \text{(D4)}$$

In this last equation, λ is a parameter to weight the relevance of the image gradient regularization. In the case of noisy measurements we repeat the minimization for different values of the parameter λ and we report the E_2 error from the best reconstruction. Since norms are convex functions and the Laplacian matrix is positive semidefinite, the solution to these two optimization problems can be found using convex optimization techniques. For the cases studied in this work, we used the `Convex.jl` and `Gurobi` optimization packages [25,26] to find the solution $\boldsymbol{x}^*$.

3. Belief propagation

The reconstructions using the belief propagation algorithm were obtained with the implementation referenced in Ref. [5], which can be found in Ref. [27].

APPENDIX E: DETAILS IN THE GENERATION OF THE PHANTOM IMAGES

For binary images we used the procedure described in Ref. [5] and the script provided in [27] to generate a sample of 50 images of size 50×50, whose number of clusters is controlled by an integer parameter p in the following manner: The algorithm generates images within a circle, as shown in the insets of Fig. 2, and therefore the effective number of pixels to be reconstructed is reduced to 1959. Once the value of p is specified, p^2 pixels are chosen randomly as centroids for a Gaussian filter. Once the filter is applied, only the pixels that have a value above the image average value are set to 1, and the rest of them are set to 0. On the other hand, when dealing with gray-scale images, the same procedure is used to generate binary clusters, but once they have been constructed, we set the value of all the pixels within one of them to a random integer inside the interval [105,255]. This is done for each of the formed clusters.

[1] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE Press, Los Alamitos, CA, 1988).

[2] S. Kaczmarz, Approximate solution of systems of linear equations, Int. J. Control **57**, 1269 (1993).

[3] R. Gordon, R. Bender, and G. T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography, J. Theor. Biol. **29**, 471 (1970).

[4] E. Y. Sidky, C.-M. Kao, and X. Pan, Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT, J. X-Ray Sci. Technol. **14**, 119 (2006).

[5] E. Gouillart, F. Krzakala, M. Mezard, and L. Zdeborová, Belief-propagation reconstruction for discrete tomography, Inv. Probl. **29**, 035003 (2013).

[6] C. Bouman and K. Sauer, A generalized Gaussian image model for edge-preserving MAP estimation, IEEE Trans. Image Process. **2**, 296 (1993).

[7] W. T. Freeman, T. R. Jones, and E. C. Pasztor, Example-based super-resolution, IEEE Comput. Graph. Appl. **22**, 56 (2002).

[8] S. Sardy and P. Tseng, On the statistical analysis of smoothing by maximizing dirty markov random field posterior distributions, J. Am. Stat. Assoc. **99**, 191 (2004).

[9] K. Tanaka, M. Yasuda, and D. M. Titterington, Bayesian image modeling by means of a generalized sparse prior and loopy belief propagation, J. Phys. Soc. Jpn. **81**, 114802 (2012).

[10] A. Montanari and T. Rizzo, How to compute loop corrections to the Bethe approximation, J. Stat. Mech.: Theory Exp. (2005) P10011.

[11] A. Pelizzola, Cluster variation method in statistical physics and probabilistic graphical models, J. Phys. A: Math. Gen. **38**, R309 (2005).

[12] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, MAP estimation via agreement on trees: message-passing and linear programming, IEEE Trans. Inf. Theory **51**, 3697 (2005).

[13] G. Parisi and F. Slanina, Loop expansion around the Bethe-Peierls approximation for lattice models, J. Stat. Mech.: Theory Exp. (2006) L02003.

[14] M. Opper and O. Winther, Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling, Phys. Rev. E **64**, 056131 (2001).

[15] T. P. Minka, Expectation propagation for approximate Bayesian inference, in *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence* (Morgan Kaufmann Publishers Inc., San Francisco, CA, 2001), pp. 362–369.

[16] M. Opper and O. Winther, Expectation consistent free energies for approximate inference, in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, Massachusetts, 2004), pp. 1001–1008.

[17] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, Approximate inference techniques with expectation constraints, J. Stat. Mech.: Theory Exp. (2005) P11015.

[18] E. Y. Sidky, R. Chartrand, and X. Pan, Image reconstruction from few views by non-convex optimization, in *Proceedings of the 2007 IEEE Nuclear Science Symposium Conference Record*, Vol. 5 (IEEE, Los Alamitos, CA, 2007), pp. 3526–3530.

[19] H. Ishwaran and J. S. Rao, Spike and slab variable selection: Frequentist and Bayesian strategies, Ann. Stat. **33**, 730 (2005).

[20] G. T. Herman and A. Kuba, *Discrete Tomography: Foundations, Algorithms, and Applications* (Springer Science & Business Media, New York, 2012).

[21] K. J. Batenburg and J. Sijbers, DART: A practical reconstruction algorithm for discrete tomography, IEEE Trans. Image Process. **20**, 2542 (2011).

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc: Ser. B **39**, 1 (1977).

[23] Note that the number of pixels to reconstruct, N, is not L^2, since only the values inside the circular region are to be inferred, thus leaving $N = 1959$.

[24] L. A. Shepp and B. F. Logan, The Fourier reconstruction of a head section, IEEE Trans. Nucl. Sci. **21**, 21–43 (1974).

[25] M. Udell, K. Mohan, D. Zeng, J. Hong, S. Diamond, and S. Boyd, Convex optimization in Julia, in *Proceedings of the 2014 First Workshop for High Performance Technical Computing in Dynamic Language* (IEEE, New Orleans, LA, 2014), pp. 18–28.

[26] L. L. C. Gurobi Optimization, Gurobi Optimizer Reference Manual (2016), https://www.gurobi.com.

[27] Belief-propagation for binary tomography reconstruction (2013), https://github.com/eddam/bp-for-tomo.