A New Convex Edge-Preserving Median Prior with Applications to Tomography

Ing-Tsung Hsiao, *Member IEEE*, Anand Rangarajan, *Member IEEE*, and Gene Gindi*, *Member IEEE*

Abstract

In a Bayesian tomographic maximum *a posteriori* (MAP) reconstruction, an estimate of the object $\mathbf{f}$ is computed by iteratively minimizing an objective function that typically comprises the sum of a log-likelihood (data consistency) term and prior (or penalty) term. The prior can be used to stabilize the solution and to also impose spatial properties on the solution. One such property, preservation of edges and locally monotonic regions, is captured by the well-known median root prior (MRP) [1][2], an empirical method that has been applied to emission and transmission tomography. We propose an entirely new class of convex priors that depends on $\mathbf{f}$ and also on $\mathbf{m}$, an auxiliary field in register with $\mathbf{f}$. We specialize this class to our median prior (MP). The approximate action of the median prior is to draw, at each iteration, an object voxel towards its own local median. This action is similar to that of MRP and results in solutions that impose the same sorts of object properties as does MRP. Our-MAP method is not empirical, since the problem is stated completely as the minimization of a joint (on $\mathbf{f}$ and $\mathbf{m}$) objective. We propose an alternating algorithm to compute the joint MAP solution and apply this to emission tomography, showing that the reconstructions are qualitatively similar to those obtained using MRP.

Index Terms

Bayesian tomographic reconstruction, median, edge-preserving prior, iterative algorithm, emission tomography

This work is supported under grant R01-NS32879 from NIH-NINDS, and grant CMRP-1334 from Chang Gung Memorial Hospital Research Fund (Taiwan) *Asterisk indicates corresponding author.*

I.-T. Hsiao was with the Departments of Radiology and Electrical & Computer Engineering, SUNY Stony Brook, Stony Brook, NY 11784. He is now with the School of Medical Technology, Kwei-Shan, Tao-Yuan 333, Taiwan, email: ihsiao@mail.cgu.edu.tw

A. Rangarajan is with Department of Computer & Information Science and Engineering, University of Florida, Gainesville, FL 32611, email:anand@cise.ufl.edu.

G. Gindi* is with the Department of Radiology, SUNY Stony Brook, Stony Brook, NY 11784-8460, Tel: (631) 444-2539, Fax: (631) 444-6450, email: gindi@clio.rad.sunysb.edu.

I. Introduction

Maximum penalized likelihood, a statistical approach to inverse problems such as those encountered in emission tomographic reconstruction, may, in some cases lead to improved image quality relative to that obtained with alternative approaches. A penalized likelihood reconstruction can easily incorporate noise and imaging models into the reconstruction, and, by using a penalty, resolves problems of ill-conditioning that may arise. The penalty also allows incorporation of object properties into the solution. One such property of interest is spatial smoothness punctuated by significant edges. The notion of "spatial smoothness" has been used to imply nearly piecewise constant regions, piecewise ramplike regions, or piecewise monotonic regions. In this paper, we focus on a novel formulation of a new kind of penalty that incorporates a form of preservation of edges and locally monotonic regions in a manner reminiscent of the well-know median root prior (MRP) [1][2]. We show that our formulation is a true maximum penalized-likelihood method where, unlike MRP, the solution is obtainable by optimizing an objective function. In our ensuing discussion, we will consider a generic case of penalized likelihood reconstruction, then specialize it to emission tomography in illustrating our results.

Let $(f_j; j = 1 \dots N)$ denote the object quantity to be estimated at voxel j in the object, and $(\hat{f}_j; j = 1 \dots N)$ its estimate derived from a reconstruction procedure. The multidimensional object and object estimate are lexicographically ordered into N-element vectors $\mathbf{f}$ and $\hat{\mathbf{f}}$, respectively, with components f_j and $\hat{f}_j$, respectively. The detected data has elements $(g_i; i = 1 \dots M)$, again lexicographically ordered into an M-element vector $\mathbf{g}$.

A maximum penalized likelihood estimate $\hat{\mathbf{f}}$ of the object $\mathbf{f}$ can be computed by optimizing an objective function $\Phi(\mathbf{f})$

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \Phi(\mathbf{f}) = \arg\min_{\mathbf{f}} \{\Phi_L(\mathbf{g}; \mathbf{f}) + \lambda \Phi_P(\mathbf{f})\}. \tag{1}$$

where $\Phi_L(\mathbf{g}; \mathbf{f})$, the $-\log$ of the likelihood, captures system and noise models, and the object properties are captured in the penalty given by $\Phi_P(\mathbf{f})$. The scalar $\lambda > 0$ is a global weight controlling the relative influence of the penalty. This same approach is often described loosely using a Bayesian terminology wherein (1) is a maximum *a posteriori*

(MAP) estimate with $\Phi(\mathbf{f})$ considered as the $-\log$ of a posterior probability on $\mathbf{f}$ and $\lambda\Phi_P(\mathbf{f})$ considered to be proportional to the $-\log$ of a prior probability on the object. In the MAP context, $\mathbf{f}$ and $\hat{\mathbf{f}}$ are random vectors. For the ensuing discussion, we shall retain this Bayesian terminology, using terms "prior", "posterior" and "MAP", since this slight abuse is by now well entrenched in the medical imaging literature.

In this paper, we formulate a new "median prior" (MP), that, in a MAP context, attempts to emulate the performance obtainable with the MRP. The median root prior (MRP) [1][2], attempts to impose the behavior that each voxel is attracted to its local median while, through the action of the likelihood term, satisfying data consistency. This action of the MRP captures significant edges while encouraging preservation of locally monotonic regions. Like MRP, our MP preserves edges, and also preserves monotonic regions. However, as we will discuss, the MRP algorithm is a heuristic empirical method; it is difficult to analyze since it lacks an associated objective function. It is impossible to formulate an MRP reconstruction as the minimization of an objective as in (1). By contrast, the MP has an associated objective and can be used within a MAP context. Furthermore, the objective associated with the median prior will be shown to be convex; thus if the likelihood objective Φ_L is also convex (as is the case in many applications including emission tomography), then the posterior objective $\Phi(\mathbf{f})$ is also convex and the MAP solution is a global minimum of $\Phi(\mathbf{f})$.

In Section II, we formulate an entire new class of convex priors and specialize this to our MP. We show how the associated MAP estimate can be obtained using an alternating algorithm. We compare/contrast the MP with MRP. In Section III we apply the MP to emission tomography and present anecdotal results showing qualitatively similar reconstructions for MP and MRP. In Section IV we conclude with a discussion.

II. Theory

First, we define a new class of priors $\Phi_P(\mathbf{f}, \mathbf{m})$ that will depend on *two* vector variables $\mathbf{f}$ and $\mathbf{m}$. The "auxiliary" vector $\mathbf{m}$ has components $(m_{j'}; j' = 1 \ldots N)$ in register with the components of $\mathbf{f}$. The new prior penalizes differences between the voxels f_j and

components $m_{j'}$ of $\mathbf{m}$

$$\Phi_P(\mathbf{f},\mathbf{m}) = \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'} \phi(f_j - m_{j'}) \tag{2}$$

where $\mathcal{N}(j)$ is the local neighborhood system of voxel j including j itself. The auxiliary vector $\mathbf{m}$ (see Fig. 1) is in register with $\mathbf{f}$, and the neighborhood weights $w_{jj'}$ link f_j *not* with its $f_{j'}$-neighbors, but with $\{m_{j'}; j' \in \mathcal{N}(j)\}$. The term $\phi(\cdot)$ is a potential function whose form will determine the action of the prior. Note that an argument z to the potential function $\phi(z)$ is usually associated with penalizing the neighboring voxels' difference, $z = f_j - f_{j'}$, but here we use $z = f_j - m_{j'}$. Note also that in (2), the vector $\mathbf{m}$ is a variable to be estimated, and should not be confused with a "line process" [3] used in some edge-preserving priors. The weights $w_{jj'} \geq 0$. Later, we will introduce the specific form of ϕ to specialize (2) to our new median prior.

Note that the prior (2) depends on *two* vector variables $\mathbf{f}$ and $\mathbf{m}$, so that the posterior in (1) must be generalized to the form of a *joint* posterior:

$$\Phi(\mathbf{f},\mathbf{m}) = \Phi_L(\mathbf{g};\mathbf{f}) + \lambda \Phi_P(\mathbf{f},\mathbf{m}) \tag{3}$$

For the case where $\Phi_L(\mathbf{g};\mathbf{f})$ is convex, then we must prove the convexity of $\Phi_P(\mathbf{f},\mathbf{m})$ to ensure that the joint posterior $\Phi(\mathbf{f},\mathbf{m})$ is globally convex.

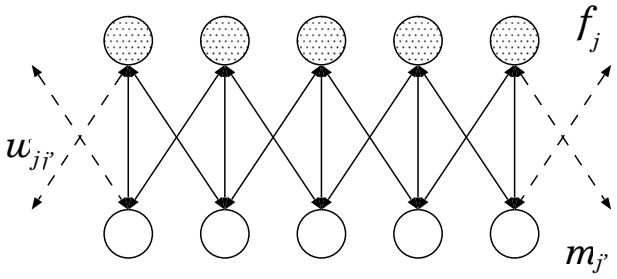


Fig. 1. One-dimensional illustration of $\mathbf{f}$ and auxiliary field $\mathbf{m}$ with 3 nearest-neighbor links for $w_{jj'}$.

A. Convexity

We will show that *any* prior of the form in (2) is convex subject only to the condition that ϕ is convex. Furthermore, the proof of convexity for (2) does not depend on the differentiability of ϕ.

In (2), weights w comprise a graph with adjacency matrix entries $\{w_{jj'}\}$ encapsulating the neighborhood topology. Our task ahead is to show that this prior is convex w.r.t. both $\mathbf{f}$ and $\mathbf{m}$ given that ϕ is convex. If ϕ is convex, we have

$$\phi(\alpha x + (1-\alpha)y) \leq \alpha\phi(x) + (1-\alpha)\phi(y) \tag{4}$$

for $\alpha \in [0,1]$. Given two points $(\mathbf{f}^{(1)}, \mathbf{m}^{(1)})$ and $(\mathbf{f}^{(2)}, \mathbf{m}^{(2)})$ in the domain of $\Phi_P(\mathbf{f}, \mathbf{m})$, we may write a convexity condition for $\Phi_P(\mathbf{f}, \mathbf{m})$ in (2) as follows:

$$\sum_{jj'} w_{jj'}\phi[\alpha f_j^{(1)} + (1-\alpha)f_j^{(2)} - \alpha m_{j'}^{(1)} - (1-\alpha)m_{j'}^{(2)}] \leq \alpha \sum_{jj'} w_{jj'}\phi(f_j^{(1)} - m_{j'}^{(1)}) + (1-\alpha)\sum_{jj'} w_{jj'}\phi(f_j^{(2)} - m_{j'}^{(2)}) \tag{5}$$

for $\alpha \in [0,1]$. Since

$$\sum_{jj'} w_{jj'}\phi[\alpha f_j^{(1)} + (1-\alpha)f_j^{(2)} - \alpha m_{j'}^{(1)} - (1-\alpha)m_{j'}^{(2)}] = \sum_{jj'} w_{jj'}\phi[\alpha(f_j^{(1)} - m_{j'}^{(1)}) + (1-\alpha)(f_j^{(2)} - m_{j'}^{(2)})]$$

and since $w_{jj'} \geq 0$, we have from the convexity of $\phi(\cdot)$ in (4):

$$\sum_{jj'} w_{jj'}\phi[\alpha(f_j^{(1)} - m_{j'}^{(1)}) + (1-\alpha)(f_j^{(2)} - m_{j'}^{(2)})] \leq \alpha \sum_{jj'} w_{jj'}\phi(f_j^{(1)} - m_{j'}^{(1)}) + (1-\alpha)\sum_{jj'} w_{jj'}\phi(f_j^{(2)} - m_{j'}^{(2)})$$

which is the right side of (5). We have shown that $\Phi_P(\mathbf{f}, \mathbf{m})$ is convex if $\phi(\cdot)$ is convex. No assumptions have been made regarding the differentiability of $\phi(\cdot)$.

B. Median Prior

By choosing a suitable convex form for $\phi(\cdot)$, we can specialize (2) to one version of the MP, denoted by $\Phi_P^{abs}(\mathbf{f}, \mathbf{m})$. This is a provisional version useful for comparison to MRP, but we will further modify this version, for practical computational reasons, to our final version $\Phi_P^{med}(\mathbf{f}, \mathbf{m})$.

There is a connection between an absolute value potential $\phi(z) = |z|$ and the median. Note that by optimizing the absolute function within a local neighborhood, the solution turns out to be the local median. For example, $median\ \{2,2,5\} = 2 = \arg\min_\zeta\{|\zeta - 2| + |\zeta - 2| + |\zeta - 5|\}$. The median is of interest because the median of a dataset preserves edges and locally monotonic regions, a property advocated in [1][2]. Now we define $\Phi_P^{abs}(\mathbf{f}, \mathbf{m})$

using $\phi(z) = |z|$, the absolute value potential function, so that

$$\begin{aligned} \Phi_P^{abs}(\mathbf{f}, \mathbf{m}) &= \sum_j \sum_{j' \in \mathcal{N}(j)} w_{jj'} \phi(f_j - m_{j'}) \\ &= \sum_j \sum_{j' \in \mathcal{N}(j)} |f_j - m_{j'}| \end{aligned} \tag{6}$$

where $w_{jj'} = 1$ for $j' \in \mathcal{N}(j)$ and is zero otherwise.

While we have shown that the general form (2) is convex, we still need to show that $\phi(z) = |z|$ is convex. We require

$$|\alpha x + (1-\alpha)y| \leq \alpha|x| + (1-\alpha)|y|.$$

Since $|a+b| \leq |a| + |b|$, we have

$$|\alpha x + (1-\alpha)y| \leq |\alpha x| + |(1-\alpha)y| = \alpha|x| + (1-\alpha)|y|$$

for $\alpha \in [0,1]$. This proves the result. Note that no differentiability conditions have been invoked.

Since the prior Φ_P^{abs} is a function of both $\mathbf{f}$ and $\mathbf{m}$, the MAP optimization in (1) now becomes a *joint* estimation, and should be rewritten as:

$$\hat{\mathbf{f}}, \hat{\mathbf{m}} = \arg\min_{\mathbf{f},\mathbf{m}} \{\Phi_L(\mathbf{g}; \mathbf{f}) + \lambda \Phi_P^{abs}(\mathbf{f}, \mathbf{m})\}. \tag{7}$$

where $\hat{\mathbf{m}}$ is the MAP estimate of $\mathbf{m}$. We implement this joint estimation by a form of coordinate (on $\mathbf{f}$ and $\mathbf{m}$) descent, which for iteration k takes the general form:

$$\hat{\mathbf{f}}^{k+1} = \arg\min_{\mathbf{f}} \{\Phi_L(\mathbf{g}; \mathbf{f}) + \lambda \Phi_P^{abs}(\mathbf{f}, \hat{\mathbf{m}}^k)\} \tag{8}$$

$$\hat{\mathbf{m}}^{k+1} = \arg\min_{\mathbf{m}} \{\Phi_P^{abs}(\hat{\mathbf{f}}^{k+1}, \mathbf{m})\} \tag{9}$$

(After each pass through the alternation, $k \to k+1$.)

We embark on an interpretation of (8) and (9). We use the notation $j \in \mathcal{N}(j')$ to designate the f_j neighbors of $m_{j'}$. By our discussion relating the median to the absolute value objective, the solution $\hat{\mathbf{m}}^{k+1}$ of (9) is a set of *local medians* of $\hat{\mathbf{f}}^{k+1}$. That is, $\hat{\mathbf{m}}^{k+1}$ is obtained as the result of a median window filtering operation on $\hat{\mathbf{f}}^{k+1}$, where the window size is determined by the neighborhood structure $\mathcal{N}(j')$. Given this interpretation of (8)

and (9), we rewrite the alternation as:

$$\hat{\mathbf{f}}^{k+1} = \arg\min_{\mathbf{f}} \{\Phi_L(\mathbf{g};\mathbf{f}) + \lambda \Phi_P^{abs}(\mathbf{f}, \hat{\mathbf{m}}^k)\} \quad (10)$$

$$\hat{m}_{j'}^{k+1} = \text{median}(\hat{f}_j^{k+1}; j \in \mathcal{N}(j')) \quad (11)$$

To continue our interpretation, consider a specific 1-D example where the neighborhood structure is as in Fig. 1, $k = 2$ and $j' = 6$. For this case, from (11), $\hat{m}_6^3 = \text{median}\{\hat{f}_5^3, \hat{f}_6^3, \hat{f}_7^3\}$. Now examine (10). It is itself a MAP estimate of $\mathbf{f}$ given the current set of medians $\hat{\mathbf{m}}^k$. The prior in (10) takes the form $\Phi_P^{med}(\mathbf{f}, \hat{\mathbf{m}}^k) = \sum_j \sum_{j' \in \mathcal{N}(j)} |f_j - \hat{m}_{j'}^k|$. The action of the prior in this expression is somewhat subtle: By the relationship between the median and the absolute value objective, each $\hat{f}_j^{k+1}$ is attracted to the local median of the $\hat{m}_{j'}^k$'s. But the $\hat{m}_{j'}^k$'s are *themselves* local medians of the $\hat{\mathbf{f}}^k$'s. Continuing with our example, consider $j = 6$. Then $\hat{f}_6^4$ is drawn by the prior towards:
$\text{median}\Big[\hat{m}_5^3, \hat{m}_6^3, \hat{m}_7^3\Big] = \text{median}\Big[\text{median}\{\hat{f}_4^3, \hat{f}_5^3, \hat{f}_6^3\}, \text{median}\{\hat{f}_5^3, \hat{f}_6^3, \hat{f}_7^3\}, \text{median}\{\hat{f}_6^3, \hat{f}_7^3, \hat{f}_8^3\}\Big]$.
Thus one can say that in the above example, each $\hat{f}_j^4$ is attracted to the median of the set of local medians of $\hat{f}_j^3$. Or, more generally, the action of the MP prior is to attract, at each iteration, the object estimate $\hat{\mathbf{f}}$ to its own median of local medians. This "median of medians" interpretation, though slippery, will allow us to relate our MP to MRP.

C. Comparison of MP to MRP

Two versions of the algorithm that implements MRP are presented in [1] and [2]. For emission tomography, the MRP algorithm takes its most familiar form as an expectation-maximization MAP one-step-late update, which appears as (21) in Section III-A where we discuss our application to emission tomography. For a clear comparison of MRP to MP, however, we now reexpress the MRP algorithm in a more general form.

The MRP algorithm can be expressed in a form that resembles our MAP optimization (10) and (11):

$$\hat{\mathbf{f}}^{k+1} = \arg\min_{\mathbf{f} \geq 0} \{\Phi_L(\mathbf{g};\mathbf{f}) + \lambda \Phi_P^{mrp}(\mathbf{f}; \hat{\mathbf{m}}^k)\} \quad (12)$$

$$\hat{m}_{j'}^{k+1} = \text{median}(\hat{f}_j^{k+1}; j \in \mathcal{N}(j')) \quad (13)$$

where (12) is in itself a MAP optimization (on $\mathbf{f}$ alone) using the MRP data-dependent

prior $\Phi_P^{mrp}(\mathbf{f};\mathbf{m})$ given by:

$$\Phi_P^{mrp}(\mathbf{f};\hat{\mathbf{m}}^k) = \sum_j \frac{(f_j - \hat{m}_j^k)^2}{2\hat{m}_j^k} \tag{14}$$

Also note that (13) is identical to (11).

The action of the MRP algorithm can be understood by examining (12), (13) and (14). The term $\hat{m}_{j'}^{k+1}$ is a local median as in (11). From the simple quadratic form in (14), the $\hat{f}_j^{k+1}$ does not interact with any of its $\hat{f}$ neighbors, but is instead drawn towards its own local median, $\hat{m}_j^{k+1}$, i.e. it is drawn towards $m_{j'}^{k+1}$ with $j' = j$. The $\hat{f}_j^{k+1}$ simultaneously tries to satisfy data consistency through the action of the likelihood term. This differs from (10), where each $\hat{f}_j^{k+1}$ is attracted *not* to its local median, but to its median of local medians. Because in MRP each object voxel is attracted to its own local median, locally monotonic patches of the object are left unaffected by the action of the term Φ_P^{mrp}. That is, a locally monotonic patch is a "root" of the MRP.

The MP and MRP algorithms, despite resemblances, differ crucially. Unlike (10) and (11), the alternation (12) and (13) is *not* derivable as the minimization of an overall joint objective. Note that while (12) is technically a MAP optimization on $\mathbf{f}$ given $\mathbf{m}$, the pair (12) (13) has no associated joint objective. Indeed, in (12) and (13) , the quantity $\mathbf{m}$ is not a variable, but is instead an empirical data-dependent quantity. Thus our MP algorithm is a joint MAP optimization, and the MRP algorithm is empirical in that the "prior" is updated empirically at each iteration.

Note that the MP and MRP alternations differ only in the definition of the priors. Whereas from (12) above, we see that f_j is drawn to its local median, we had noted, by contrast, that from (10) and (11), that f_j is attracted to the the median of a set of local medians. Surprisingly, the MP prior Φ_P^{abs} acts to preserve locally monotonic patches just as for MRP. One way to understand this is to note that in a locally monotonic region of $\mathbf{f}$, the operation of taking a median of a local neighborhood of medians leaves $\mathbf{f}$ unchanged. A second way to see this is as follows: Observe that the gradient of the absolute value is ± 1, so that f_j is attracted with the same "force" $(= +\lambda)$ towards any neighbor median of greater value $(m_{j'} > f_j)$, attracted with an equal and opposite force $(= -\lambda)$ to any neighbor median with a lesser value $(m_{j'} < f_j)$, and attracted to its own median $(m_j = f_j)$ with zero force. (This zero attraction to itself would be hard to observe numerically due

to roundoff errors.) If f_j sits in a locally monotonic patch comprising $2N+1$ neighbors, then it will be attracted with a force proportional to $+N\lambda$ to the N median neighbors of greater values, a force proportional to $-N\lambda$ to the N median neighbors with lesser values, and zero force to its local median $m_j = f_j$. The forces thus cancel and the value of f_j is undisturbed. Thus a locally monotonic patch is left unaltered by the prior, and MP behaves just like MRP !

D. A More Practical Median Prior

Though useful for comparison to MRP, Φ_P^{abs} is impractical for computational purposes in its present form. While (11) is easily accomplished by a simple median-window filtering operation, the optimization in (10) runs into problems if gradient methods are used. Because of the absolute value form of $\Phi_P^{abs}(\mathbf{f}, \mathbf{m})$, the prior is convex but is not differentiable at points where the argument to an absolute value term is zero (i.e. where $f_j = m_{j'}$ in (6)). As such, it is difficult to incorporate into the most common forms of optimization schemes that rely on the differentiability of the objective. To circumvent this problem, we use the log cosh function to approximate $|z|$ by

$$\frac{1}{\eta}\log\cosh(\eta z) \approx |z| \tag{15}$$

where η is a parameter that adjusts the accuracy of the approximation. A plot of $\frac{1}{\eta}\log\cosh(\eta z)$ vs. $|z|$, illustrated in Fig. 2 for two different values of η, shows the effect of η on the approximation. As η increases, the absolute value function is better approximated. Note that (15) is but one choice for a differentiable approximation to the absolute value. [1]

With this approximation, we define our final form of the median prior, $\Phi_P^{med}(\mathbf{f}, \mathbf{m})$ as

$$\Phi_P^{med}(\mathbf{f}, \mathbf{m}) \equiv \frac{1}{\eta}\sum_j \sum_{j' \in \mathcal{N}(j)} \log\cosh(\eta(f_j - m_{j'})) \tag{16}$$

To ensure that Φ_P^{med} remains convex, we need only note that if $\phi(z) = \log\cosh(z)$ is convex, then by our proof in Section II-A, Φ_P^{med} is also convex. It is easy to demonstrate the convexity of $\phi(z) = \log\cosh(z)$ by noting that its second derivative is given by $\mathrm{sech}^2(z) \geq 0$.

[1] A more convenient approximation to $|x|$, suggested by an anonymous referee, is $\sqrt{x^2+\epsilon}$

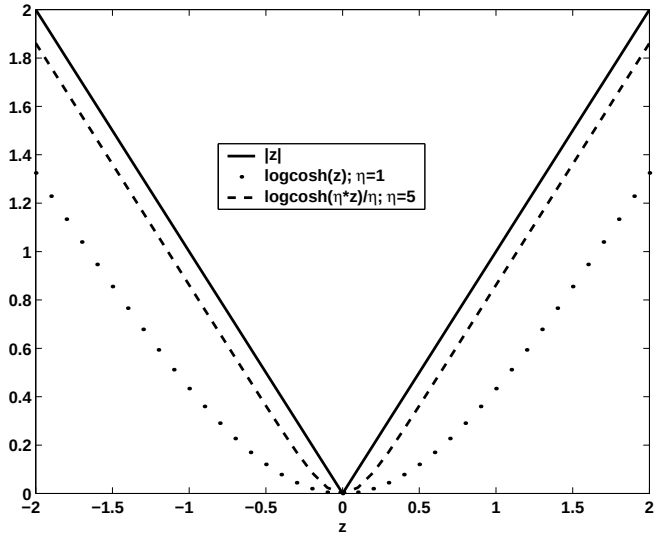


Fig. 2. This illustrates the effect of the η parameter on the approximation of the absolute value function by log cosh. The solid line is the absolute value function, while the dotted (..) line and the dashed (–) line indicate the log cosh function with $\eta = 1$ and $\eta = 5$, respectively.

With this modified form of the MP prior, the MAP problem becomes

$$\hat{\mathbf{f}}, \hat{\mathbf{m}} = \arg\min_{\mathbf{f},\mathbf{m}} \{\Phi_L(\mathbf{g}; \mathbf{f}) + \lambda \Phi_P^{med}(\mathbf{f}, \mathbf{m})\}. \tag{17}$$

and the corresponding alternation becomes

$$\hat{\mathbf{f}}^{k+1} = \arg\min_{\mathbf{f}} \{\Phi_L(\mathbf{g}; \mathbf{f}) + \lambda \Phi_P^{med}(\mathbf{f}, \hat{\mathbf{m}}^k)\} \tag{18}$$

$$\hat{\mathbf{m}}^{k+1} = \arg\min_{\mathbf{m}} \{\Phi_P^{med}(\hat{\mathbf{f}}^{k+1}, \mathbf{m})\} \tag{19}$$

Note that the minimization in (19) is now only an approximation to a median-window filtering operation, and in an implementation, the actual minimization must itself be computed by some iterative procedure. Again, Φ_P^{med} approximates Φ_P^{abs} in that MRP-like behavior is better approximated as $\eta \to \infty$. The minimization in (18) is itself an iterative MAP estimation (on $\mathbf{f}$). Hence for each outer-loop iteration k, inner loop iterations for (18) and for (19) must be performed, and the overall joint MAP estimation would appear to be computationally intense. However, in Section III-A, we show how to implement the joint estimation (18) and (19) so that the net computational complexity is approximately that of more conventional MAP reconstructions in emission tomography.

III. Application to Tomography

In this section, we apply MAP reconstruction with our MP prior (16) to the problem of emission tomography (ET) . We compare ET reconstructions with those obtained using an MRP method, and observe that the reconstructions are qualitatively similar. Our own interest lies in ET, and we note that MRP was first proposed for application to emission and transmission tomography.

To specialize our general formulations to ET, we need to define $\mathbf{f}$ and $\mathbf{g}$ and then state our imaging and noise models through specification of the log likelihood $\Phi_L(\mathbf{g};\mathbf{f})$. For ET, f_j is the mean emission rate (counts/sec/vol) from voxel j, and g_i is the number of counts received at detector bin i, i.e. the sinogram. Let the $M \times N$ system matrix $\mathcal{H}$ have elements $\mathcal{H}_{ij}$ proportional to the probability of receiving a count in detector element i from voxel j. In our simplified model for ET, we will use $\mathcal{H}$ to model a discrete Radon transform and ignore the myriad physical effects often modelled in $\mathcal{H}$ for both SPECT and PET. In this simplified model, the sinogram data g_i are independently Poisson distributed so that the likelihood is $P(\mathbf{g}|\mathbf{f}) = \text{Poisson}(\bar{\mathbf{g}})$ with mean $\bar{\mathbf{g}} = \mathcal{H}\mathbf{f}$. Given this likelihood, we may now write $\Phi_L(\mathbf{g};\mathbf{f})$ as the familiar expression

$$\Phi_L(\mathbf{g};\mathbf{f}) = -\sum_i \{g_i \log[\sum_j \mathcal{H}_{ij} f_j] - \sum_j \mathcal{H}_{ij} f_j\}. \tag{20}$$

A. Implementational Details for Joint MAP MP Reconstruction (18) and (19)

The MAP-MP reconstruction is given by the alternation (18) and (19). Note that, as written, each of these equations is itself a minimization that must be solved by an iterative method, and so each involves a sub-iteration of the main k-iteration. To solve (19), we note that the problem breaks up into a series of separate 1-D minimizations for each $m_{j'}^{k+1}$, and we use a Newton method for each 1-D minimization. For (18), we needn't perform the complete minimization (inner loop) of $\Phi(\mathbf{f}, \mathbf{m}^k)$ for each outer loop k. Instead, we simply do a single iteration of the inner loop, in this case using one step of an unconstrained Polak-Ribiere form of a preconditioned conjugate-gradient (PCG) method [4] with a simple diagonal preconditioner. The overall convexity of the joint objective guarantees that this strategy works, and in practice, this strategy leads to a reasonably fast reconstruction. The computational complexity of a single PCG step is dominated by the operations of one

projection and one backprojection. Since the computing time in (19) is negligible with respect to a single PCG step used for (18), the overall computing time is determined by K, the number of k-iterations until convergence, and is approximately equal to that of $2K$ projection operations. We have observed in our simulations that K turns out to be roughly equal to the number of iterations used for a conventional MAP reconstruction, using for instance, a conventional quadratic smoothing prior. Thus the joint MAP reconstruction using our median prior takes about as long as a conventional MAP reconstruction.

We used an implementation of the median root prior as formulated in [1], in which, for emission tomography, the MRP is incorporated into an iterative EM-OSL (expectation-maximization one-step-late) algorithm [5]. The resulting MRP update equation becomes

$$\begin{aligned} \hat{f}_j^{k+1} &= \hat{f}_j^k \frac{\sum_i \mathcal{H}_{ij} \frac{g_i}{\sum_l \mathcal{H}_{il} \hat{f}_l^k}}{\sum_i \mathcal{H}_{ij} + \lambda \frac{\partial}{\partial f_j} \Phi^{mrp}(\mathbf{f} = \hat{\mathbf{f}}^k)} \\ &= \hat{f}_j^k \frac{\sum_i \mathcal{H}_{ij} \frac{g_i}{\sum_l \mathcal{H}_{il} \hat{f}_l^k}}{\sum_i \mathcal{H}_{ij} + \lambda \frac{\hat{f}_j^k - m_j^k}{m_j^k}} \end{aligned} \tag{21}$$

where $\Phi^{mrp}(\mathbf{f}) \equiv \sum_j \frac{(f_j - m_j)^2}{2m_j}$ and m_j is *heuristically* chosen to be a local median window on the neighbors of voxel j as in (13). In fact, with this definition of $\mathbf{m}$, (21) is a single-step realization of the general MRP alternation (12) and (13). This OSL method is not guaranteed to converge, but in all of the trials described below, it did converge.

B. Results

In this section, we present 2D anecdotal ET reconstructions using MP and MRP to illustrate qualitative comparisons. We generated two noisy sinograms (of 500k and 100k counts) using the 2D 64×64 phantom in Fig. 3. The phantom consists of a disk background, two hot lesions and two cold lesions with contrast ratio of 1:4:8 (cold:background:hot). The two sinograms are then MP-reconstructed using (18) and (19). We used a 5-neighborhood system for $\mathcal{N}(j)$, setting $w_{jj'} = 1$ for the center pixel and its 4 nearest neighbors. The results are shown in Fig. 4 for (a) noisy (500K) reconstruction with $\lambda = 0.5$ and (b) noisy (100K) reconstruction with $\lambda = 0.8$. The value of η is fixed at 20 for both reconstructions.

For comparison, we also reconstructed the same two sinograms using the median root

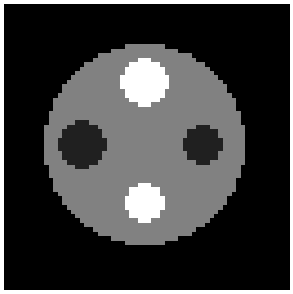


Fig. 3. The 64x64 digital phantom used in the simulations.

prior and the EM-OSL algorithm (21). A 9-neighborhood system was used for the median computation in (21). The anecdotal reconstructions of the median root prior are shown in Fig. 4 for (c) high-count (500K) reconstruction with $\lambda = 30$, and (d) low-count (100K) reconstruction with $\lambda = 80$.

We intentionally used different sized neighborhoods, 9 for MRP and 5 for MP. As may be seen in the MP alternation, each m interacts with 5 f's through (19), and then each f interacts with 5 m's via (18). Thus even though $w_{jj'}$ in (16) encodes only 5 neighbors, the *effective* neighborhood for the MP is about twice that size. So, to get qualitatively similar effects in the MRP and MP reconstructions, we intentionally mismatched the neighborhood sizes. In all cases, the value of λ was chosen empirically to achieve visually appealing and similar (MP vs MRP) reconstructions. Our aim was to show that our median prior and MRP have qualitatively similar behavior. We draw no conclusions on MP vs MRP performance comparisons.

As shown in Fig. 4(a)(b), the reconstructions using the median prior have qualitatively similar properties as those for the median root prior of Fig. 4(c)(d). However, the reconstructions seem smoother in Fig. 4(a)(b) than those in Fig. 4(c)(d), in particular, for the low-count data.

To see the effect of the value η on the results of the median prior, we also reconstructed the low-count data using (18) and (19) with two different η's as shown in Fig. 5(a) for $\eta = 6$ and (b) for $\eta = 30$. The edge-preserving effect is stronger in Fig. 5(b) than in (a), as expected.

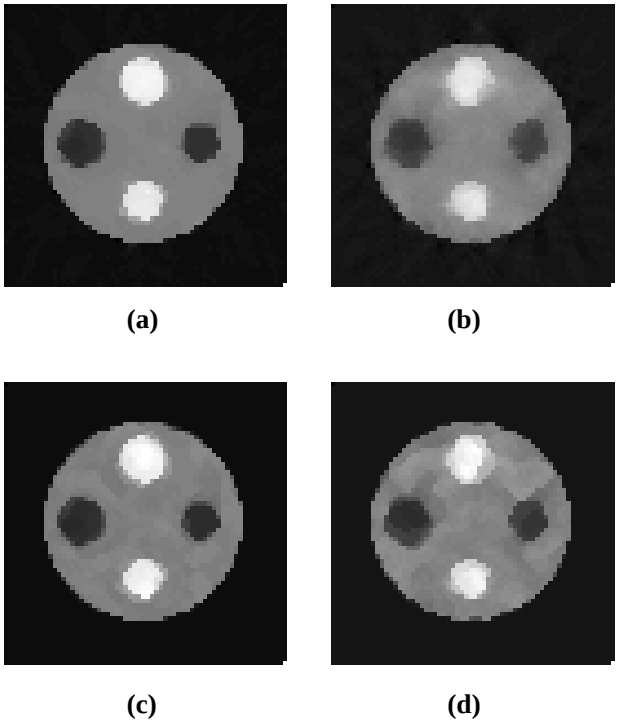
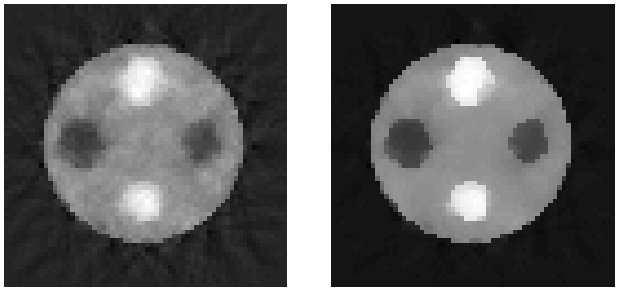


Fig. 4. The anecdotal reconstructions using the median prior for (a) high-count (500k) and (b) low-count (100k) data, and the results using the median root prior for (c) high-count and (d) low-count data. The results of the median prior show the edge-preserving properties similar to those of the median root prior. All 4 figures share a common gray scale.

IV. Discussion

We presented in (2) a general new class of prior that makes use of an auxiliary field $\mathbf{m}$, and, furthermore, showed that it is convex if the potential $\phi(\cdot)$ is convex. For $\phi(z) = |z|$, we used the relationship between the absolute value and the median to specialize (2) to a form of median prior Φ_P^{abs} that preserves locally monotonic regions and edges in a manner similar to that of MRP. We furthermore approximated Φ_P^{abs} by a modified version Φ_P^{med} whose differentiability ensures that it can be used in a practical optimization algorithm, and whose behavior approximates Φ_P^{abs} more closely as a parameter $\eta \to \infty$. Unlike MRP algorithms, our MP prior can be placed in a MAP context.

The use of the log cosh function for the potential $\phi(\cdot)$ recalls earlier work in [5] in which this potential was used to penalize the difference between neighboring voxels, $\phi(f_j - f_{j'})$, rather than the difference between the voxel and a auxiliary field, $\phi(f_j - m_{j'})$, as in our



(a) **(b)**

Fig. 5. The anecdotal reconstructions using the median prior with (a) $\eta = 6$, and (b) $\eta = 30$ for the low-count data (100k). Both figures share a common gray scale.

median prior. There is no median computation in [5]. (Other related work appeared in [6].) It would be interesting to compare performance aspects of our median prior with these seemingly incommensurate approaches, but that is beyond the scope of the present work.

We have argued that if both $\Phi_L(\mathbf{g};\mathbf{f})$ and $\Phi_P(\mathbf{f},\mathbf{m})$ are convex, then the solution $(\hat{\mathbf{f}},\hat{\mathbf{m}})$ is a global minimum of the associated posterior objective (3). Yet this global minimum solution may not be unique if both $\Phi_L(\mathbf{g};\mathbf{f})$ and $\Phi_P(\mathbf{f},\mathbf{m})$ are not *strictly* convex. If the solution is not unique, then it may depend on the initial condition $\hat{\mathbf{f}}^0$. For a previously studied [7] case of MAP reconstruction using the ET likelihood in (20) and a simple quadratic smoothing prior on $\mathbf{f}$ alone, the authors observed that while the likelihood and prior were each not strictly convex, the MAP solution was indeed unique. (It is easy to see that a simple quadratic smoothing prior is not strictly convex since any $\mathbf{f}$ that is constant everywhere yields zero for the value of the prior.) They argued that the null spaces associated with the likelihood and prior do not overlap for any reasonable tomographic system, so that the MAP solution must be unique. In our present case, Φ_L is not strictly convex, and it is easy to see that Φ_P^{med} or Φ_P^{abs} are not strictly convex and thus have a nonempty null space. (For example, when $\mathbf{f} = \mathbf{m}$, both forms of prior are zero.) Unfortunately, we are not yet able to prove that the null spaces do not overlap, but we suspect nevertheless that the MAP solution $(\hat{\mathbf{f}},\hat{\mathbf{m}})$ is indeed unique. In many simulations, we have not observed any dependence of the MAP solution on the initial estimate. Further work is required here.

The contribution in this work lies mainly in the proposed use of an auxiliary field $\mathbf{m}$ to define priors for use in joint MAP reconstructions. For the median prior proposed here, the use of an auxiliary field has allowed us to formulate an MRP-like method but place it into a MAP context, avoiding the use of an empirical data-dependent penalty. The use of an auxiliary field is far more flexible, however, than our MP application implies. For example, as seen in our previous work [8], we used a form of $\Phi_P(\mathbf{f}, \mathbf{m})$, completely different from that in (2), whose net effect was to obtain a smoothed and positivity constrained solution to a tomographic MAP reconstruction. It will be interesting to further explore different forms for $\Phi_P(\mathbf{f}, \mathbf{m})$ to derive novel and useful priors.

References

[1] S. Alenius and U. Ruotsalainen, "Bayesian image reconstruction for emission tomography based on median root prior," *E. J. Nuc. Med.*, vol. 24, no. 3, pp. 258–265, Mar. 1997.

[2] S. Alenius, U. Ruotsalainen, and J. Astola, "Using local median as the location of the prior distribution in iterative emission tomography image reconstruction," *IEEE Tran. Nuc. Sci.*, vol. 45, no. 6, pp. 3097–3107, Dec. 1998.

[3] S. J. Lee, A. Rangarajan, and G. R. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Transactions on Medical Imaging*, vol. vol. 14, pp. 669–680, 1995.

[4] D. Luenberger, *Linear and Nonlinear Programming - 2nd ed.*, Addison-Wesley, Menlo Park, CA, 1984.

[5] P. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[6] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Proc.*, vol. 2, no. 3, pp. 296–310, July 1993.

[7] J. A. Fessler and L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomography," *IEEE Trans. Img. Proc.*, vol. 5, pp. 1346–1358, Sept. 1996.

[8] I. Hsiao, A. Rangarajan, Y. Xing, and G. Gindi, "A smoothing prior with embedded positivity constraint for tomographic reconstruction," in *Sixth International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, R. Huesman and J. Qi, Eds. 2001, pp. 81 – 84, Kluwer.