

Penalized-Likelihood Image Reconstruction for Emission Tomography Using Higher Order Convex-Nonquadratic Priors

Soo-Jin Lee

Department of Electronic Engineering
Paichai University
439-6 Doma 2-Dong, Seo-Ku, Taejon, 302-735, Korea

ABSTRACT

Penalized-likelihood methods using Bayesian smoothing priors have formed the core of the development of reconstruction algorithms for emission tomography. In particular, there has been considerable interest in edge-preserving prior models, which are associated with smoothing penalty functions that are nonquadratic functions of nearby pixel differences. Our early work¹ used a higher-order nonconvex prior that imposed piecewise smoothness on the first derivative of the solution to achieve results superior to those obtained using a conventional nonconvex prior that imposed piecewise smoothness on the zeroth derivative. In spite of several advantages of the higher-order model – the weak plate, its use in routine applications has been hindered by several factors, such as the computational expense due to the nonconvexity of its penalty function and the difficulty in the selection of hyperparameters involved in the model. We note that, by choosing a penalty function which is nonquadratic but is still convex, both the problem of nonconvexity involved in some nonquadratic priors and the oversmoothness of edge regions in quadratic priors may be avoided. In this paper, we use a class of two-dimensional smoothing splines with first and second spatial derivatives applied to convex-nonquadratic penalty functions. The convex-nonquadratic priors used in this work provide not only the edge-preserving ability, but also a globally convergent optimization algorithm. To validate the quantitative performance of our new priors, we use the quantitation of bias/variance and total squared error over noise trials using the Monte Carlo method. Our experimental results show that a linear combination of low (first) and high (second) orders of spatial derivatives applied to convex-nonquadratic penalty functions improves the reconstruction in terms of total squared error.

Keywords: Emission tomography, Penalized-likelihood reconstruction, Bayesian method, Gibbs priors

1. INTRODUCTION

The primary problem in emission computed tomography (ECT) is to reconstruct spatial distribution of a radionuclide in a given cross section of the human body from the observed projection measurements. Unfortunately, since the observed data in ECT systems are contaminated by noise due to low count rate and several physical factors, it has been a difficult problem to reconstruct images with good accuracy. For example, the quality of reconstructed images is poor when the measurements are assumed to be made with infinite accuracy, which is the case for the conventional filtered backprojection (FBP) method. Statistical reconstruction methods, on the other hand, are known to overcome these problems by naturally expressing accurate system models of physical effects and by accurately modeling the statistical character of the data. A significant advance for these methods was made with the introduction of the well-known ML-EM (maximum-likelihood expectation-maximization) approach. Unfortunately, however, this approach also suffers from its instability due to the ill-posed nature of the reconstruction problem.

Recently, penalized likelihood methods in the context of a Bayesian maximum *a posteriori* (MAP) framework have proven useful for problems in image reconstruction for ECT, such as single-photon emission computed tomography (SPECT) and positron emission tomography (PET). In these methods, the Bayesian model is especially apt for two reasons: (i) Accurate statistical modeling, via the likelihood function, of the Poisson noise associated with gamma-ray projection data is essential since noise is severe, and (ii) Given the ill-posed nature of tomographic inversion and the poor quality of projection data, a regularization needed to stabilize ML solutions may be provided by MAP estimates that can be formulated using Bayesian approaches that incorporate suitable prior models.

Further author information:

Email: sjlee@mail.paichai.ac.kr; Telephone: +82-42-520-5711; Fax: +82-42-520-5663

The prior information in Bayesian approaches reflects assumptions about the spatial properties of the underlying source distribution. Conventional smoothing priors use quadratic penalties, which are mathematically simple and lead to an algorithm whose solution is easier to compute. However, the conventional quadratic prior, such as the membrane prior (MM), tends to oversmooth discontinuities and incurs large bias error. To overcome this problem, we introduced a simple modification of the MM prior to one less sensitive to variations in first spatial derivatives - the thin plate (TP),² in which Gibbs potentials were applied to spatial *second* derivative terms better able to model gradual transition regions that separate anatomical regions. When the performance of TP was compared with that of MM in terms of quantitation of bias and variance over noise trials using the Monte Carlo method, the TP prior yielded improved reconstructions in the sense of low bias at little change in variance. It was also observed that the use of the TP prior revealed considerably less sensitivity in bias to the variations of smoothing parameter than the MM prior.³ In spite of the improvement of the MM prior to the TP prior, its quadratic penalty function has a fundamental limitation in the preservation of discontinuities.

In recent years, there has been considerable interest in edge-preserving prior models, which are associated with smoothing penalty functions that are *nonquadratic* functions of nearby pixel differences. Nonquadratic priors may be classified into two categories depending on the type of their penalty functions – nonconvex priors and convex priors. In our early work,¹ we proposed nonconvex prior models, such as the weak membrane and the weak plate derived from properties of ideal physical materials. Here, the weak membrane (WM) is a low order (1st derivative) spline allowed to have breaks (0th-order discontinuities) and the weak plate (WP) is the thin-plate spline that is allowed to have breaks and creases (discontinuities in both 0th and 1st derivatives). The weak plate is thus an extension of weak membrane to higher order. When tested on digital “ground-truth” phantoms derived from autoradiographic data, the WP prior illustrated superior quantitative performance; for comparable performance to ML-EM and WM in bias, WP yielded reduced variance in most of regions including edge regions.¹ In spite of such good performance of the WP prior, any enthusiasm for the use of this model is tempered by the fact that nonconvex priors suffer difficulties in optimization since an annealing process, such as deterministic annealing based on the continuation method or simulated annealing, has to be incorporated into the algorithm to achieve a good solution.

In this paper, we investigate a class of convex-nonquadratic penalty functions, which not only permit the recovery of discontinuities in the reconstructed image, but also provide for globally convergent optimization algorithms. Our motivation of considering such penalty functions stems from the comparison of the quantitative performance of the weak plate and the thin plate priors reported in^{1,2}; the advantages of each prior model in our previous work encouraged us to investigate whether those advantages might be merged in one form of the prior model by choosing a convex version of nonquadratic penalty functions. We note that, by choosing a penalty function which is nonquadratic but is still convex, both the problem of nonconvexity involved in some nonquadratic priors and the oversmoothness of edge regions in quadratic priors may be avoided.

The remainder of this paper overviews the Bayesian MAP approach, develops new Bayesian reconstruction models with convex-nonquadratic priors, and presents our experimental results showing good performance of our new priors on a realistic digital phantom.

2. BAYESIAN MODELS WITH CONVEX-NONQUADRATIC PRIORS

Before preceding to present our new prior models, we first define our notation. We model our problem on a two-dimensional (2-D) finite discrete lattice indexed by (i, j) . Lowercase bold quantities denote 2-D vector fields and the corresponding lowercase, italicized quantities denote the elements of the vector field. Similarly, uppercase bold quantities denote 2-D random fields. Thus, $\Pr(\mathbf{F} = \mathbf{f})$ denotes the probability that the random field \mathbf{F} takes the value \mathbf{f} and f_{ij} denotes an element of \mathbf{f} at the location (i, j) .

The MAP approach in the context of a Bayesian framework is to estimate the underlying source field \mathbf{f} by maximizing the posterior probability, given as

$$\Pr(\mathbf{F} = \mathbf{f} \mid \mathbf{G} = \mathbf{g}) = \frac{\Pr(\mathbf{G} = \mathbf{g} \mid \mathbf{F} = \mathbf{f})\Pr(\mathbf{F} = \mathbf{f})}{\Pr(\mathbf{G} = \mathbf{g})}, \quad (1)$$

where \mathbf{f} and \mathbf{g} are 2-D vector fields for the source intensities and projection data, respectively, and \mathbf{F} and \mathbf{G} are the associated random fields. Given the posterior distribution in (1), maximizing the posterior distribution is equivalent to minimizing $-\log$ of the posterior probability and the MAP estimation reduces to

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} [-\log \Pr(\mathbf{G} = \mathbf{g} \mid \mathbf{F} = \mathbf{f}) - \log \Pr(\mathbf{F} = \mathbf{f})], \quad (2)$$

where the two terms on the right side are the likelihood and the prior, respectively. For the likelihood, Poisson statistics are applied in a conventional way, which is given by

$$\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) = \prod_{t\theta} \frac{\bar{g}_{t\theta}^{g_{t\theta}} \exp(-\bar{g}_{t\theta})}{g_{t\theta}!}, \quad (3)$$

where $\bar{g}_{t\theta} \stackrel{\text{def}}{=} \sum_{ij} \mathcal{H}_{t\theta,ij} f_{ij}$. In (3), $g_{t\theta}$ is the number of detected counts and $\bar{g}_{t\theta}$ is the expected number of counts for a particular source \mathbf{f} and a forward projection matrix $\mathcal{H}_{t\theta,ij}$. This projection is from source location (i, j) to detector bin t at angle θ .

To incorporate our new priors in a MAP approach, we model the prior probability as a Gibbs distribution given as

$$\Pr(\mathbf{F} = \mathbf{f}) = \frac{1}{Z} \exp[-\lambda E_P(\mathbf{f})],$$

where λ the positive hyperparameter that weights the prior relative to the likelihood term, Z a normalization of no concern here. The function $E_P(\mathbf{f})$ is the sum of energies of individual ‘‘cliques’’ which are defined by adjacent pixels in a local neighborhood \mathcal{N}_{ij} of a pixel located at (i, j) :

$$E_P(\mathbf{f}) = \sum_{ij} \sum_{(i',j') \in \mathcal{N}_{ij}} \phi(f_{ij} - f_{i'j'}). \quad (4)$$

The penalty (or potential) function $\phi(\cdot)$ is usually defined so that its argument $(f_{ij} - f_{i'j'})$ is a discrete first spatial derivative. In order to model gradual transition regions, however, the argument of the potential function can also take the form of a discrete *second* spatial derivative, which will be discussed later in this section. One widely used penalty function takes a simple quadratic form (e.g., $\phi(\xi) = \xi^2$). Although the priors derived from the quadratic penalty function can stabilize the unstable ML problem, they have no ability to model larger changes, such as the sharp edges, in activity that may occur between regions corresponding to different tissue types.

In order to explicitly model the existence of edges or boundaries, Geman and Geman⁴ introduced the powerful idea of including unobservable ‘‘line processes’’ in the image model. Versions of the line processes have been proposed for medical imaging.^{5-7,1} The line processes are binary variables (0 or 1) that act to suspend smoothness constraints at sites where they are turned on. By embedding the line processes into the image model, conventional smoothness constraints using the quadratic penalty function are extended to include *weak* continuity constraints.⁸ In our early work,¹ we showed that the the existence of the line processes can be expressed implicitly in a *nonconvex* penalty function, which is often referred to as the ‘‘broken parabola’’ function. In this case, the problem of minimization over both the continuous source intensities and the binary line processes reduces to a minimization over the source intensities only.

Table 1. Convex-Nonquadratic Penalty Functions

	Penalty Function
Huber (HB)	$\phi_{HB}(\xi) = \begin{cases} \xi^2, & \xi \leq \alpha \\ 2\alpha \xi - \alpha^2, & \xi > \alpha \end{cases}$
Bouman & Sauer (BS)	$\phi_{BS}(\xi) = \xi ^\alpha, (1.0 \leq \alpha \leq 2.0)$
Green (GR)	$\phi_{GR}(\xi) = \log \left(\cosh \left(\frac{\xi}{\alpha} \right) \right)$
Lange (LN)	$\phi_{LN}(\xi) = \frac{1}{2} \left(\xi/\alpha + \frac{1}{1+ \xi/\alpha } - 1 \right)$

Although the above nonconvex priors can accurately model the edges or boundaries, they suffer from difficulties in optimization as described in Sec. 1. One possible way of retaining the edge-preserving ability of nonconvex priors, while keeping the simplicity of convex priors in optimization, may be to choose a convex-nonquadratic penalty function. Representative convex-nonquadratic (CNQ) penalty functions, which have been widely used in computer

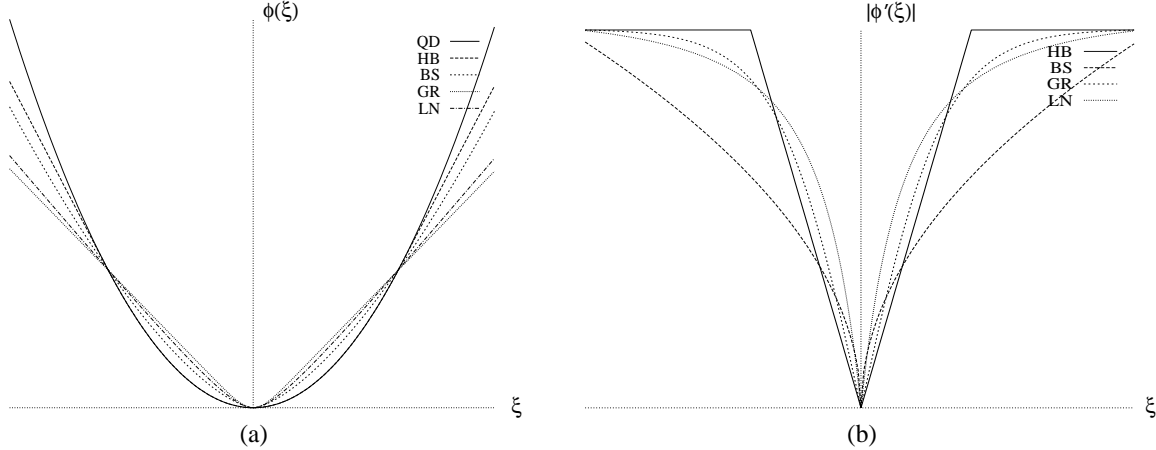


Figure 1. Qualitative plots for the shapes of convex penalty functions (a) and magnitudes of their derivatives (b).

vision and medical imaging, are listed in Table 1 as ϕ_{HB} , ϕ_{BS} , ϕ_{GR} , and ϕ_{LN} , which are the function of Huber from robust statistics⁹ (referred to hereafter as HB), the priors adopted by Bouman and Sauer¹⁰ (BS), Green¹¹ (GR), and Lange¹² (LN), respectively. The qualitative shapes of these functions are plotted in Fig. 1(a). Compared to the quadratic penalty function, one can qualitatively notice that the penalty functions listed in Table 1 and plotted in Fig. 1 provide a degree of edge preservation by penalizing large pixel differences less than small pixel differences.

Stevenson *et al.*¹³ summarizes the properties of convex-nonquadratic penalty functions that can have good behaviors for edge preservation as follows: (i) convex, (ii) symmetric, (iii) $\phi(\xi) < \xi^2$ for large $|\xi|$, and (iv) controlled by a parameter. However, it may be more intuitive to interpret CNQ penalties using the derivative of $\phi(\xi)$ rather than using the penalty function itself since the magnitude of $\phi'(\xi) = \frac{d\phi(\xi)}{d\xi}$ specifies the strength of smoothing. Li¹⁴ presents an extensive study on the convex-nonquadratic priors by using a function derived from $\phi'(\xi)$. In general, a CNQ function must have bounded influence in order to preserve edges. Therefore, the necessary condition for a CNQ penalty function to preserve edges may be summarized as

$$\lim_{\xi \rightarrow \infty} \left| \frac{d\phi(\xi)}{d\xi} \right| = C \quad (5)$$

where $C \geq 0$ is a constant. When $C = 0$, smoothing at discontinuities is entirely prohibited, which is the case for nonconvex priors such as the broken parabola function¹ used in our early work. When $C > 0$, on the other hand, limited smoothing is allowed at discontinuities, which is the case for the CNQ priors discussed here. There are many other CNQ functions that can be useful for edge-preserving reconstruction. In this paper, however, we restrict attention to the ϕ functions described in Table 1. The details on more restrictive conditions for well-defined algorithms and global convergence can be found in.⁹

According to the quantitative results reported in,¹⁴ the Huber function is the most favorable in edge preservation, when it is applied to synthetic data with ideal step edges. In fact, this function has been useful in applications to computer vision, such as optical flow¹⁵ and surface reconstruction.¹³ Unlike the Huber function, where the magnitude of $\phi'(\xi)$ is strictly constant for $|\xi| > \alpha$, the magnitude of $\phi'(\xi)$ of both ϕ_{GR} and ϕ_{LN} converges slowly. These two penalty functions are qualitatively similar to each other and have been used for Bayesian priors in ECT reconstruction.^{11,12}

Unlike the three CNQ functions described above, ϕ_{BS} does not strictly satisfy the necessary condition for edge preservation in (5), unless $\alpha = 1$ (i.e., $\phi_{BS}(\xi) = |\xi|$), and may not be as good as other functions listed in Table 1 for the recovery of ideal step edges. However, it is important to point out that, although ϕ_{BS} does not strictly satisfy the edge-preserving condition, it may work well for our tomographic reconstruction problem in which ideal step edges are hardly found. In fact, according to our quantitative comparison, ϕ_{BS} turns out to be the most favorable in terms of total squared error over noise trials using the Monte Carlo method (see Sec. 4).

Since it is often important to regularize gradual transition regions as well as flat regions, we may consider higher-order models for CNQ priors as done in our previous work for quadratic priors² (e.g., the thin plate) and for nonconvex

priors¹ (e.g., the weak plate), where the higher-order models were motivated by our observations of ground-truth spatial variation of radionuclide density in primate autoradiography.^{16,1} A general form of the second-order prior energy may be written as

$$E_P(\mathbf{f}) = \sum_{ij} \phi \left(\sqrt{V_{ij}} \right), \quad (6)$$

where V_{ij} is the discrete form of quadratic variation, which is given by

$$V_{ij} \stackrel{\text{def}}{=} f_{hh}^2(i, j) + 2f_{hv}^2(i, j) + f_{vv}^2(i, j).$$

Here, $f_{vv}(i, j)$ and $f_{hh}(i, j)$ denote the discrete second partial derivatives of the source distribution in the vertical and horizontal directions, respectively, and $f_{hv}(i, j)$ is the second partial cross derivative. Our choices for discretization of the derivatives are:

$$\begin{aligned} f_{hh}(i, j) &\stackrel{\text{def}}{=} f_{i,j+1} - 2f_{i,j} + f_{i,j-1} \\ f_{vv}(i, j) &\stackrel{\text{def}}{=} f_{i+1,j} - 2f_{i,j} + f_{i-1,j} \\ f_{hv}(i, j) &\stackrel{\text{def}}{=} f_{i+1,j+1} - f_{i+1,j} - f_{i,j+1} + f_{i,j}. \end{aligned}$$

When the above second partial derivatives are applied to the CNQ penalty functions listed in Table 1, they encourage smoothness even in ramplike regions without incurring a penalty. Therefore, the discontinuities in this case correspond to discontinuities in the source gradient in addition to those in the source itself. The major difference between our new convex version of higher-order model and the nonconvex version of higher-order model, such as the weak plate, is that, while the nonconvex version entirely prohibits smoothing discontinuities (in both 0th and 1st spatial derivatives), our convex version allows limited smoothing at discontinuities.

One could also consider a generalized form of the prior energy by a linear combination of the first-order energy (E_{P_1}) and the second-order energy (E_{P_2}), which is given by

$$E_P(\mathbf{f}) = (1 - \tau)E_{P_1}(\mathbf{f}) + \tau E_{P_2}(\mathbf{f}), \quad (7)$$

where $\tau \in [0, 1]$ is the weight parameter; $\tau = 0$ reduces (7) to the first-order energy and $\tau = 1$ to the second-order energy. For example, a linearly combined CNQ prior energy with ϕ_{BS} as its penalty function (referred to hereafter as CNQ-BS) may be written as

$$E_P(\mathbf{f}) = (1 - \tau) \sum_{ij} \left[\left(\sqrt{f_h^2(i, j)} \right)^\alpha + \left(\sqrt{f_v^2(i, j)} \right)^\alpha \right] + \tau \sum_{ij} \left[\sqrt{f_{hh}^2(i, j) + 2f_{hv}^2(i, j) + f_{vv}^2(i, j)} \right]^\alpha, \quad (8)$$

where $f_v(i, j)$ and $f_h(i, j)$ are the discrete first partial derivatives of the source distribution in the vertical and horizontal directions, respectively, and the equality, $|\xi|^\alpha = \left(\sqrt{\xi^2} \right)^\alpha$, is used for both E_{P_1} and E_{P_2} . The CNQ prior energy with second partial derivatives in (8) can also be written so that each of the three penalty functions is associated with each second partial derivative.¹⁷ Note that the case for $\alpha = 2$ in (8) corresponds to the quadratic prior energy with a linear combination of the membrane and the thin plate energies.

Our task now reduces to finding the minimum of the following overall energy:

$$E(\mathbf{f}) = E_L(\mathbf{f}) + \lambda E_P(\mathbf{f}), \quad (9)$$

where $E_L(\mathbf{f})$ is the likelihood energy defined by the first term in the right side of (2) and $E_P(\mathbf{f})$ is the prior energy defined by (6). A variety of algorithms are available to find the minimum of the overall energy in (9). However, since we are evaluating and comparing the performance of several different priors listed in Table 1, we use the iterative MAP-EM OSL (one-step-late) algorithm derived by Green¹¹ for simplicity. The OSL algorithm is not derivable directly from a MAP principle, but can be shown to converge to the MAP solution if it converges at all. The OSL algorithm is given by:

$$\hat{f}_{ij}^{k+1} = \frac{\hat{f}_{ij}^k}{\sum_{t\theta} \mathcal{H}_{t\theta, ij} + \lambda \left. \frac{\partial E_P(\mathbf{f})}{\partial f_{ij}} \right|_{f_{ij} = \hat{f}_{ij}^k}} \sum_{t\theta} \frac{\mathcal{H}_{t\theta, ij} g_{t\theta}}{\sum_{kl} \mathcal{H}_{t\theta, kl} \hat{f}_{kl}^k},$$

where \hat{f}_{ij}^k is the object estimate at location (i, j) and iteration k . Note that the derivative energy function $\frac{\partial E_P(\xi)}{\partial \xi_{ij}}$ in the above update equation acts as a penalty term to enforce the conditions required by the prior; the larger the magnitude of $\frac{\partial \phi(\xi)}{\partial \xi_{ij}}$, the smaller the estimate of \hat{f}_{ij} becomes, thereby suppressing pixel intensity differences. Unlike the quadratic priors, where the increment of the penalty with respect to the increment of the pixel difference is the same for all pixel differences, CNQ priors have less increment of the penalty for large pixel differences as shown in Fig. 1(b). Note also that the familiar ML-EM algorithm is obtained simply by setting $\lambda = 0$.

3. SIMULATION PROCEDURE FOR QUANTITATIVE PERFORMANCE TEST

To evaluate and compare the quantitative performance of the CNQ priors listed in Table 1, we used a 128×128 digital phantom derived from primate autoradiograph.¹⁸ Note that the phantom shown in Fig. 2 contains a variety of realistic edge structures. We generated 50 Monte Carlo noise trials by adding independent realizations of Poisson noise to the noiseless projection data obtained from the phantom. The total number of projection counts in this case was approximately 500,000.

Since our algorithms are iterative, we needed to choose a sufficient number of iterations after which the change in reconstruction was negligible. Thus, iteration number was removed as a parameter in comparisons. For the choice of hyperparameters involved in our MAP algorithms, such as λ and α , while this is an area of active research, we simply chose the values that led to a best reconstruction in terms of minimum root-mean squared error (RMSE) relative to the phantom. A Hanning window was applied to FBP for high-frequency apodization and was also adjusted to lead to a minimum RMSE. For the ML-EM algorithm, we chose two stopping criteria and designated the resulting reconstructions as EM-1 and EM-2, respectively. EM-1 was chosen by observing the iteration number at which reconstructions minimized RMSE. Since EM-1 yields only one choice, we also included a second criterion based on the simple heuristic of choosing the iteration number that optimized qualitative resemblance of reconstruction and phantom, based on our own subjective impression.

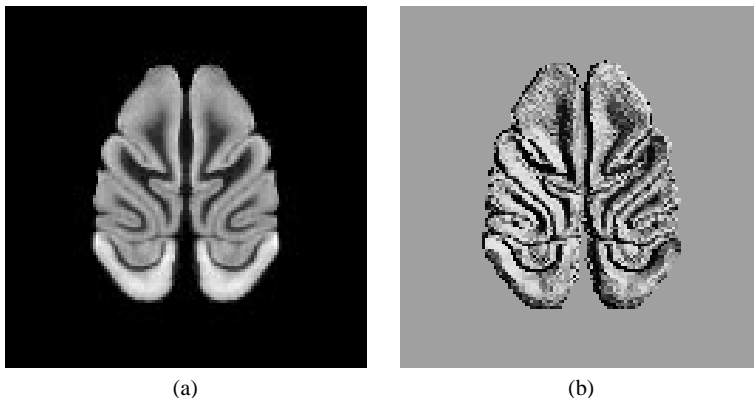


Figure 2. The digital autoradiograph phantom used in the simulation (a) and its surface plot (b).

To evaluate the reconstructions quantitatively, we computed bias and standard deviation (STD) images. A bias image, b_{ij} , is defined as

$$b_{ij} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N (\hat{f}_{ij}^n - f_{ij}) \quad (10)$$

where \hat{f}_{ij}^n is the n th reconstruction of phantom f at the location (i, j) and the summation is over $N = 50$ independent noise trials. To display the bipolar bias image, an intermediate grey scale value of 128 out of 256 levels was used as zero bias. A standard deviation image, s_{ij} , is defined as

$$s_{ij} \stackrel{\text{def}}{=} \sqrt{\frac{1}{N-1} \sum_{n=1}^N (\hat{f}_{ij}^n - \bar{f}_{ij})^2} \quad (11)$$

where \bar{f}_{ij} is the mean of \hat{f}_{ij} over the noise trials defined as $\bar{f}_{ij} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \hat{f}_{ij}^n$.

4. RESULTS

Figures 3 and 4 show anecdotal reconstructions for each of the estimators (FBP, ML-EM, and MAP with several different priors). Figure 3(a) shows an FBP result, (b) and (c) the typical behavior of ML-EM for different iteration numbers. For MAP reconstructions, we used the 15 different priors; three different values of τ in (7) (the first order ($\tau = 0$), a combination of the first and the second orders ($\tau = 0.5$), and the second order ($\tau = 1$)) were applied to the four penalty functions listed in Table 1 and to the quadratic (QD) function. In Fig. 4 the transition from the first row to the third row shows qualitative effects of using higher-order spatial derivatives applied to each of the penalty functions. For quadratic priors, while the MM prior tends to oversmooth the discontinuities and incurs large bias error, the TP prior exhibits overshoots around discontinuities as reported in.¹⁹ On the other hand, the combination of MM and TP using $\tau = 0.5$ in (7) seems to compromise the two extreme cases (see Fig. 4(b)). Comparison of the results in the first column (Fig. 4(a)-(c)) with those in the second column (Fig. 4(d)-(f)) shows that the use of CNQ penalty functions enhances edges in the reconstructed images. For the noise level, the phantom, and the values of hyperparameters used in our experiments, the visual quality of reconstructions at a fixed value of τ looks similar. For example, while the first-order spatial derivatives ($\tau = 0$) applied to CNQ functions exhibit sharp edges yielding reconstructions which are often artificially patchy in gradual transition regions, the second-order spatial derivatives ($\tau = 1$) preserve those regions well, though they are not as good as the cases for the first-order derivatives in edge preservation. In addition, as in the case for the quadratic priors, the linear combination of the first- and the second-order spatial derivatives using $\tau = 0.5$ compromises the above two extreme cases.

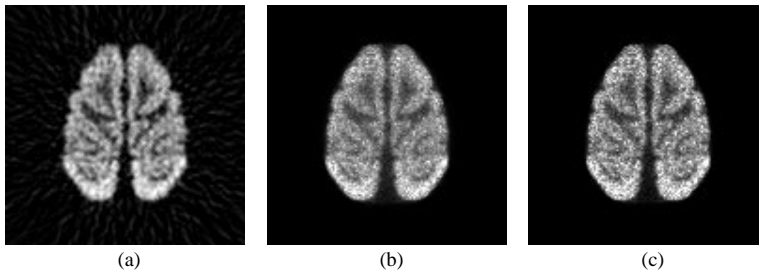


Figure 3. Anecdotal reconstructions for FBP and ML-EM: (a) FBP; (b) EM-1 (20 iterations); (c) EM-2 (30 iterations).

Figures 5 and 6 show our bias and STD images. Bias images are bipolar, with a value of zero displayed as an intermediate grey, and with darker/lighter regions corresponding to negative/positive bias. In STD images, intensity (lighter means greater) codes the positive STD value. A given figure comprises images displayed with same grey scale to allow visual comparisons. In general, the bias images show a negative bias for high-signal regions and a positive bias for low signal regions.

Figure 5 shows pointwise bias/STD images for FBP and ML-EM, where (b)(c) and (e)(f) show the usual bias/variance tradeoff inherent in ML-EM; fewer iterations lead to lower variance but larger bias, and the opposite is true for the larger number of iterations. Note that the FBP algorithm reveals the patterning artifact in its bias image, and “spreads” variance across the entire image area.

Figure 6 shows pointwise bias/STD images for MAP with 9 different priors – the three different values of τ ($\tau = 0$, $\tau = 0.5$, $\tau = 1$) for each of the three penalty functions (ϕ_{QD} , ϕ_{BS} , ϕ_{GR}). Since the results obtained from both the ϕ_{HB} and ϕ_{LN} functions were visually similar to those from ϕ_{BS} in our experiments, only the results from ϕ_{BS} are reported in Figure 6. (The bias/variance behavior of both CNQ-HB and CNQ-LN described below thus follows that of CNQ-BS.) In general, the transition of bias images from $\tau = 0$ to $\tau = 1$ for CNQ priors is similar to that of bias images for quadratic priors; the first-order derivatives ($\tau = 0$) yield relatively larger bias error than the second-order derivatives ($\tau = 1$), and the combination ($\tau = 0.5$) of the first-order and the second-order derivatives compromises the bias errors for $\tau = 0$ and $\tau = 1$.

Interesting results are seen in the case of STD images. The STD image for CNQ-BS with $\tau = 0$ shows high variance in sharp edge regions. This effect appears to be due to the fact that the first order derivatives applied to the CNQ-BS penalty function results in unstable estimates of edge location; in different noise realizations, the locations of continuity breaks can shift. In fact, the similar effect in STD images was also observed when the weak membrane

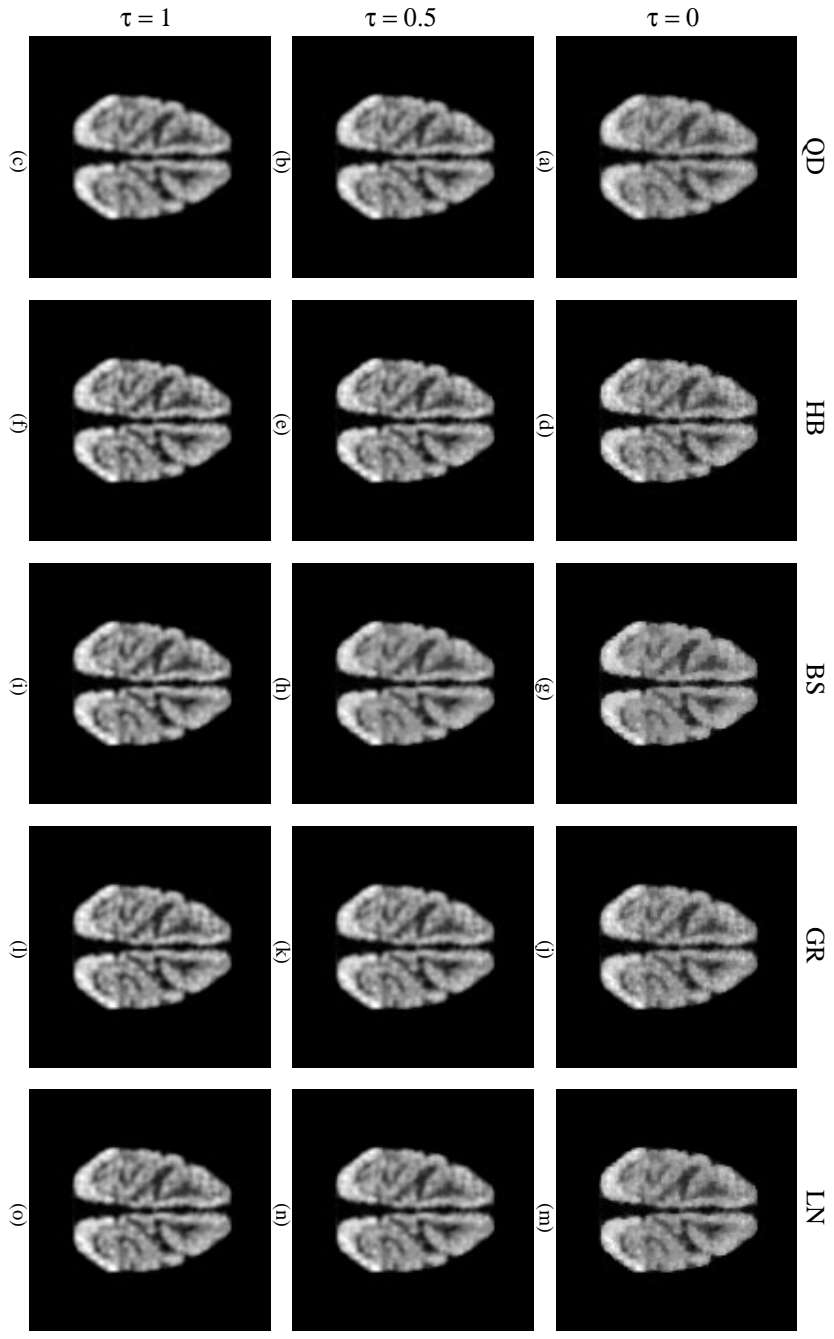


Figure 4. Anecdotal MAP reconstructions with 15 different priors using 3 different values of τ ($\tau = 0, 0.5, 1$) and 5 different penalty functions ($\phi_{QD}, \phi_{HB}, \phi_{BS}, \phi_{GR}, \phi_{LN}$): (a)-(c) quadratic (QD); (d)-(f) CNQ-HB; (g)-(i) CNQ-BS; (j)-(l) CNQ-GR; (m)-(o) CNQ-LN.

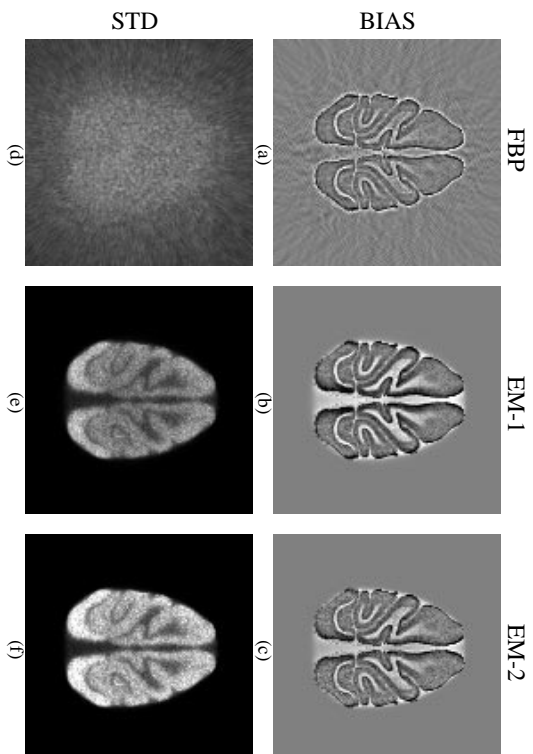


Figure 5. Pointwise bias and STD images for FBP and ML-EM reconstructions: (a),(b),(c), bias images for FBP, EM-1, and EM-2, respectively; (d),(e),(f), STD images for FBP, EM-1, and EM-2, respectively.

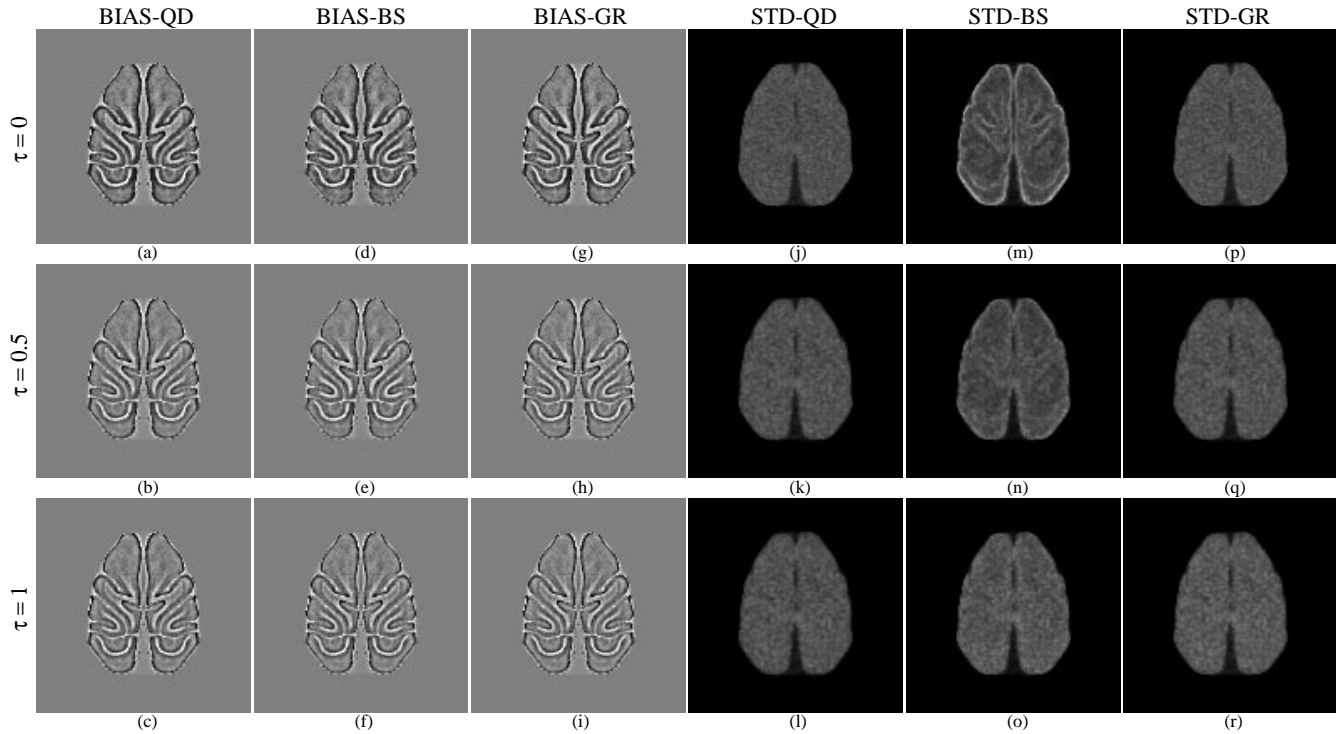


Figure 6. Pointwise bias and STD images for MAP reconstructions with 9 different priors using 3 different values of τ ($\tau = 0, 0.5, 1$) and 3 different penalty functions ($\phi_{QD}, \phi_{BS}, \phi_{GR}$): (a)-(i) bias images; (j)-(r) STD images. The first row is for $\tau = 0$, the second row for $\tau = 0.5$, and the third row for $\tau = 1$.

prior was used in our early work.¹ Compared to the results from the weak membrane prior, however, two different aspects of the CNQ-BS prior with $\tau = 0$ may be noted here. First, since the nonconvex broken parabola function¹ for the WM prior entirely prohibits smoothing at discontinuities, bias error can be very large if ramplike steep edges are treated as 0th-order discontinuities, which can also yield high variance for edge shifts in different noise realizations. On the other hand, variance for CNQ priors in edge regions is not as high as that for nonconvex priors since CNQ priors allow limited smoothing at discontinuities. Second, the bias for CNQ priors with $\tau = 0$ appears relatively more correlated with image intensity than the WM case, which is also true for quadratic priors. This is presumably due to the fact that, unlike nonconvex priors, the effect of (limited) smoothing at discontinuities in CNQ priors affects low-contrast regions near the discontinuities and incurs relatively larger bias error. In other words, the bias behavior of CNQ priors is inherited from quadratic priors.

Although the CNQ-GR priors theoretically have the edge-preserving ability as other CNQ priors, our results shown in Fig. 6(p)-(r) are similar to those for the quadratic priors in Fig. 6(j)-(l). In other words, for the given noise level and the values of hyperparameters, the CNQ-GR priors act like quadratic priors without showing their edge-preserving behavior.

To compare our results more quantitatively, we computed total squared error (TSE), t^2 , which is defined by

$$t^2 \stackrel{\text{def}}{=} \sum_{ij} (b_{ij}^2 + s_{ij}^2),$$

where b_{ij} and s_{ij} are the bias and standard deviation quantities defined in (10) and (11), respectively. Table 2 summarizes the total squared errors for the 18 different algorithms used in our experiments. Note that, for a given penalty function, $\tau = 0.5$ reveals the smallest TSE in MAP reconstructions. It is also interesting to note that, for a given penalty function, TSE for $\tau = 1$ is always smaller than that for $\tau = 0$. According to the above quantitative results, the CNQ-BS prior with $\tau = 0.5$ seems to be the most favorable among the priors considered in our experiments in terms of total squared error.

Table 2. Total squared errors of reconstruction noise trials

	Quadratic	CNQ-HB	CNQ-BS	CNQ-GR	CNQ-LN	FBP	EM-1	EM-2
MAP ($\tau = 0$)	127.780	128.376	121.931	127.020	121.619	-	-	-
MAP ($\tau = 0.5$)	121.644	120.440	114.596	121.272	118.173	-	-	-
MAP ($\tau = 1$)	122.299	124.874	121.432	123.651	121.535	-	-	-
Others	-	-	-	-	-	164.586	234.139	296.309

5. SUMMARY AND CONCLUSION

We have considered new convex-nonquadratic priors, which provide various advantages over conventional quadratic priors and nonconvex priors. Although the CNQ priors considered here do not preserve edges as well as those derived from nonconvex penalty functions, such as the weak membrane and the weak plate, their bias/variance behavior turns out to be a good compromise between nonconvex priors and quadratic priors. In other words, since the CNQ penalty functions do not entirely prohibit smoothing at discontinuities, the variance of CNQ priors around edges over the different noise realizations is not as high as that of nonconvex priors, whereas their bias behavior is inherited from quadratic priors. For example, our experimental results show that the CNQ-BS prior with $\tau = 0$ compromises the membrane and the weak membrane, and the CNQ-BS prior with $\tau = 1$ compromises the thin plate and the weak plate in the quantitation of bias and variance.

We have also experimentally shown that the generalization of the prior energy using a control parameter (τ) provides additional information compared to the above two extreme constituent models. According to our numerical results, when tested with the three different values of τ ($\tau = 0, 0.5, 1$), $\tau = 0.5$ always revealed the smallest total squared error for a given penalty function, which indicates that a combination of low (first) and high (second) orders of a CNQ prior exhibits improved reconstruction in terms of total squared error.

Since our evaluations were made using stopping rules and values of hyperparameters based on knowledge of the known phantom, results would differ if the hyperparameters had to be estimated based only on observed projection data.²⁰ Our future work will include the use of meaningful task-specific criteria for choices of number of iterations (for EM) and λ and α (for MAP), the implementation of a parallelizable algorithm, which can easily accommodate our CNQ priors with a moderate number of iterations, and compare models via results in terms of meaningful task quantities such as quantitation metrics.

ACKNOWLEDGMENTS

This work was supported in part by a grant (HMP-98-E-1-0008) of the Good Health R&D Project from the Ministry of Health and Welfare, Korea.

REFERENCES

1. S.J. Lee, A. Rangarajan, and G. Gindi, "Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors", *IEEE Trans. Med. Imaging*, MI-14(4), pp. 669–680, Dec. 1995.
2. S.J. Lee, I.T. Hsiao, and G.R. Gindi, "The Thin Plate as a Regularizer in Bayesian SPECT Reconstruction", *IEEE Trans. on Nuclear Science*, NS-44(3), pp. 1381–1387, Jun. 1997.
3. S.J. Lee, I.T. Hsiao, and G.R. Gindi, "Quantitative Effects of Using Thin-Plate Priors in Bayesian SPECT Reconstruction", In *Proc. SPIE Image Reconstruction and Restoration II*, pp. 252–263, Jul. 1997.
4. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6(6), pp. 721–741, November 1984.
5. V. E. Johnson, W. H. Wong, X. Hu, and C. T. Chen, "Image Restoration Using Gibbs Priors: Boundary Modeling, Treatment of Blurring, and Selection of Hyperparameter", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(5), pp. 413–425, May 1991.

6. R. Leahy and X. Yan, "Incorporation of Anatomical MR Data for Improved Functional Imaging with PET", In A. C. F. Colchester and D. J. Hawkes, editors, *Information Processing in Medical Imaging*, pp. 105–120, Springer–Verlag, 1991.
7. G. Gindi, M. Lee, A. Rangarajan, and G. Zubal, "Bayesian Reconstruction of Functional Images Using Anatomical Information as Priors", *IEEE Trans. Med. Imaging*, MI-12, pp. 670–680, Dec. 1993.
8. G. E. Hinton, *Relaxation and Its Role In Vision*, PhD thesis, University of Edinburgh, 1977.
9. P. Huber, *Robust Statistics*, John Wiley, New York, NY, 1981.
10. C. Bouman and K. Sauer, "A Generalized Gaussian Image Model for Edge Preserving MAP Estimation", *IEEE Trans. Image Processing*, 2, pp. 296–310, July 1993.
11. P.J. Green, "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm", *IEEE Trans. Med. Imaging*, MI-9(1), pp. 84–93, March 1990.
12. K. Lange, "Convergence of EM Image Reconstruction Algorithms with Gibbs Smoothing", *IEEE Trans. Med. Imaging*, MI-9(4), pp. 439–446, December 1990.
13. R.L Stevenson, B.E. Schmitz, and E.J. Delp, "Discontinuity Preserving Regularization of Inverse Visual Problems", *IEEE Trans. on Systems, Man and Cybernetics*, SMC-24, pp. 455–469, Mar. 1994.
14. S.Z. Li, "Close-Form Solution and Parameter Selection for Convex Minimization-Based Edge-Preserving Smoothing", *IEEE Trans. Patt. Anal. Mach. Intell.*, 20(9), pp. 916–932, Sep. 1998.
15. D. Shulman and J.Y. Herve, "Regularization of Discontinuous Flow Fields", In *Proc. Workshop Visual Motion*, pp. 81–86, 1989.
16. S. J. Lee, G. R. Gindi, I. G. Zubal, and A. Rangarajan, "Using Ground-Truth Data to Design Priors in Bayesian SPECT Reconstruction", In Y. Bizais, C. Barillot, and R. D. Paola, editors, *Information Processing in Medical Imaging*, pp. 27–38, Kluwer Academic Publishers, 1995.
17. D. Geman and G. Reynolds, "Constrained Restoration and the Recovery of Discontinuities", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-14(3), pp. 367–383, March 1992.
18. J. L. Lear, "Principles of Single and Multiple Radionuclide Autoradiography", In M. E. Phelps, J. C. Mazziotta, and H. R. Schelbert, editors, *Positron Emission Tomography and Autoradiography*, chapter 5, Raven Press, New York, NY, 1986.
19. M. Gokmen and A.K. Jain, " $\lambda\tau$ -Space Representation of Images and Generalized Edge Detector", *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(6), pp. 545–563, Jun. 1997.
20. Z. Zhou and R. Leahy, "Approximate Maximum Likelihood Parameter Estimation for Gibbs Priors", Technical Report TR-285, Signal and Image Processing Institute, University of Southern California, June 1995.