

A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography

J. Nuyts, D. Bequé, P. Dupont, L. Mortelmans^{1,2}

Abstract

A well known problem with maximum likelihood reconstruction in emission tomography is the excessive noise propagation. To prevent this, the objective function is often extended with a Gibbs prior favoring smooth solutions. We hypothesize that the following three requirements should produce a useful and conservative Gibbs prior for emission tomography: 1) the prior function should be concave to ensure that the posterior has a unique maximum; 2) the prior should penalize relative differences rather than absolute differences; 3) the prior should be tolerant for “large” differences between neighboring pixels. The second requirement should avoid tuning problems caused by the large dynamic range of activity values in the reconstructed image. A simple function has been derived that meets these three requirements. Our initial evaluations indicate that the prior behaves as intended.

Keywords— PET, maximum likelihood, maximum a posteriori reconstruction

I. INTRODUCTION

An effective and elegant way to suppress noise propagation in emission and transmission tomography is to implement the reconstruction using a maximum-a-posteriori approach. However, there is no theoretically correct way to implement the prior, since the available a-priori knowledge is intuitive and fuzzy. As a result, many different priors have been proposed. Usually, a Gibbs distribution is used which penalizes differences between actual pixel values and “ideal” values. The simplest ideal model is a uniform distribution, which requires penalizing differences between neighboring pixels. A less restrictive but more complex prior is obtained by encouraging smoothness of the first spatial derivative of the pixel values. If piecewise smoothness is preferred over global smoothness, the prior should be relatively tolerant for large differences, which are assumed to be due to the signal rather than to the noise.

In the case of transmission tomography, the amount of a-priori knowledge is considerable because the attenuation coefficients in the human body have a limited dynamic range, predictable values and are known to be locally smooth. Consequently, it is acceptable to use relatively restrictive priors. In contrast, the characteristics of the emission image are more difficult to predict, and the dynamic range is much larger than in transmission tomography. Thus, the design of a prior for emission tomography is a more delicate matter. Here, we only consider priors favoring piecewise smoothness, using the

uniform image as the ideal model.

The priors most often used penalize absolute differences between neighboring pixels. Hence, the logarithm of the prior can be written as:

$$M(\lambda) = -\beta \sum_j \sum_{k \in N_j} \Phi(\lambda_j - \lambda_k), \quad (1)$$

where λ_j is the activity in pixel j , N_j is the set of neighbors of pixel j and β controls the weight of the prior. Typical choices of the function $\Phi(x)$ are [1, 2]:

$$\text{Quadratic: } \frac{x^2}{2\sigma^2} \quad (2)$$

$$\text{Huber: } |x| < \sigma : \frac{x^2}{2\sigma^2}, \quad |x| > \sigma : \frac{|x| - \sigma/2}{\sigma} \quad (3)$$

$$\text{Geman: } \frac{x^2}{2\sigma^2 + x^2} \quad (4)$$

It is well known that a quadratic prior strongly smooths over edges, which is caused by the fact that the gradient of the prior is proportional to the difference between the pixel values. To avoid this, the Huber and Geman priors compare the difference between neighboring pixels with the value of a parameter σ . Values smaller than σ are considered small and for those values, the prior is nearly quadratic. Values larger than σ are considered large and the strength of the prior is decreased relative to that of the quadratic prior. The Geman prior is extremely tolerant for large edges (its gradient converges to zero), but as a result it is non-concave. The Huber prior is less tolerant but preserves concavity.

In transmission tomography, selection of σ is relatively straightforward, because the typical attenuation values are known in advance. In contrast, selecting σ is far from obvious in emission tomography. To avoid this problem, we propose a new prior which penalizes relative rather than absolute differences.

In the next section, we propose three requirements which are expected to yield a conservative and useful prior for emission tomography. In the third section, a prior meeting these requirements is described. The results of the evaluation on simulations and on a clinical image are presented in section four, and the work is discussed in section five.

II. REQUIREMENTS

We hypothesize that the following three requirements should result in a useful Gibbs prior for maximum-a-posteriori reconstruction in single photon and positron emission tomography:

1. the Gibbs prior is concave,

¹Nuclear Medicine, K.U.Leuven, B3000 Leuven, Belgium

²This work is supported by the Flemish Fund for Scientific Research (FWO), grant number G.0106.98, and by OT/00/32 of K.U.Leuven. P. Dupont is post-doctoral researcher of the FWO.

2. the Gibbs prior penalizes relative differences between neighboring pixel values,
3. the gradient of the Gibbs prior should level off for large differences between neighboring pixel values.

The first requirement ensures that the posterior has a unique maximum. This avoids dependence of the result on the initial image and on possible acceleration techniques.

The second requirement eliminates the problem of selecting parameter σ to discriminate between small and large absolute differences.

The third requirement is intended to obtain maximum tolerance for edges without sacrificing concavity, similar to the Huber prior.

III. THE GIBBS PRIOR

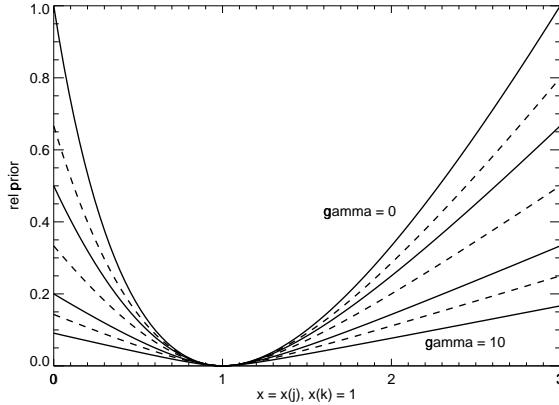


Figure 1: A plot of $-M_{jk}$ (5) as a function of λ_j , for $\lambda_k = 1$ and for different values of γ .

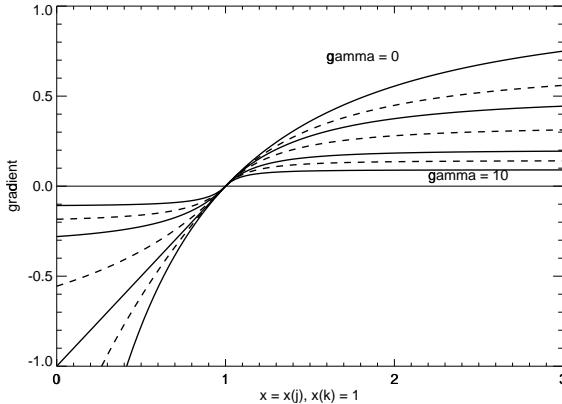


Figure 2: A plot of $-\frac{\partial M_{jk}}{\partial \lambda_j}$ (6) as a function of λ_j/λ_k , for different values of γ .

The logarithm of the new prior (figure 1) is defined as

$$\begin{aligned} M &= \sum_j \sum_{k \in N_j} M_{jk} \\ M_{jk} &= -\beta \frac{(\lambda_j - \lambda_k)^2}{(\lambda_j + \lambda_k) + \gamma |\lambda_j - \lambda_k|} \text{ if } k \in N_j \\ &= 0 \text{ otherwise,} \end{aligned} \quad (5)$$

where N_j is set of neighbors of pixel j , parameter β controls the weight of the prior and γ controls the shape. The gradient (figure 2) equals

$$\begin{aligned} \frac{\partial M_{jk}}{\partial \lambda_j} &= -\beta \frac{(\lambda_j - \lambda_k)(\gamma|\lambda_j - \lambda_k| + \lambda_j + 3\lambda_k)}{(\lambda_j + \lambda_k + \gamma|\lambda_j - \lambda_k|)^2} \\ &= -\beta \frac{(r_{jk} - 1)(\gamma|r_{jk} - 1| + r_{jk} + 3)}{(r_{jk} + 1 + \gamma|r_{jk} - 1|)^2}, \end{aligned} \quad (6)$$

where $r_{jk} = \lambda_j/\lambda_k$.

A. Concavity

The matrix of second derivatives contains the following elements:

$$\begin{aligned} \frac{\partial^2 M}{\partial \lambda_j^2} &= -\sum_{k \in N_j} \frac{16\lambda_k^2}{(\lambda_j + \lambda_k + \gamma|\lambda_j - \lambda_k|)^3} \\ \frac{\partial^2 M}{\partial \lambda_j \partial \lambda_k} &= \frac{16\lambda_j \lambda_k}{(\lambda_j + \lambda_k + \gamma|\lambda_j - \lambda_k|)^3} \text{ if } k \in N_j \\ &= 0 \text{ if } k \notin N_j \text{ and } k \neq j. \end{aligned} \quad (7)$$

It follows that

$$\begin{aligned} \sum_j \sum_k \frac{\partial^2 M}{\partial \lambda_j \partial \lambda_k} x_j x_k \\ = \sum_j \sum_{k \in N_j} \frac{-8(\lambda_k x_j - \lambda_j x_k)^2}{(\lambda_j + \lambda_k + \gamma|\lambda_j - \lambda_k|)^3} \leq 0, \end{aligned} \quad (8)$$

for all nonzero x_j, x_k and non-negative λ_j and λ_k . Consequently, the matrix of second derivatives is negative semi-definite, implying that the prior is concave.

In maximum-a-posteriori reconstruction for emission tomography, the logarithm of the posterior is obtained by adding the logarithms of the prior and the likelihood. The likelihood describes the Poisson nature of the emission data and its logarithm is concave under reasonable conditions [3]. Consequently, if also the prior is concave, the posterior does not have multiple local maxima.

B. Relative differences

The maximum-a-posteriori expectation maximization (MAPEM) reconstruction is given by [4]

$$\lambda_j^{\text{old}} = \frac{\lambda_j^{\text{old}}}{\sum_i c_{ij} - \frac{\partial M}{\partial \lambda_j}} \sum_i c_{ij} \frac{y_i}{\sum_k c_{ik} \lambda_k^{\text{old}}}, \quad (9)$$

where λ_j and λ_j^{old} are the new and current activity values in pixel j , y_i is the measured count in i , c_{ij} is proportional to the detection probability and the partial derivative of the prior M must be evaluated in the new (unknown) reconstruction λ . At the point of convergence, $\lambda_j = \lambda_j^{\text{old}}$ and we have:

$$\frac{\partial M}{\partial \lambda_j} = -\frac{\partial L}{\partial \lambda_j} = \sum_i c_{ij} \left(1 - \frac{y_i}{\sum_k c_{ik} \lambda_k} \right), \quad (10)$$

where L is the logarithm of the likelihood.

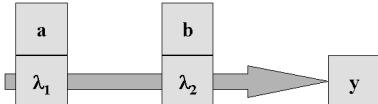


Figure 3: Simplified reconstruction problem.

Figure 3 represents a simplified reconstruction problem involving a single measurement y which depends on two pixel values λ_1 and λ_2 . Both pixels have a neighbor fixed at the values a and b respectively.

Assuming (scaled) detection probabilities of 1, the maximum likelihood (MLEM) algorithm reduces to

$$\lambda_j = \lambda_j^{\text{old}} \frac{y}{\lambda_1^{\text{old}} + \lambda_2^{\text{old}}}, \quad j = 1, 2 \quad (11)$$

implying that the likelihood is maximized by any pair of non-negative numbers with sum equal to y . Introducing the Gibbs prior M and applying equation (10) produces:

$$\frac{\partial M}{\partial \lambda_1} = \frac{\partial M}{\partial \lambda_2} = 1 - \frac{y}{\lambda_1 + \lambda_2}. \quad (12)$$

Since the gradient of the prior (6) is a function of the ratio of the pixel values, equation (12) is satisfied when

$$\frac{\lambda_1}{a} = \frac{\lambda_2}{b} \quad \text{or} \quad \frac{\lambda_1 - a}{\lambda_1 + a} = \frac{\lambda_2 - b}{\lambda_2 + b}. \quad (13)$$

Thus, with prior (5) the posterior has a solution with equal relative differences, as required.

C. Avoiding excessive smoothing over edges.

For a quadratic prior, the gradient increases with increasing (absolute) difference between neighboring pixels, which results in strong smoothing over edges. Some tolerance for steep edges is obtained by fixing the gradient to a constant for large differences. This approach is used in the Huber prior [1], which is quadratic for small differences and linear for large ones.

A similar effect for relative differences between neighboring pixels is obtained with the parameter γ in the new prior. The gradient of the prior (6) converges to a constant value when $\gamma|\lambda_j - \lambda_k|$ is larger than $(\lambda_j + \lambda_k)$. Consequently, the parameter γ controls which relative differences are considered as “large”: a relative difference of $2/\gamma$ is “intermediate”.

D. Implementation of the MAPEM algorithm

P.J. Green [4] has proposed to implement the MAPEM algorithm by evaluating the partial derivative $\partial M / \partial \lambda_j$ of equation (9) in the known current reconstruction λ^{old} instead of in the yet unknown reconstruction λ . In our experience, this approximation works well for low to moderate strengths β of the prior. However, with large β , the denominator in (9) may become very small or even negative, causing numerical problems. To avoid this, we use a different approach. The MLEM algorithm can be written as a preconditioned gradient ascent algorithm:

$$\lambda_j = \lambda_j^{\text{old}} + \frac{\lambda_j^{\text{old}}}{\sum_i c_{ij}} \left. \frac{\partial L}{\partial \lambda_j} \right|_{\lambda^{\text{old}}} \quad (14)$$

where L is the likelihood. The prior is included by adding its gradient to that of the likelihood, and adapting the preconditioner to control convergence when the prior dominates:

$$\lambda_j = \lambda_j^{\text{old}} + \left(\frac{\partial L}{\partial \lambda_j} + \frac{\partial M}{\partial \lambda_j} \right) / \left(\frac{\sum_i c_{ij}}{\lambda_j^{\text{old}}} - \frac{\partial^2 M}{\partial \lambda_j^2} \right), \quad (15)$$

where all partial derivatives are evaluated in the current reconstruction λ^{old} . This expression reduces to MLEM if the likelihood dominates, and to an approximation of Newton’s method if the prior dominates. Note that the second derivative of the prior is always negative for non-negative images (equation (7)), so the denominator is the sum of two positive terms if $\lambda^{\text{old}} \geq 0$.

In absence of the prior, (15) is guaranteed to be non-negative. Because the prior is designed to reduce relative differences, it is not expected to introduce negative values. However, the preconditioner of the prior does not guarantee non-negativity. To make sure, our program checks for negative values and replaces them with zero.

IV. EVALUATION

Three simulation experiments have been carried out, and the prior was also applied to a clinical PET whole body study.

A. Simulation 1

A simulation experiment was carried out to verify if the new prior preserves relative differences as intended. The simulated object consists of three disks with a small hot spot near the center (fig. 4). The ratio of the hot spot activity to the activity in the disk was equal to 3 for all disks. Thus, the hot spots have the same relative difference to the surrounding background, but a different absolute difference. When reconstructed with MAP, the recovery coefficient of the small hot spots will be less than unity, because the hot spot violates the assumption of uniformity.

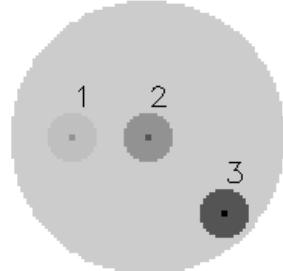


Figure 4: Software phantom. The regions are numbered from left to right. In the first experiment, the background in the enclosing circle was set to zero.

In the first simulation, the background activity was set to zero, and unattenuated noise-free projections were computed. MLEM and three MAP reconstructions were computed from the resulting sinograms; for MAP we applied the new prior, the Huber prior and the quadratic prior. For all priors, the

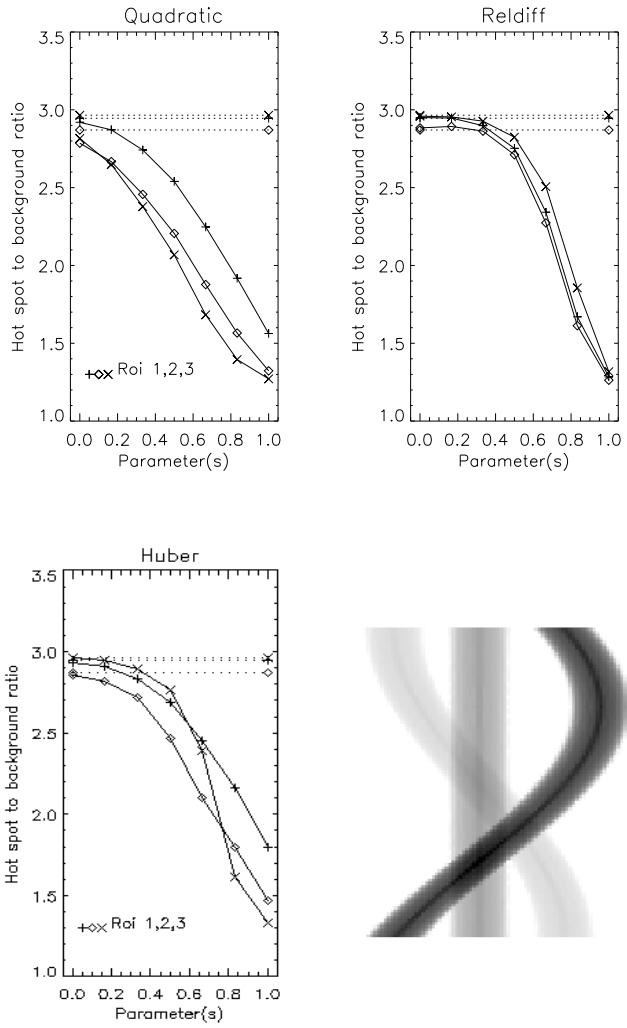


Figure 5: The object to background ratio of the 3 hot spots after MAP reconstruction, as a function of the prior weight β . Background activity and attenuation were zero. Top left: quadratic prior, top right: relative difference prior. Bottom left: Huber prior, bottom right: the simulated sinogram. The horizontal lines in the graphs are the values obtained with MLEM. In all cases, 180 iterations were applied.

parameters were varied. The results shown in figure 5 are for a single value of γ (new prior) and σ (Huber prior). However, similar results were obtained for different values of those parameters. The new prior always yielded similar recovery coefficients for the three hot spots. In contrast, the other priors produced different recovery coefficients. The quadratic prior smooths more aggressively in regions with higher activity. The Huber prior is in its linear mode for region 3, which explains its different behavior for that region. This result confirms that the new prior penalizes relative differences rather than absolute ones.

B. Simulation 2

In a second experiment, background activity was added and non-zero attenuation (constant attenuation along the projection line, as in PET) was incorporated. A uniform attenuating object was assumed, coinciding with the background activity in figure

4. Both effects make the strength of the likelihood position dependent (a measure of this strength can be obtained from the Fisher information matrix, see e.g. [5, 6]). The presence of background activity reduces the strength of the likelihood most near low count regions. Attenuation reduces the strength of the likelihood more for regions closer to the center of the enclosing attenuating object. Because the strength of the prior is uniform, the recovery coefficient should decrease with decreasing strength of the likelihood. Figure 6 shows the results. The

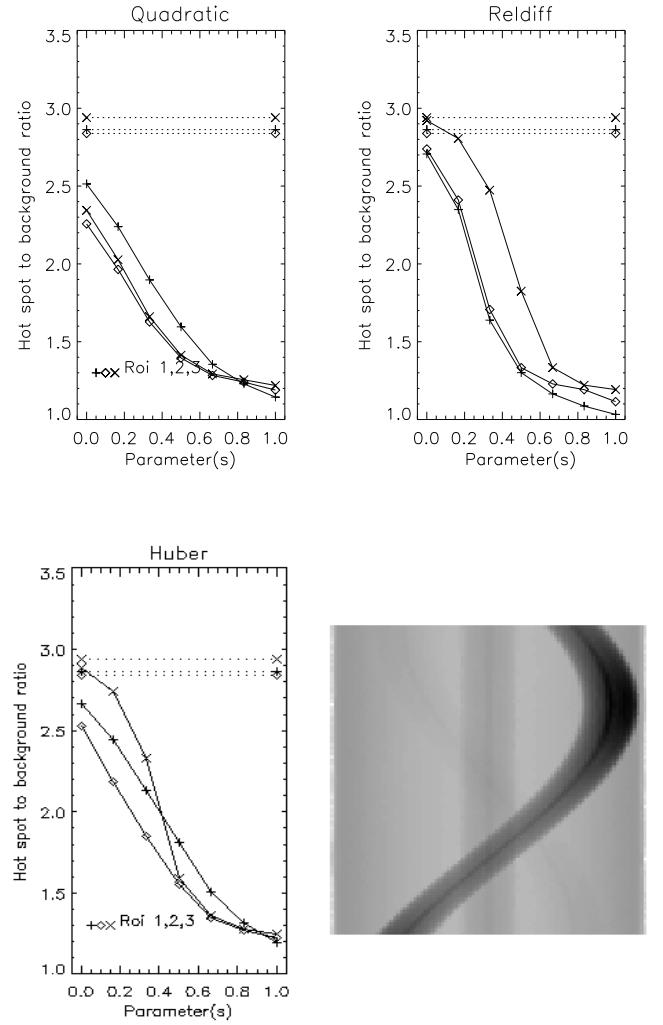


Figure 6: Same information as shown in figure 5, but now for the simulation with attenuation and background activity

relative difference prior produces the best recovery coefficient for region 3. This is as expected: because this region has the highest activity and is least attenuated during the measurement, the sinogram provides more information about this region than about the other two. Applying the quadratic prior causes more smoothing in region 3 than in region 1.

C. Simulation 3

A simple object was simulated to analyze the recovery coefficient as a function of the contrast for each of the priors. The object consists of a single disk with a hot spot in the center (figure 7). The ratio of the hot spot activity to the disk activity

was varied from 1.1 to 10. Unattenuated noise-free projections were calculated and a MAP reconstruction was computed with the three priors, fixing the parameters to an arbitrary value. Figure 8 shows the ratio in the reconstructed image as a

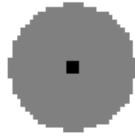


Figure 7: The simulated object of simulation 3. The contrast of the hot spot was varied.

function of the true ratio. The figure confirms that the relative differences prior has edge preserving characteristics somewhat similar to that of the Huber prior. The main difference is that the Huber prior switches from quadratic to linear mode at a fixed point, whereas the new prior gradually moves from one mode to the other.

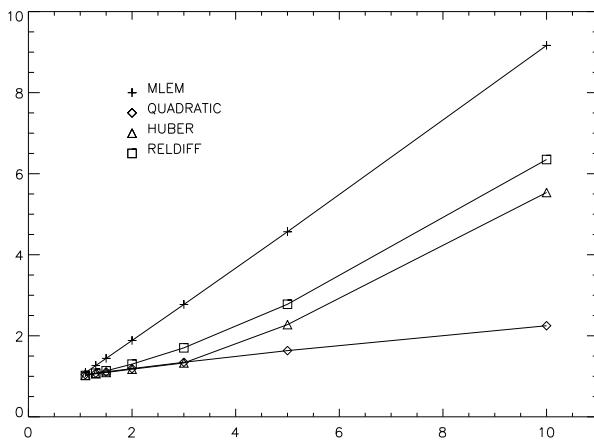


Figure 8: The hot spot to background ratio as a function of the true ratio. The symbols are as follows: MLEM: +, new prior: squares, Huber: triangles, quadratic: diamonds.

D. Clinical study

As an illustration, figure 9 compares the MLEM and MAP reconstructed image for a 2D PET fluorodeoxyglucose (FDG) whole body study with attenuation correction. The sixty three planes of one bed position were reconstructed simultaneously, the prior was applied using the six nearest neighbors in three dimensions using $\gamma = 10$. The study consists of 7 bed positions. This patient had a gastro-esophageal cancer, which is clearly visible in the ML and MAP images. The prior strongly suppresses the noise in the liver (the spots are believed to be noise, since there was no clinical evidence for increased FDG uptake in the liver), and preserves the paravertebral lesion in the neck. At this time, the parameters have not yet been optimized for clinical application, and the effect on the diagnostic process remains to be evaluated.

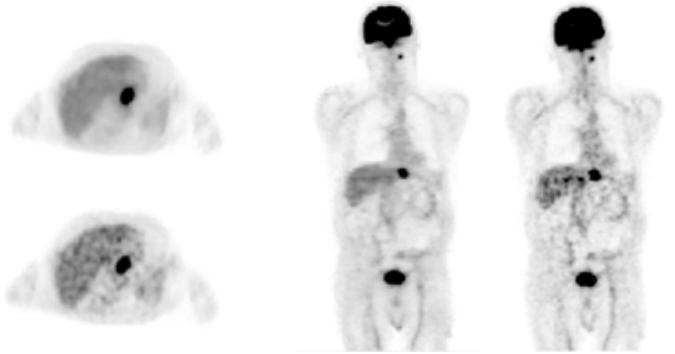


Figure 9: Reconstruction of a PET whole body study with MLEM and with MAP using the new prior. The left panel shows transaxial images obtained with MAP (top) and MLEM (bottom), the right panel compares coronal slices of MAP (left) and MLEM (right).

V. DISCUSSION

The requirements for the new prior are, at least to some extent, based on intuition rather than on strong scientific arguments. We believe that physicians focus on relative rather than on absolute differences when inspecting PET or SPECT images. As a result, it becomes very difficult to tune a prior that penalizes absolute differences, because the parameter σ specifies a threshold separating small from large absolute differences. Our new prior also needs tuning, but its parameter γ separates relative differences. The transition to the edge preserving mode occurs when $\gamma|\lambda_j - \lambda_k| \simeq (\lambda_j + \lambda_k)$, so the critical relative difference equals $2/\gamma$.

The Huber prior has a sharp transition between quadratic and linear mode. In contrast, our new prior has a gradual transition. It is not clear if this should be regarded as an advantage or a disadvantage. The value of this prior for clinical applications will have to be evaluated with observer studies.

Fessler and Rogers [5] have proposed a weighting strategy for quadratic priors that results in a uniform local impulse response. The weights are proportional to the diagonal elements of the Fisher information matrix. As argued by Qi et al [6], the variance of an MLEM pixel value is approximately proportional to the inverse of the corresponding Fisher information value. Wilson et al [7] have shown that in absence of attenuation, the variance of an MLEM pixel value is proportional to the reconstructed value. Consequently, one would expect that in absence of attenuation, multiplying the quadratic difference with the Fisher information value is approximately equivalent to dividing it by the reconstructed pixel value. Our new prior does exactly that. In absence of attenuation, it results in uniform recovery, suggesting a uniform local impulse response.

In presence of noise and background activity, the inverse of the reconstructed count is no longer a good estimate for the Fisher information. Consequently, the effect of our prior becomes non-uniform. One can argue that this is what should be expected from a prior: it smooths more in regions where less information can be obtained from the data.

VI. CONCLUSION

We have hypothesized that a concave Gibbs prior, which penalizes relative differences between neighboring pixels and which avoids excessive smoothing over large edges should be useful for MAP-reconstruction in emission tomography. A prior meeting these requirements has been derived. Our initial evaluation indicates that the prior behaves as expected.

VII. REFERENCES

- [1] EÜ Mumcuoglu, R Leahy, SR Cherry. "Bayesian reconstruction of PET images: methodology and performance analysis," *Phys Med Biol*, vol 41, pp. 1777-1807, 1996.
- [2] S. Geman, D.E. McClure. "Statistical methods for tomographic image reconstruction." *Bull. Int. Stat. Inst.* vol 52-4, pp. 5-21, 1987
- [3] LS Shepp, Y Vardi. "Maximum likelihood reconstruction for emission tomography," *IEEE Trans Med Imaging*, vol MI-1, pp. 113-122, 1982.
- [4] PJ Green. "Bayesian reconstruction from emission tomography data using a modified EM algorithm", *IEEE Trans Med Imaging* vol 9, pp. 84-93, 1990.
- [5] JA Fessler, WL Rogers. "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomography", *IEEE Trans Image Processing*, vol 5, pp. 1346-1358, 1996.
- [6] J Qi, RM Leahy. "A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data", *IEEE Trans Med Imaging* vol 18, pp. 293-305, 1999.
- [7] DW Willson, BMW Tsui, HH Barrett. "Noise properties of the EM algorithm: II. Monte Carlo simulations", *Phys Med Biol* vol 39, pp. 847-872, 1994.