# MAP Tomographic Reconstruction with a Spatially Adaptive Hierarchical Image Model

Christophoros Nikou

Department of Computer Science and Engineering
University of Ioannina, Greece
Email: cnikou@cse.uoi.gr

*Abstract*—A method for penalized likelihood tomographic reconstruction is presented which is based on a spatially adaptive stochastic image model. The model imposes onto the image a smoothing Gaussian prior whose parameters follow a Gamma distribution. Three variations of the model are examined: (i) a stationary model, where the Gamma distribution has the same constant parameter for the entire image, (ii) a non stationary model, where this parameter varies with respect to location and (iii) a non stationary directional model where the parameter varies also with respect to orientation (horizontal or vertical direction). In all cases, the MAP criterion provides a closed form solution for both the unknown image and the parameters of the model. Numerical experiments showed that the reconstructions obtained using the proposed image priors outperform the state of the art EM based methods.

## I. INTRODUCTION

Several maximum likelihood (ML) or maximum *a posteriori* (MAP) iterative algorithms for reconstructing tomographic images have been proposed in the last three decades [1]. Among them, a well-known family of methods leads to solutions that minimize certain combinations of the Kullback-Leibler (KL) distances between the observed photon counts and the projected unobserved image [2]. The solution is derived from alternating minimization of related KL distances between convex sets.

MAP or penalized maximum likelihood tomographic reconstruction methods impose a prior probability density function (pdf) on the image to be reconstructed which usually aims to encourage the image to be smooth in order to suppress the effect of noise. This assumption is based on the knowledge that the system (projection) matrix suppresses image detail due to its blurring effect. Therefore, any such detail present in the reconstruction is more probably to have arisen from noise [3]. For instance, in [4], to avoid over-fitting to noisy data, a penalty function was employed involving also the KL distance between a prior estimate of the unobserved image and the current estimate at each iteration.

A common model for the prior is the Markov random field (MRF) expressed by the Gibbs distribution [5]. Many methods were proposed in that framework differing on the choice of the potential function [6], [7], [8], [9]. Among them, the Gauss MRF (MRF) has the advantage of estimating its parameters from the data [10]. More recently, in [11], the notion of clustered intensity histogram is introduced in a penalized

likelihood method. The prior pdf is a mixture of Gamma distributions enforcing positivity of the reconstructed image intensities. An alternating optimization is performed where a likelihood estimate is followed by inference of the mixture model [11]. A monotonically decreasing surrogate objective function resulting in a closed form expression is proposed in [12] while the median root prior was also used to impose spatial smoothness and stabilize the solution [13]. Finally, a nonlocal prior was designed [14] where the definition of a pixel's neighborhood is broadened.

In this paper, we propose a sparse, edge-preserving, spatially adaptive model for the image to be reconstructed which relies on a two level Gaussian non-stationary prior. The prior assumes that the image intensities follow a Gaussian distribution whose variance is spatially adaptive. Thus, the strength of the prior depends on local image statistics, leading to distinct variances at each pixel. To avoid over-parameterization, a Gamma hyper-prior [15] is imposed on the spatially varying variances of the model. Model inference is obtained by a maximum *a posteriori* (MAP) formulation which yields closed form updates for the image and the spatially varying variances of the involved distributions.

Priors similar in spirit to those proposed in this work, have been applied to image restoration [16] and non-rigid image registration [17]. The related model is characterized by an intrinsic flexibility depending on the degree of detail carried by its hyper-parameters. We follow this step-by-step decomposition in its description. At first, the simplest version of the model is presented, where a unique hyper-parameter controls the variances of the pixel intensities yielding a stationary image model. Secondly, the model is refined to capture each pixel's probability with a different hyper-parameter. thus, the model becomes spatially varying or *non stationary*. Finally, we employ different hyper-parameters for the horizontal and vertical image directions at each pixel in order to make the model more adaptive to local edge directions.

The proposed reconstruction method is successfully compared to state-of-the art tomographic reconstruction algorithms. Furthermore, we compared the different versions of the model among themselves to investigate how the augmented model complexity affects the quality of the reconstruction.

In the remaining of the paper, the proposed image model is presented in section II, experimental results are shown in section III and conclusions are drawn in section IV.

## II. THE IMAGE MODEL

Let $\mathbf{f}$ be the vectorized form of the image to be reconstructed. Let also $\mathbf{g}$ be the observed projections (sinogram), also in vectorized form and let $\mathbf{H}$ represent the projection matrix. Penalized likelihood models rely on the stochastic interpretation of Tikhonov regularization [18] by introducing an appropriate prior $p(\mathbf{f})$ for the image $\mathbf{f}$. The likelihood function $p(\mathbf{g}|\mathbf{f})$ is related to the posterior probability $p(\mathbf{f}|\mathbf{g})$ by the Bayes rule $p(\mathbf{f}|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f})p(\mathbf{f})$.

In tomography, the likelihood $p(\mathbf{g}|\mathbf{f})$ is a Poisson distribution assuming independence between counts

$$p(\mathbf{g}|\mathbf{f}) = \prod_{i=1}^{N} ([\mathbf{Hf}]_i)^{g_i} \frac{\exp(-[\mathbf{Hf}]_i)}{g_i!}, \qquad (1)$$

where $N$ is the number of projection measures, $g_i$ is the $i$-th component of $\mathbf{g}$ and $[\mathbf{Hf}]_i$ is the $i$-th component of vector $\mathbf{Hf}$. A common approach for the likelihood is to approximate it by a Gaussian which is considered to be valid if the mean $[\mathbf{Hf}]_i$ is greater than 20 [19]. Therefore, the log-likelihood is expressed by

$$\log p(\mathbf{g}|\mathbf{f}) \approx C(\mathbf{g}) - \frac{1}{2}(\mathbf{Hf} - \mathbf{g})^T \mathbf{D}^{-1}(\mathbf{Hf} - \mathbf{g}), \quad (2)$$

where $C(\mathbf{g})$ collects the terms not depending on the image $\mathbf{f}$ and $\mathbf{D}$ is the diagonal covariance matrix.

Having defined the likelihood and using an appropriate prior, MAP estimates for the image $\mathbf{f}$ may be obtained by minimizing the negative log-posterior

$$-\log p(\mathbf{f}|\mathbf{g}) = -\log p(\mathbf{g}|\mathbf{f}) - \log p(\mathbf{f}) \qquad (3)$$

with respect to $\mathbf{f}$.

### A. Stationary model

Based on the assumption that the image to be reconstructed should be smooth, we assume that the prior for the image $\mathbf{f}$ is Gaussian and emphasizes low frequency information. This assumption may be expressed by employing a high frequency operator represented by matrix $\mathbf{Q}$. In our case, $\mathbf{Q}$ applies the Laplacian operator and $\mathbf{Qf}$ is the vectorized form of the Laplacian of image $\mathbf{f}$. Hence, the prior for an image pixel may be written as

$$p(f_i) = \frac{(\alpha|\mathbf{Q}^T\mathbf{Q}|)^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}([\mathbf{Qf}]_i)^T \alpha([\mathbf{Qf}]_i)\right), \quad (4)$$

where $[\mathbf{Qf}]_i$ is the Laplacian of the image at the $i$-th pixel. Considering the pixels to be independent, we come up with the probability for the entire image $p(\mathbf{f}) = \prod_{i=1}^{N} p(f_i)$ which is given by

$$p(\mathbf{f}) = \frac{(\alpha|\mathbf{Q}^T\mathbf{Q}|)^{N/2}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}\alpha(\mathbf{Qf})^T(\mathbf{Qf})\right), \quad (5)$$

The above zero-mean normal distribution assigns a high probability to images not exhibiting rich edge information. Parameter $\alpha$ controls the precision matrix $\mathbf{Q}^T\mathbf{Q}$ (inverse covariance) and consequently the shape of the distribution (it is generally called *hyper-parameter*). The simplest approach is to

consider parameter $\alpha$ spatially constant, yielding a stationary model for the whole image. This implies that the statistics for $\mathbf{Qf}$ are Gaussian, independent and identically distributed. Applying the prior in (5) to the MAP expression (3) and minimizing it with respect to $\mathbf{f}$ provides a linear system with unknowns the elements of the image $\mathbf{f}$:

$$(\mathbf{H}^T\mathbf{D}^{-1}\mathbf{H} + \alpha\mathbf{Q}^T\mathbf{Q})\mathbf{f} = \mathbf{H}^T\mathbf{D}^{-1}\mathbf{g} \qquad (6)$$

This is also the solution to the standard weighted least squares problem:

$$\arg\min_{\mathbf{f}}\{||\mathbf{Qf}||^2\} \text{ subject to } \mathbf{D}^{-\frac{1}{2}}\mathbf{Hf} = \mathbf{D}^{-\frac{1}{2}}\mathbf{g}. \quad (7)$$

However, in the least squares approach, parameter $\alpha$ has to be determined empirically while in the MAP methodology developed here this parameter is naturally estimated from the data. This is obtained by maximizing (3) with respect to $\alpha$, thus obtaining:

$$\alpha = \frac{N - 1}{||\mathbf{Qf}||^2} \qquad (8)$$

The linear system in (6) may be solved iteratively using the conjugate gradient method. At each iteration the precision parameter $\alpha$ is also updated by eq. (8). Notice that the term $\mathbf{H}^T\mathbf{D}^{-1}\mathbf{H}$ represents a projection operation (represented by matrix $\mathbf{H}$), followed by a multiplication by the diagonal matrix $\mathbf{D}^{-1}$, followed by a backprojection operation (represented by matrix $\mathbf{H}^T$). Also, the operator $\mathbf{Q}^T\mathbf{Q}$ may be effectively computed in the Fourier domain.

### B. Non stationary model

In order to make the model more flexible we impose spatial adaptivity. In other words, the precision parameter of the model depends now on the pixel location and we have one precision parameter for each pixel. The pixel prior now becomes

$$p(f_i) = \frac{(\alpha_i|\mathbf{Q}^T\mathbf{Q}|)^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}([\mathbf{Qf}]_i)^T \alpha_i([\mathbf{Qf}]_i)\right), \quad (9)$$

where we have introduced the spatially varying parameter $\alpha_i$, $i = 1, ..., N$. Hence, the probability for the entire image is expressed by

$$p(\mathbf{f}) = \frac{(|\mathbf{Q}^T\mathbf{Q}|)^{N/2}}{(2\pi)^{N/2}} \prod_{i=1}^{N} \alpha_i^{1/2} \exp\left(-\frac{1}{2}(\mathbf{Qf})^T\mathbf{A}(\mathbf{Qf})\right),$$
$$(10)$$

where $\mathbf{A} = \text{diag}\{\alpha_1, \alpha_2, ..., \alpha_N\}$ is a $N \times N$ diagonal matrix. Let us also notice that this prior is *proper* since it integrates to 1.

The drawback of this prior is that it introduces $N$ parameters $\alpha_i$ that have to be estimated and data overfitting may occur. Thus, we follow the Bayesian paradigm [20] and add one more layer to our model and consider $\alpha_i$, $i = 1, ..., N$, as random variables and introduce a Gamma hyper-prior for them. The Gamma distribution is a natural choice for the model parameters because it is conjugate to the Gaussian

distribution, thus, facilitating model inference. We consider the following parametrization of the Gamma distribution [15]:

$$p(\alpha_i) \propto (\alpha_i)^{(l-2)/2} \exp\left(-m(l-2)\alpha_i\right) \qquad (11)$$

Following this parametrization, the mean and variance of Gamma, are given by $E[\alpha_i] = l\left(2m(l-2)\right)^{-1}$ and $Var[\alpha_i] = l\left(2m^2(l-2)^2\right)^{-1}$. We employ this parametrization because the value of parameter $l$ may be interpreted as the level of confidence to the prior knowledge provided by the Gamma hyper-prior. More specifically, as $l \to \infty$, $E[\alpha_i] \to (2m)^{-1}$ and $Var[\alpha_i] \to 0$. This means that the prior becomes very informative and restrictive resulting in $\alpha_i = (2m)^{-1}$, $\forall i$ yielding a stationary image model. On the other hand, as $l \to 2$, then $E[\alpha_i] \to \infty$ and $Var[\alpha_i] \to \infty$ and the prior becomes uninformative and does not play any role in the determination of the values of $\alpha_i$. In that case, the model becomes highly non stationary. In other words, parameter $l$ controls the degree of non stationarity of the image model.

Minimizing the MAP expression (3) with respect to $\mathbf{f}$ this time leads to the following linear system of equations:

$$(\mathbf{H}^T \mathbf{D}^{-1} \mathbf{H} + \mathbf{Q}^T \mathbf{A} \mathbf{Q})\mathbf{f} = \mathbf{H}^T \mathbf{D}^{-1} \mathbf{g} \qquad (12)$$

Moreover, minimization with respect to the spatially varying parameters $\alpha_i$, $\forall i$ provides the following update:

$$\alpha_i = \frac{l-1}{[\mathbf{Q}\mathbf{f}]_i^2 + 2m(l-2)} \qquad (13)$$

The role of the Gamma pdf becomes apparent by observing the above equation. For example, when the image $\mathbf{f}$ is smooth the first term in the denominator of (13) becomes zero. Thus, without the Gamma pdf ($m = 0$ and $l = 2$) the estimates for $\alpha_i$ become unstable. Parameter $l$ may take values within the interval $(2, +\infty)$. Its choice affects our model in the following way. When $l$ is very close to 2, the second term in the denominator ensures that $\alpha_i$ depends only on $[\mathbf{Q}\mathbf{f}]_i^2$ and thus the Gamma hyper-prior is non-informative since the estimates of $\alpha_i$ depend only on the data. On the other hand, if we assign a large values to $l$, the second term in the denominator of (13) dominates. Then, the estimates of $\alpha_i$ do not depend on the data and have the same value for all spatial locations. As a result, our model degenerates to a spatially invariant model. Parameter $m$ is extracted from the data, since it turns out that it is proportional to the variance of $\mathbf{Q}\mathbf{f}$ [15].

*C. Non stationary directional model*

We now present a more flexible form of the previously presented image models which imposes a distinct prior to each image direction (horizontal, vertical and two diagonals) for a given pixel. However, as we will see, the corresponding prior is *improper*.

We assume that the first order differences of the image $\mathbf{f}$ in matrix vector form are given by $\varepsilon^k = \mathbf{Q}^k \mathbf{f}$, for $k = 1, 2, 3, 4$, in four directions respectively, where $\mathbf{Q}^k$ is now the directional difference operator. Gaussian statistics are assumed for the local differences $\varepsilon_i^k \sim \mathcal{N}(0, (\alpha_i^k)^{-1})$ with a spatially varying set of parameters $\mathbf{A}^k = \text{diag}\{\alpha_1^k, \alpha_2^k, ..., \alpha_N^k\}$ for each image

direction $k$. The whole set of variables may be represented by the $4N \times 4N$ matrix $\tilde{\mathbf{A}} = \text{diag}\{\mathbf{A}^1, \mathbf{A}^2, \mathbf{A}^3, \mathbf{A}^4\}$ and the pdf of the differences $\varepsilon = [\varepsilon^1, \varepsilon^2, \varepsilon^3, \varepsilon^4]^T$, where $\varepsilon^k = [\varepsilon_1^k, \varepsilon_2^k, ..., \varepsilon_N^k]^T$ is expressed by

$$p(\tilde{\varepsilon}) \propto \prod_{k=1}^{4} \prod_{i=1}^{N} (\alpha_i^k)^{1/2} \exp\left(-\frac{1}{2}\left(\tilde{\varepsilon}^T \tilde{\mathbf{A}} \tilde{\varepsilon}\right)\right) \qquad (14)$$

which assumes independence between locations as well as between directions at each location. To relate $\tilde{\varepsilon}$ with the image $\mathbf{f}$, we define the $4N \times N$ operator $\tilde{\mathbf{Q}} = \text{diag}\{(\mathbf{Q}^1)^T, (\mathbf{Q}^2)^T, (\mathbf{Q}^3)^T, (\mathbf{Q}^4)^T\}$. Then, the relation between the image and the directional differences is $\tilde{\varepsilon} = \tilde{\mathbf{Q}}\mathbf{f}$ and the corresponding image prior is given by

$$p(\mathbf{f}) \propto \prod_{k=1}^{4} \prod_{i=1}^{N} (\alpha_i^k)^{1/8} \exp\left(-\frac{1}{2}\left(\tilde{\mathbf{Q}}\mathbf{f}\right)^T \tilde{\mathbf{A}} \left(\tilde{\mathbf{Q}}\mathbf{f}\right)\right) \qquad (15)$$

The above prior is *improper* since it is not scaled to integrate to 1 because its normalizing constant depends on the non-square matrix $\tilde{\mathbf{Q}}$ and the corresponding determinant cannot be obtained analytically as a matrix product (see [16] for more details).

To complete the model in a Bayesian sense, in order to avoid overfitting, we impose a Gamma hyper-prior to the spatially varying parameters at each direction with the corresponding pdfs being of the form

$$p(\alpha_i^k) \propto (\alpha_i^k)^{(l_k-2)/2} \exp\left(-m_k(l_k-2)\alpha_i^k\right) \qquad (16)$$

where we have four distinct Gamma priors parameterized by the pairs $(m_k, l_k)$, for each direction $k = 1, 2, 3, 4$. To summarize, this non stationary directional image model may be viewed as an *improper* generalization of both the stationary and the non stationary models presented above.

Substituting the image prior in (15) to the expression in (3) and maximizing it with respect to $\mathbf{f}$, we obtain

$$\left[\mathbf{H}^T \mathbf{D}^{-1} \mathbf{H} + \sum_{k=1}^{4} \left(\mathbf{Q}^k\right)^T \mathbf{A}^k \mathbf{Q}^k\right] \mathbf{f} = \mathbf{H}^T \mathbf{D}^{-1} \mathbf{g} \qquad (17)$$

Similarly, maximization of (3) with respect to the spatially varying and directional parameters $\alpha_i^k$ gives the following update

$$\alpha_i^k = \frac{\frac{1}{8} + \frac{1}{2}(l_k - 2)}{\frac{1}{2}(\epsilon_i^k)^2 + m_k(l_k - 2)} \qquad (18)$$

### III. RESULTS

The performances of the proposed image models were examined using the well known Shepp-Logan phantom. The compared methods were evaluated in terms of the improvement in signal to noise ratio (ISNR) with respect to a reconstruction obtained by a simple filtered back-projection using the Ram-Lak filter:

$$\text{ISNR} = 10 \log_{10} \frac{||\mathbf{f} - \mathbf{f}_{FBP}||^2}{||\mathbf{f} - \hat{\mathbf{f}}||^2} \qquad (19)$$

where $\mathbf{f}$ is the ground truth image, $\mathbf{f}_{FBP}$ is the reconstructed image by filtered back-projection and $\hat{\mathbf{f}}$ is the reconstructed

image using the proposed image model. Practically, ISNR measures the improvement (or deterioration) in the quality of the reconstruction of the proposed method with respect to the reconstruction obtained by filtered back-projection. To this end, degraded images were generated from the initial image by modifying the total photon counts in order to simulate a Poisson distribution. More specifically, the total photon counts in a $256 \times 256$ image of the phantom varied between $5 \times 10^6$ and $5 \times 10^8$ which corresponds, on average, to a range between 100 and 10000 photons per pixel. In all of the experiments we have set the hyper-parameter of the Gamma distribution $l = 2.01$.



(a)



(b)
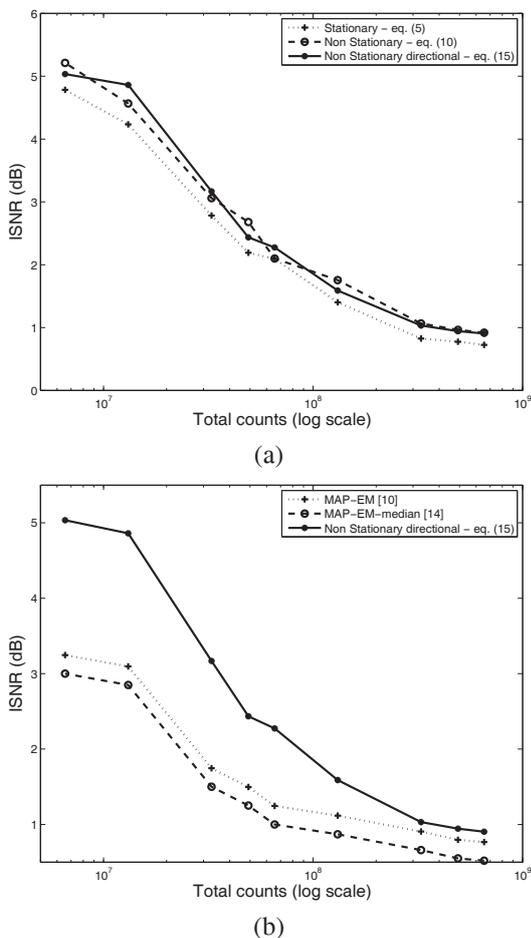
Fig. 1. ISNR comparative curves between (a) the three versions of the proposed model and (b) the non stationary directional model (10), the MAP-EM algorithm [6] and the MAP-EM with median prior algorithm (MAP-EM-median) [13]. The test image is the $256 \times 256$ Shepp-Logan phantom.

In figure 1(a), the three image models are compared. The points on the curves represent the average ISNR for each noise level (total photon counts) computed by 40 random realizations of the experiment. As it can be observed, as the noise increases (the number of photon counts per pixel decreases) the ISNR becomes larger. This means that the improvement with respect to the filtered back-projection reconstruction increases. The simpler stationary model has consistently lower ISNR val-

ues compared to the non stationary versions of the method. However, the differences are relatively small (0.1 to 0.25 dB). The non stationary methods have similar results and their differences are practically insignificant. This behavior may be explained by the general piecewise smooth form of the phantom image and emission tomography images in general. The model complexity added by the directional prior may overfit image data with Poisson statistics. This prior may be appropriate for image restoration problems (e.g. [16]) where blurring, Gaussian noise and image content may increase the problem's complexity. Nevertheless, in the majority of the cases in fig. 1(a) the no stationary directional model is slighltly superior.

Using the same experimental framework, the proposed models were also compared to established reconstruction methods such as the MAP-EM algorithm [6] and the robust MAP-EM with median prior algorithm (MAP-EM-median) [13]. The results are summarized in figure 1(b) where only the non stationary directional model is shown for clarity of presentation purposes (this is the same curve with fig. 1(a)). As it can be seen, the proposed model provides largely better ISNR values than the other MAP-based algorithms. The difference is more pronounced in low SNR (low number of total counts) and becomes smaller as the image becomes less degraded.

In order to compare the statistical properties of the proposed MAP methods, we also considered the bias (BIAS) and the variance (VAR) of the reconstructed images. These quantities are estimated through Monte-Carlo simulations by the following expressions:

$$\text{BIAS} = ||\mathbf{f} - \bar{\mathbf{f}}||, \quad \text{VAR} = \sum_{k=1}^{M} ||\bar{\mathbf{f}} - \hat{\mathbf{f}}_k||^2,$$

with $\bar{\mathbf{f}} = \frac{1}{M} \sum_{k=1}^{M} \hat{\mathbf{f}}_k$, where $\mathbf{f}$ is the ground truth image and $\hat{\mathbf{f}}_k$, for $k = 1, ..., M$, is the $k$-th reconstructed image, obtained from $M$ different experiments. For each noise level, the degraded images were corrupted with $M = 40$ different noise realizations. The results are summarized in figure 2. The bias (fig. 2(a)), measuring the distance of the mean estimated image from the ground truth, is approximately constant for all the versions of the model. However, the simpler, stationary version provides consistently a higher bias value with respect to the non stationary variants as it was the case for the ISNR values. To highlight the small order of magnitude of the values of the bias, let us notice, that, these values concern all the image pixels as they are computed by (III). To obtain an idea of the *per pixel* bias, these values should be divided by $256^2$.

The variances of the three versions of the proposed model, shown in figure 2(b)), decrease as the total counts increase. Moreover, for low photon counts, the differences between the stationary and the non stationary variants of the model are more pronounced. These differences disappear as the photon counts increase. Notice also that the variance values correspond to the whole image and divided by $256^2$ provide the order of magnitude *per pixel*.

(a)



(b)
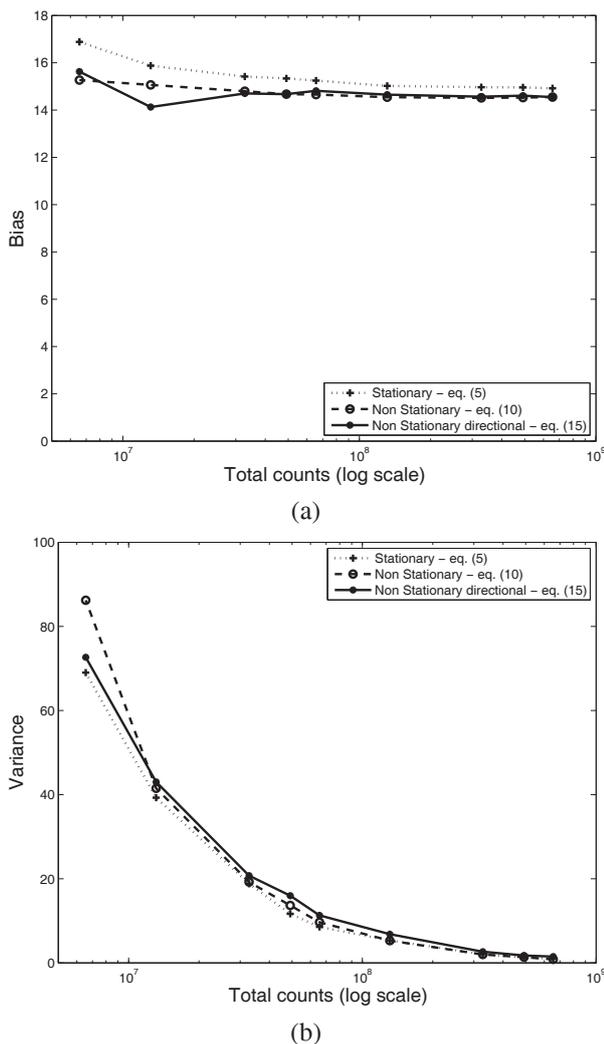
Fig. 2. (a) Bias and (b) variance comparative curves between the three versions of the proposed model. The test image is the $256 \times 256$ Shepp-Logan phantom.

## IV. CONCLUSION

We have presented a MAP methodology for tomographic image reconstruction using a family of stochastic image priors which gradually refine the image model from stationary to non stationary. The proposed approach is favorably compared to established EM-based reconstruction methods. Moreover, numerical experiments showed that the non stationary versions of the model are consistently superior with respect to the stationary version. Between them, the non stationary methods do not exhibit significant differences due to the piecewise smooth nature of tomographic images.

Many tomographic reconstruction algorithms employ the linearization in (2) in order to facilitate the integration of image priors into the problem. Although this approximation is generally admissible and valid, a perspective of this work is to integrate the image priors (5), (10) and (15) directly into an Expectation-Maximization (EM) framework for image reconstruction without using eq. (2). By these means, matrix inversions involved in (6), (12) and (17) which are now handled by the conjugate gradient method would be avoided.

## REFERENCES

[1] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Transactions on Medical Imaging*, vol. 6, no. 2, pp. 106–114, 1987.

[2] C. Byrne, "Iterative image reconstruction algorithms based on cross entropy minimization," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 96–103, 1993.

[3] M. Wernick, *Emission Tomography: The Fundamentals of PET and SPECT*. Prentice Hall, 2004.

[4] J. A. Case, T.-S. Pan, M. A. King, L. Der-Shan, B. C. Penney, and M. S. Z. Rabin, "Reaction of truncation artifacts in fan beam transmission imaging using a spatially varying gamma prior," , *IEEE Transactions on Nuclear Science*, vol. 42, no. 6, pp. 2260 –2265, 1995.

[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 721–741, 1984.

[6] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Transactions on Medical Imaging*, vol. 9, no. 1, pp. 84–93, 1990.

[7] T. J. Hebert and R. Leahy, "Statistic-based MAP image reconstruction from Poisson data using Gibbs priors," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2290–2303, 1992.

[8] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaczszak, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 516–526, 1997.

[9] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 687–701, 1994.

[10] S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields with applications to Bayesian tomography," , *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1029 –1044, 1998.

[11] I. T. Hsiao, A. Rangarajan, and G. Gindi, "Joint MAP Bayesian tomographic reconstruction with a Gamma-mixture prior," *IEEE Transactions on Medical Imaging*, vol. 11, no. 12, pp. 1466–14 775, 2002.

[12] J. H. Chang and J. M. M. Anderson, "Regularized image reconstruction algorithms for positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 23, no. 9, pp. 1165–1175, 2004.

[13] I. T. Hsiao, A. Rangarajan, and G. Gindi, "A new convex edge-preserving median prior with apllications to tomography," *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 580–585, 2003.

[14] Y. Chen, J. Ma, Q. Feng, L. Luo, P. Shi, and W. Chen, "Nonlocal prior Bayesian tomographic reconstruction," *Journal of Mathematical Imaging and Vision*, vol. 30, pp. 133–146, 2008.

[15] R. Molina, "On the hierarchical Bayesian approach to image restoration. applications to astronomical images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 11, pp. 1122–1128, 1994.

[16] J. Chantas, N. P. Galatsanos, and A. Likas, "Bayesian restoration using a new non-stationary edge-preserving image prior," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2987–2997, 2006.

[17] E. Fotiou, C. Nikou, and N. Galatsanos, "A spatially adaptive hierarchical stochastic model for non-rigid image registration," in *Proceedings of the $16^{th}$ European Signal Processing Conference (EUSIPCO'08)*, Lausanne, Switzerland, 2008.

[18] A. Tikhonov and V. Arsenin, *Solution of Ill-posed Problems*. Washington, USA: Winston & Sons, 1977.

[19] D. S. Lalush and B. M. W. Tsui, "A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm," *Medical Physics*, vol. 22, pp. 1273–1284, 1995.

[20] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.