Tomographic Reconstruction Using a New Voxel-Domain Prior and Gaussian Message Passing

Nour Zalmai, Clément Luneau, Carina Stritt*, and Hans-Andrea Loeliger

ETH Zurich, Dept. of Information Technology & Electrical Engineering
*EMPA, Dübendorf, Switzerland, Reliability Science & Technology Laboratory
{zalmai, loeliger}@isi.ee.ethz.ch, cluneau@student.ethz.ch, carina.stritt@empa.ch

***Abstract*—The paper proposes a new prior model for gray-scale images in 2D and 3D, and a pertinent algorithm for tomographic image reconstruction. Using ideas from sparse Bayesian learning, the proposed prior is a Markov random field with individual unknown variances on each edge, which allows for sharp edges. Such a prior model remarkably captures and preserves both the edge structures and continuous regions of natural images while being computationally attractive. The proposed reconstruction algorithm is an efficient EM (expectation maximization) algorithm where the actual computations essentially reduce to scalar Gaussian message passing. Simulation results show that the proposed approach works well even with few projections, and it yields (slightly) better results than a state-of-the art method.**

I. INTRODUCTION

Tomographic reconstruction—the estimation of a 2D or 3D image from noisy projections—has been a challenge for decades due to both the 2D (or 3D) nature of the problem and the huge dimensions of the corresponding mathematical objects. As of today, the filtered back projection (FBP) method [1] is still the most widely used reconstruction algorithm in real-world applications, mainly because of the high computational complexity of more advanced methods. However, FBP has many drawbacks including a blurred reconstruction, the requirement of a large number of uniformly-distributed projections, and a high sensitivity to noise in the projections. In particular, FBP with linear pre- and/or post-filtering is intrinsically ill-suited to reconstruct sharp edges in the image.

In order to cope with sharp edges, a number of more advanced methods have been proposed, such as [2]–[7] and many others. Some of these methods exploit sparsity in the gradient [4], [5] while others exploit sparsity in a wavelet transform of the image [6]. In particular, regularization involving the image total variation greatly improves on FBP and can reconstruct sharp edges [4], [5]. In addition, these techniques can deal with a small number of projections, and the projections need not be uniformly distributed. These methods often use gradient-based optimization algorithms of reasonable complexity.

Several authors have advocated a Bayesian approach with a prior in form of a Markov random field (MRF) [7]–[9]. Inference in such models has often been carried out by Markov chain Monte Carlo techniques that are too slow for most practical applications. However, in very recent work, fast Gaussian message passing methods have successfully been demonstrated [7].

In this paper, we propose a new version of an (improper) MRF prior with sparsifying NUV terms (normal with unknown variance). NUV priors are a central idea from sparse Bayesian learning and automatic relevance determination as in [10]–[13], see also [14]. The proposed prior is similar to the prior in [9] (cf. the remark at the end of Section II-C); it is effectively (but not mathematically) similar also to the spike-and-slab prior of [7].

We also propose an efficient reconstruction algorithm based on expectation maximization (EM) (cf. Section III) to estimate the unknown variances and iterative scalar Gaussian message passing (cf. Section IV). The actual computations are somewhat similar to those in [7]. In our simulations (cf. Section V), we obtain better reconstruction results than with the state-of-the-art method from [4].

II. STATISTICAL MODEL

Each voxel (or pixel) is associated with a real random variable X_{s_ℓ}, $\ell \in \{1, \dots, L\}$ with L the total number of voxels and $s_\ell = (i, j, k)$ the spatial index of that voxel (in 2D, omit the index k). For simplicity, we assume that each dimension of the mesh has the same number of units ($L^{\frac{1}{3}}$ in 3D or $L^{\frac{1}{2}}$ in 2D). We denote by $\mathbf{X} = (X_{s_1}, \dots, X_{s_L})$ the random vector in $\mathbb{R}^L$. We define an order among the spatial indices: for all spatial indices $s_\ell = (i, j, k)$ and $s_{\ell'} = (i', j', k')$

$$\begin{aligned} s_\ell < s_{\ell'} \Leftrightarrow (i < i') \text{ or } ((i = i') \;\&\; (j < j')) \\ \text{or } ((i = i') \;\&\; (j = j') \;\&\; (k < k')) \; . \end{aligned} \quad (1)$$

We say that two voxels X_{s_ℓ} and $X_{s_{\ell'}}$ (with $s_\ell = (i, j, k)$ and $s_{\ell'} = (i', j', k')$) are neighbors if and only if

$$|i - i'| + |j - j'| + |k - k'| = 1 \; . \quad (2)$$

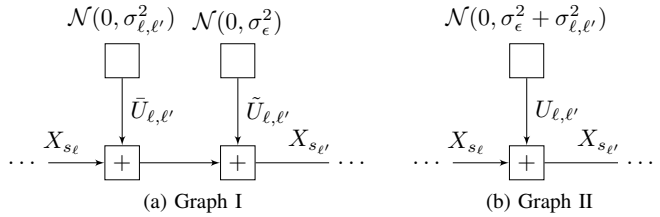
Let Δ denote the set of index pairs (ℓ, ℓ') such that the voxels indexed by s_ℓ and $s_{\ell'}$ are neighbors and $s_\ell < s_{\ell'}$.

In the following, we use factor graphs as in [15], [16] to represent the statistical model and (later on) to describe Gaussian message passing computations.

A. Measurement Model

Let $Q \in \mathbb{N}$ denote the number of pixels of the detector. Each measurement $\mathbf{y}_m \in \mathbb{R}^Q$, $m \in \{1, \dots, M\}$, is obtained by projecting the 3D object $\mathbf{x} \in \mathbb{R}^L$ onto the detector and is characterized by the (sparse) projection matrix $A_m \in \mathbb{R}^{Q \times L}$

978-0-9928-6265-7/16/$31.00 ©2016 IEEE



(a) Graph I (b) Graph II

Fig. 1. Factor graph representations of the prior model between neighboring voxels X_{s_ℓ} and $X_{s_{\ell'}}$, $(\ell, \ell') \in \Delta$.

such that $\mathbf{y}_m = A_m \mathbf{x}$. The projection matrices A_m are obtained from a projection method such as the distance-driven projection [17].

Assuming white Gaussian measurement noise of variance σ_Z^2 and denoting $N = Q \cdot M$, the likelihood of the data is

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma_Z^2)^{\frac{N}{2}}} \exp\left(-\frac{\|\mathbf{y} - A\mathbf{x}\|^2}{2\sigma_Z^2}\right) , \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^N$ contains the stacked measurements $\mathbf{y}_m$ and $A \in \mathbb{R}^{N \times L}$ contains the stacked projection matrices A_m.

B. Prior Model in 1D

For ease of exposition we first describe the proposed prior in 1D, i.e., for a discrete-time signal $X_1, X_2, \ldots \in \mathbb{R}$. In this case, our model is

$$X_{k+1} = X_k + U_k \tag{4}$$

$$U_k = \tilde{U}_k + \bar{U}_k \sim \mathcal{N}(0, \sigma_\epsilon^2 + \sigma_k^2) , \tag{5}$$

with independent $\tilde{U}_k \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\bar{U}_k \sim \mathcal{N}(0, \sigma_k^2)$, where σ_ϵ^2 is fixed and known, but $\sigma_1^2, \sigma_2^2, \ldots$ are unknown and estimated by maximum-likelihood. The first term, $\tilde{U}_k \sim \mathcal{N}(0, \sigma_\epsilon^2)$, is a random-walk model that favors similar values for neighboring variables; the NUV term $\bar{U}_k$ is sparse and allows for occasional jumps of arbitrary magnitude [14], [18].

C. Prior Model in 2D and 3D

Generalizing the 1D model of Section II-B to 2D and 3D yields

$$X_{s_{\ell'}} = X_{s_\ell} + U_{\ell,\ell'} \tag{6}$$

for each $(\ell, \ell') \in \Delta$ (cf. Fig. 1 and Fig. 2), and we have

$$\tilde{p}(\mathbf{x}; \boldsymbol{\sigma}^2) = \prod_{(\ell,\ell') \in \Delta} \frac{1}{\sqrt{2\pi(\sigma_\epsilon^2 + \sigma_{\ell,\ell'}^2)}} \exp\left(-\frac{(x_{s_{\ell'}} - x_{s_\ell})^2}{2(\sigma_\epsilon^2 + \sigma_{\ell,\ell'}^2)}\right) . \tag{7}$$

The linear transform from $\mathbf{X}$ to $\mathbf{U} \in \mathbb{R}^B$ can be written as

$$\mathbf{U} = D\mathbf{X} , \tag{8}$$

with $D \in \mathbb{R}^{B \times L}$ and $B = 2L - 2L^{\frac{1}{2}}$ in 2D or $B = 3L - 3L^{\frac{2}{3}}$ in 3D. Unlike the 1D model, this transform is no longer surjective in 2D and 3D.

We use $\tilde{p}(\mathbf{x}; \boldsymbol{\sigma}^2)$ as prior model on $\mathbf{X}$, which is (in 2D and 3D) not a probability density function but only a measure. This prior model is similar in spirit to the one of Curriero & Lele in [19] and is based on the composite likelihood of Lindsay in [20]. It can be interpreted as

$$\tilde{p}(\mathbf{x}; \boldsymbol{\sigma}^2) = \int \delta(\mathbf{u} - D\mathbf{x}) p(\mathbf{u}|\boldsymbol{\sigma}^2) \, \mathrm{d}\mathbf{u} , \tag{9}$$

by pretending that all the random variables $U_{\ell,\ell'}$ are independent of each other a priori, i.e.,

$$p(\mathbf{u}|\boldsymbol{\sigma}^2) = \prod_{(\ell,\ell') \in \Delta} \mathcal{N}(u_{\ell,\ell'} : 0, \sigma_\epsilon^2 + \sigma_{\ell,\ell'}^2) . \tag{10}$$

However, the term $\delta(\mathbf{u} - D\mathbf{x})$ in (9) projects the random vector $\mathbf{U}$ onto a subspace of linear constraints, which creates dependencies among the variables $U_{\ell,\ell'}$. Graphically, those dependencies introduce cycles as illustrated in Fig. 2.

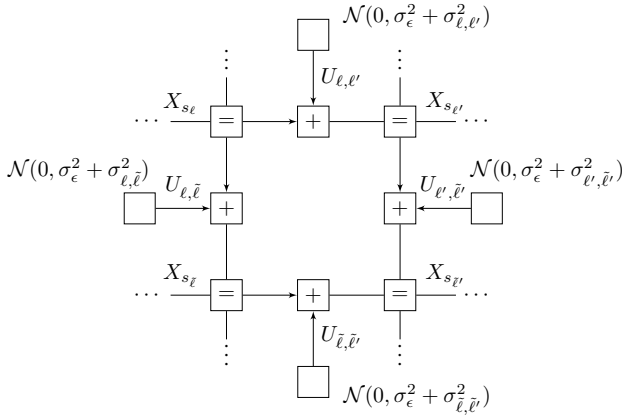


Fig. 2. Factor graph representation of a cycle in the prior model (in 2D or 3D) with $(\ell, \ell') \in \Delta$, $(\ell, \tilde{\ell}) \in \Delta$, $(\ell', \tilde{\ell}') \in \Delta$ and $(\tilde{\ell}, \tilde{\ell}') \in \Delta$.

This prior model captures a basic property of natural structures: two neighboring voxel values tend to be either approximately equal (e.g., when the voxels belong to the same material) or else they may be completely different (e.g., when the voxels belong to different materials). Furthermore, for natural structures, the number of abrupt changes should be substantially smaller than the number of smooth transitions.

A similar prior was used in [9], but with an extra prior on $\boldsymbol{\sigma}^2$ and, more importantly, without the random-walk terms. In addition, we use a different estimation approach (and a different algorithm), as described below.

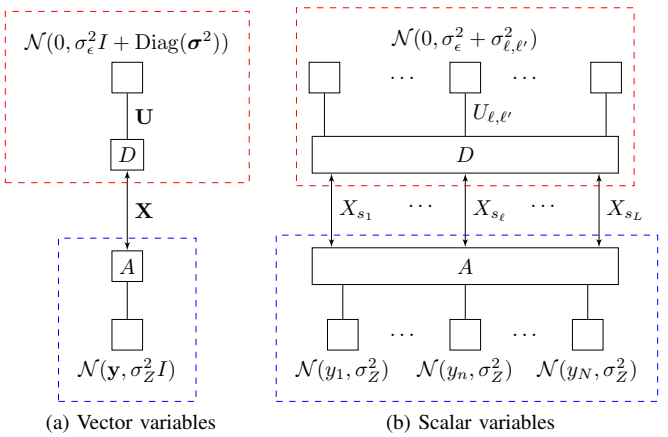
D. Maximum "Likelihood" Estimation

Inference is done by maximizing the "likelihood" [21]

$$\tilde{p}(\mathbf{y}; \boldsymbol{\sigma}^2) = \int p(\mathbf{y}|\mathbf{x}) \tilde{p}(\mathbf{x}; \boldsymbol{\sigma}^2) \, \mathrm{d}\mathbf{x} \tag{11}$$

$$= \int \int p(\mathbf{y}|\mathbf{x}) \delta(\mathbf{u} - D\mathbf{x}) p(\mathbf{u}|\boldsymbol{\sigma}^2) \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{u} . \tag{12}$$

The local maxima of $\tilde{p}(\mathbf{y}; \boldsymbol{\sigma}^2)$ (see Fig. 3a) are at sparse solutions [10], [14]: many $\sigma_{\ell,\ell'}^2$'s are zero, and sharp edges in the image are introduced exactly where $\sigma_{\ell,\ell'}^2 \neq 0$. Note that the use of $\tilde{p}(\mathbf{x}; \boldsymbol{\sigma}^2)$ instead of $p(\mathbf{x}|\boldsymbol{\sigma}^2)$ in (11) does not affect this sparsity behavior and makes the adaptation of the proofs



(a) Vector variables (b) Scalar variables

Fig. 3. Factor graph representation of $p(\mathbf{y}|\mathbf{x})\delta(\mathbf{u}-D\mathbf{x})p(\mathbf{u}|\boldsymbol{\sigma}^2)$. Top dashed box (red): $\tilde{p}(\mathbf{x};\sigma^2)$. Bottom dashed box (blue): $p(\mathbf{y}|\mathbf{x})$

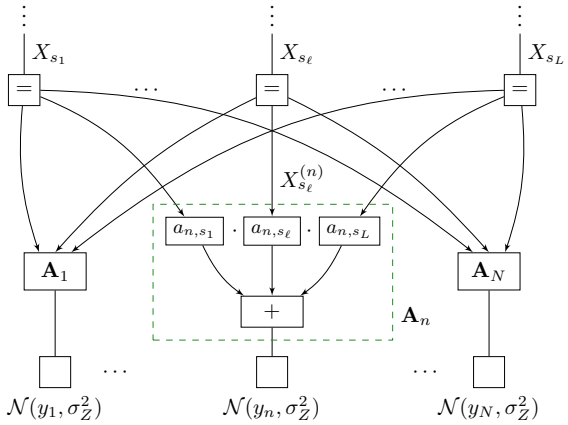


Fig. 4. Detailed factor graph representation of the multiplication by the matrix A. $\mathbf{A}_n$ denotes the n^{th} row of A and a_{n,s_ℓ} the element (n, s_ℓ) of A.

in [10], [14] straightforward. As the maximization of $\tilde{p}(\mathbf{y};\boldsymbol{\sigma}^2)$ cannot be done in closed form, we derive an EM algorithm in the next section.

III. EM ALGORITHM

Considering $\mathbf{U}$ as hidden variable and starting from an initial guess $\hat{\boldsymbol{\sigma}}^2$, the EM algorithm consists in iteratively updating the parameters according to

$$\hat{\boldsymbol{\sigma}}^2 = \underset{\boldsymbol{\sigma}^2}{\operatorname{argmax}}\, \mathbb{E}\left[\ln \tilde{p}(\mathbf{y}, \mathbf{U}; \boldsymbol{\sigma}^2)\right] \ , \tag{13}$$

where the expectation is taken with respect to the density $p(\mathbf{u}|\mathbf{y}, \hat{\boldsymbol{\sigma}}^2)$ ($\hat{\boldsymbol{\sigma}}^2$ being the previous estimate). As can be seen from (12), this optimization problem splits for each individual $\sigma_{\ell,\ell'}^2$, $(\ell, \ell') \in \Delta$

$$\hat{\sigma}_{\ell,\ell'}^2 = \underset{\sigma_{\ell,\ell'}^2}{\operatorname{argmax}}\, \mathbb{E}\left[\ln p(U_{\ell,\ell'}|\sigma_{\ell,\ell'}^2)\right] \tag{14}$$

$$= \max\left(0, \mathbb{E}\left[U_{\ell,\ell'}^2\right] - \sigma_\epsilon^2\right) \ . \tag{15}$$

Since $\mathbb{E}[U_{\ell,\ell'}^2] = m_{U_{\ell,\ell'}}^2 + \sigma_{U_{\ell,\ell'}}^2$, the EM algorithm relies on the ability to compute the posterior mean $m_{U_{\ell,\ell'}}$ and variance $\sigma_{U_{\ell,\ell'}}^2$ of $U_{\ell,\ell'}$. These computations are addressed in the next section. Note that this EM update can exactly set $\hat{\sigma}_{\ell,\ell'}^2$ to zero.

IV. GAUSSIAN MESSAGE PASSING ALGORITHM

For fixed $\boldsymbol{\sigma}^2$, we want to compute the posterior mean $m_{U_{\ell,\ell'}}$ and variance $\sigma_{U_{\ell,\ell'}}^2$ of the random variable $U_{\ell,\ell'}$, $(\ell, \ell') \in \Delta$. Looking at the factor graph in Fig. 3a, all random variables are multivariate Gaussian and by elementary probability rules

$$\mathbf{m}_\mathbf{U} = D\left(D^\mathsf{T}(\sigma_\epsilon^2 I + \mathrm{Diag}(\boldsymbol{\sigma}^2))^{-1}D + \frac{A^\mathsf{T}A}{\sigma_Z^2}\right)^{-1} A^\mathsf{T}\frac{\mathbf{y}}{\sigma_Z^2} \tag{16}$$

$$V_\mathbf{U} = D\left(D^\mathsf{T}(\sigma_\epsilon^2 I + \mathrm{Diag}(\boldsymbol{\sigma}^2))^{-1}D + \frac{A^\mathsf{T}A}{\sigma_Z^2}\right)^{-1} D^\mathsf{T} \ . \tag{17}$$

These expressions involve the inversion of an $L \times L$ matrix (where L is the total number of voxels), which is computationally unattractive.

Thus, in order to reduce the computational complexity, we decompose the cycle-free factor graph in Fig. 3a into the factor graph of Fig. 3b, where all variables are scalars. The matrix multiplication by A is further decomposed as shown in Fig. 4. In this scalar factor graph, we use iterative scalar Gaussian message passing, which has very low complexity. Note that both matrices D and A are rather sparse: in 3D, the matrix D has only $2(3L - 3L^{\frac{2}{3}})$ non-zero elements and the matrix A has only $O(N \cdot L^{\frac{1}{3}})$ non-zero elements (since each projection beam crosses $O(L^{\frac{1}{3}})$ voxels).

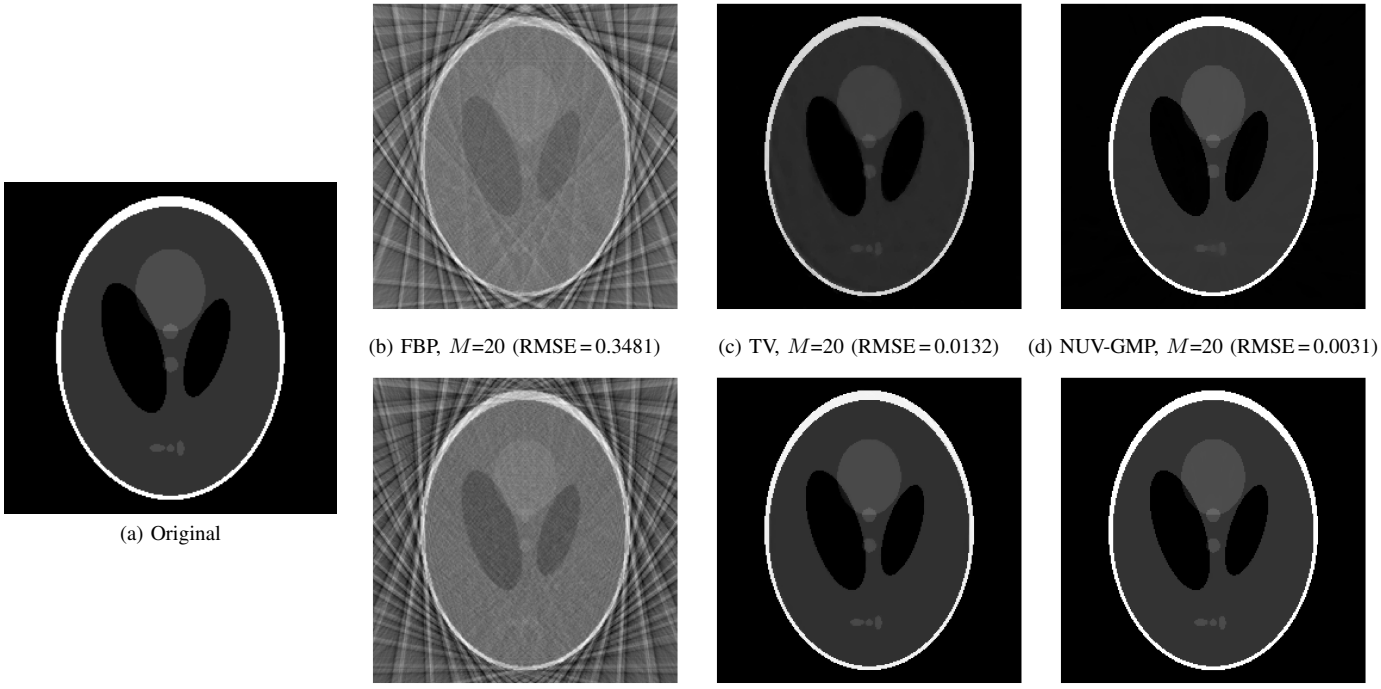
It is well known that iterative message passing in Gaussian graphs with cycles is not guaranteed to converge, but if it converges, then the means are correct [22]–[24]. However, the variances do not usually converge to the correct values, and are of doubtful quality as approximations [22], [23]. In our specific application, non-convergence can (apparently) be avoided by appropriately tuning the parameters σ_Z^2 and σ_ϵ^2. More surprisingly, perhaps, the computed variances are good enough for the EM algorithm to work well.

Each message passing iteration consists in sending a message (i.e., an inverse variance w and a weighted mean $\xi = wm$) along each edge, first from top to bottom in Fig. 3.b (downward messages $\overrightarrow{\xi}$ and $\overrightarrow{w}$) and then from bottom to top (upward messages $\overleftarrow{\xi}$ and $\overleftarrow{w}$). Initially, all messages are set to some values. Then, we treat each projection $(y_n, \mathbf{A}_n)$ sequentially followed by the treatment of the D node at once. $X_{s_\ell}^{(n)}$ denotes the replica of X_{s_ℓ} going towards the measurement y_n. $X_{s_\ell}^{(\ell,\ell')}$ denotes the replica of X_{s_ℓ} going towards $U_{\ell,\ell'}$. Specifically, one message passing iteration works as follows.

1) Process each measurement $(y_n, \mathbf{A}_n)$ sequentially. For all $\ell \in \{1, \ldots, L\}$ such that $a_{n,\ell} \neq 0$, update

$$\overrightarrow{w}_{X_{s_\ell}^{(n)}} := \overrightarrow{w}_{X_{s_\ell}} + \overleftarrow{w}_{X_{s_\ell}} - \overleftarrow{w}_{X_{s_\ell}^{(n)}} \tag{18}$$

$$\overrightarrow{\xi}_{X_{s_\ell}^{(n)}} := \overrightarrow{\xi}_{X_{s_\ell}} + \overleftarrow{\xi}_{X_{s_\ell}} - \overleftarrow{\xi}_{X_{s_\ell}^{(n)}} \ . \tag{19}$$



(a) Original

(b) FBP, M=20 (RMSE = 0.3481)

(c) TV, M=20 (RMSE = 0.0132)

(d) NUV-GMP, M=20 (RMSE = 0.0031)

(e) FBP, M=30 (RMSE = 0.3115)

(f) TV, M=30 (RMSE = 0.0043)

(g) NUV-GMP, M=30 (RMSE = 0.0008)

Fig. 5. Reconstruction results from noise-free projections for the Shepp-Logan phantom using FBP, TV [4], and the proposed method (NUV-GMP). Top: reconstruction from $M = 20$ projections. Bottom: reconstruction from $M = 30$ projections. Also shown is the root-mean squared error (RMSE).

Then, compute

$$\sigma_n^2 := \sigma_Z^2 + \sum_{\ell: a_{n,\ell} \neq 0} a_{n,\ell}^2 \vec{w}_{X_{s_\ell}^{(n)}}^{-1} \tag{20}$$

$$m_n := y_n - \sum_{\ell: a_{n,\ell} \neq 0} a_{n,\ell} \vec{w}_{X_{s_\ell}^{(n)}}^{-1} \vec{\xi}_{X_{s_\ell}^{(n)}} \ . \tag{21}$$

For all $\ell \in \{1, \ldots, L\}$ such that $a_{n,\ell} \neq 0$, update

$$\overleftarrow{w}_{X_{s_\ell}} := \overleftarrow{w}_{X_{s_\ell}} - \overleftarrow{w}_{X_{s_\ell}^{(n)}} \tag{22}$$

$$\overleftarrow{\xi}_{X_{s_\ell}} := \overleftarrow{\xi}_{X_{s_\ell}} - \overleftarrow{\xi}_{X_{s_\ell}^{(n)}} \tag{23}$$

$$\overleftarrow{w}_{X_{s_\ell}^{(n)}} := a_{n,\ell}^2 \left(\sigma_n^2 - a_{n,\ell}^2 \vec{w}_{X_{s_\ell}^{(n)}}^{-1} \right)^{-1} \tag{24}$$

$$\overleftarrow{\xi}_{X_{s_\ell}^{(n)}} := \frac{\overleftarrow{w}_{X_{s_\ell}^{(n)}}}{a_{n,\ell}} \left(m_n + a_{n,\ell} \vec{w}_{X_{s_\ell}^{(n)}}^{-1} \vec{\xi}_{X_{s_\ell}^{(n)}} \right) \tag{25}$$

$$\overleftarrow{w}_{X_{s_\ell}} := \overleftarrow{w}_{X_{s_\ell}} + \overleftarrow{w}_{X_{s_\ell}^{(n)}} \tag{26}$$

$$\overleftarrow{\xi}_{X_{s_\ell}} := \overleftarrow{\xi}_{X_{s_\ell}} + \overleftarrow{\xi}_{X_{s_\ell}^{(n)}} \ . \tag{27}$$

2) Process the D node. For each pair $(\ell, \ell') \in \Delta$, update

$$\overleftarrow{w}_{X_{s_\ell}^{(\ell,\ell')}} := \overleftarrow{w}_{X_{s_\ell}} + \vec{w}_{X_{s_\ell}} - \vec{w}_{X_{s_\ell}^{(\ell,\ell')}} \tag{28}$$

$$\overleftarrow{\xi}_{X_{s_\ell}^{(\ell,\ell')}} := \overleftarrow{\xi}_{X_{s_\ell}} + \vec{\xi}_{X_{s_\ell}} - \vec{\xi}_{X_{s_\ell}^{(\ell,\ell')}} \ , \tag{29}$$

and similarly for the upward messages on edge $X_{s_{\ell'}}^{(\ell,\ell')}$. Then, compute

$$\vec{w}_{X_{s_\ell}^{(\ell,\ell')}} := \overleftarrow{w}_{X_{s_{\ell'}}^{(\ell,\ell')}} \left(1 + (\sigma_\epsilon^2 + \sigma_{\ell,\ell'}^2) \overleftarrow{w}_{X_{s_{\ell'}}^{(\ell,\ell')}} \right)^{-1} \tag{30}$$

$$\vec{\xi}_{X_{s_\ell}^{(\ell,\ell')}} := \overleftarrow{\xi}_{X_{s_{\ell'}}^{(\ell,\ell')}} \left(1 + (\sigma_\epsilon^2 + \sigma_{\ell,\ell'}^2) \overleftarrow{w}_{X_{s_{\ell'}}^{(\ell,\ell')}} \right)^{-1} , \tag{31}$$

and similarly for the downward messages on edge $X_{s_{\ell'}}^{(\ell,\ell')}$.

3) For each $\ell \in \{1, \ldots, L\}$, compute the messages sent from node D to node A

$$\vec{w}_{X_{s_\ell}} := \sum_{\ell': (\ell,\ell') \in \Delta} \vec{w}_{X_{s_\ell}^{(\ell,\ell')}} + \sum_{\ell': (\ell',\ell) \in \Delta} \vec{w}_{X_{s_\ell}^{(\ell',\ell)}} \tag{32}$$

$$\vec{\xi}_{X_{s_\ell}} := \sum_{\ell': (\ell,\ell') \in \Delta} \vec{\xi}_{X_{s_\ell}^{(\ell,\ell')}} + \sum_{\ell': (\ell',\ell) \in \Delta} \vec{\xi}_{X_{s_\ell}^{(\ell',\ell)}} \ . \tag{33}$$

When convergence is achieved, the posterior mean $m_{U_{\ell,\ell'}}$ and variance $\sigma_{U_{\ell,\ell'}}^2$ can be computed with the formulae

$$\frac{1}{\sigma_{U_{\ell,\ell'}}^2} = \frac{1}{\sigma_{\ell,\ell'}^2 + \sigma_\epsilon^2} + \frac{\overleftarrow{w}_{X_{s_\ell}^{(\ell,\ell')}} \overleftarrow{w}_{X_{s_{\ell'}}^{(\ell,\ell')}}}{\overleftarrow{w}_{X_{s_\ell}^{(\ell,\ell')}} + \overleftarrow{w}_{X_{s_{\ell'}}^{(\ell,\ell')}}} \tag{34}$$

$$m_{U_{\ell,\ell'}} = \sigma_{U_{\ell,\ell'}}^2 \frac{\overleftarrow{w}_{X_{s_\ell}^{(\ell,\ell')}} \overleftarrow{\xi}_{X_{s_{\ell'}}^{(\ell,\ell')}} - \overleftarrow{w}_{X_{s_{\ell'}}^{(\ell,\ell')}} \overleftarrow{\xi}_{X_{s_\ell}^{(\ell,\ell')}}}{\overleftarrow{w}_{X_{s_{\ell'}}^{(\ell,\ell')}} + \overleftarrow{w}_{X_{s_\ell}^{(\ell,\ell')}}} \ . \tag{35}$$

Note that all loops over the voxel index ℓ and over the indices in Δ can be performed in parallel.

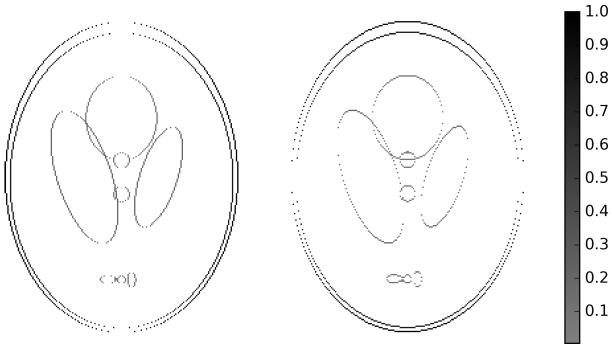


Fig. 6. Estimated parameters $\hat{\sigma}^2_{\ell,\ell'}$ for $M = 30$ projections. Left: along horizontal neighbors; right: along vertical neighbors. ($\approx 98\%$ of zeros)

V. SIMULATION RESULTS

We illustrate the proposed approach with 2D examples, but our 3D simulations (not reported in this paper) support the same conclusions. In Fig. 5, we show the reconstruction results for the Shepp-Logan phantom [25] with 256×256 pixels using $M = 20$ and $M = 30$ projections onto a line detector with $Q = 512$ pixels. The projections are obtained using the distance-driven method [17] with a fan-beam geometry. Our algorithm (NUV-GMP) is compared with a standard FBP method and the state-of-the-art total variation (TV) method [4]. For our algorithm, we used $\sigma^2_\epsilon = 10^{-4}$, $\sigma^2_Z = 10^{-2}$, and the $\sigma^2_{\ell,\ell'}$'s initialized to $0.1 \cdot \sigma^2_\epsilon$. We used 15 EM updates, and 15 message passing iterations within each EM update. The estimated image is obtained from the posterior mean of $\mathbf{X}$. For a fair comparison, we use 200 iterations for the TV method.

The results in Fig. 5 indicate that the proposed method greatly improves on FBP and yields slightly better reconstructions than the TV method.

In Fig. 6, we plot the estimated variances $\hat{\sigma}^2_{\ell,\ell'}$ from our algorithm for the case of 30 projections. We observe that the sparsity structure of $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$ directly reflects the edge structure of the reconstructed image. Note that both σ^2_Z and σ^2_ϵ can be tuned to control the sparsity level.

VI. CONCLUSION

We have proposed a new voxel-domain (or pixel-domain) prior for 3D images (or 2D images, respectively) that promotes both continuity and occasional sharp transitions, and we have also proposed a corresponding practical reconstruction algorithm based on EM and scalar Gaussian message passing. In our simulations with the Shepp-Logan phantom, the proposed approach yields slightly better reconstructions than a state-of-the-art method. Even better results (not reported here) are obtained for phantoms that are not piecewise constant.

The proposed approach is most similar in spirit to [7], of which we became aware only very recently. Preliminary simulation results indicate that the performance is similar as well, but a detailed comparison has not yet been carried out.

REFERENCES

[1] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction From Projections*. Springer-Verlag, 2009.

[2] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. on Medical Imaging*, vol. 13, no. 2, pp. 290–300, 1994.

[3] S. Maeda, W. Fukuda, A. Kanemura, and S. Ishii, "Maximum a posteriori X-ray computed tomography using graph cuts," in *Journal of Physics: Conference Series*, vol. 233, no. 1. IOP Publishing, 2010.

[4] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, no. 17, p. 4777, 2008.

[5] M. Yan and L. A. Vese, "Expectation maximization and total variation-based model for computed tomography reconstruction from undersampled data," in *Proc. SPIE 7961, Medical Imaging*. Int. Society for Optics and Photonics, 2011.

[6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[7] M. Borgerding, P. Schniter, J. Vila, and S. Rangan, "Generalized approximate message passing for cosparse analysis compressive sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 3756–3760.

[8] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. on Image Processing*, vol. 18, no. 9, pp. 2059–2070, 2009.

[9] G. K. Chantas, N. P. Galatsanos, and A. C. Likas, "Bayesian restoration using a new nonstationary edge-preserving image prior," *IEEE Trans. on Image Processing*, vol. 15, no. 10, pp. 2987–2997, 2006.

[10] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[11] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[12] R. M. Neal, *Bayesian Learning for Neural Networks*. Springer-Verlag, 2012, vol. 118.

[13] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 341–355.

[14] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai, "On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing," in *Information Theory and Applications Workshop (ITA), La Jolla, CA*, Feb 2016. [Online]. Available: http://arxiv.org/abs/1602.02673

[15] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 28–41, 2004.

[16] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.

[17] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in Medicine and Biology*, vol. 49, no. 11, p. 2463, 2004.

[18] N. Zalmai, H. Malmberg, and H.-A. Loeliger, "Blind deconvolution of sparse but filtered pulses with linear state space models," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4194–4198.

[19] F. C. Curriero and S. Lele, "A composite likelihood approach to semivariogram estimation," *Journal of Agricultural, Biological, and Environmental Statistics*, pp. 9–28, 1999.

[20] B. G. Lindsay, "Composite likelihood methods," *Contemporary Mathematics*, vol. 80, no. 1, pp. 221–39, 1988.

[21] J. Berger, *Statistical Decision Theory And Bayesian Analysis*, 2nd ed., ser. Series in statistics. Springer, 1985.

[22] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, no. 10, pp. 2173–2200, 2001.

[23] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," in *NIPS*, vol. 13, 2000, pp. 689–695.

[24] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *J. Machine Learning Research*, vol. 7, pp. 2031–2064, 2006.

[25] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. on Nuclear Science*, vol. 21, no. 3, pp. 21–43, 1974.