

# A Theoretical Study of the Contrast Recovery and Variance of MAP Reconstructions from PET Data

Jinyi Qi,\* *Member, IEEE*, and Richard M. Leahy, *Member, IEEE*

**Abstract**—We examine the spatial resolution and variance properties of PET images reconstructed using maximum *a posteriori* (MAP) or penalized-likelihood methods. Resolution is characterized by the contrast recovery coefficient (CRC) of the local impulse response. Simplified approximate expressions are derived for the local impulse response CRC's and variances for each voxel. Using these results we propose a practical scheme for selecting spatially variant smoothing parameters to optimize lesion detectability through maximization of the local CRC-to-noise ratio in the reconstructed image.

**Index Terms**—Hyperparameter estimation, image resolution, MAP estimation, variance analysis.

## I. INTRODUCTION

PET image reconstruction algorithms based on maximum likelihood (ML) or maximum *a posteriori* (MAP) principles can produce improved spatial resolution and noise properties in comparison to conventional filtered backprojection (FBP) methods. It is often important to be able to quantify this improvement in terms of the resolution (or bias) and variance of the resulting images. These measures can be used in comparing different reconstruction algorithms on a particular imaging system or for comparing different system configurations. Similarly, estimates of bias and variance from a single data set in quantitative studies are important measures of the reliability of a point estimate of tracer uptake.

Bias and variance for filtered backprojection images are readily computed using a Poisson data model since the estimator is linear [1]. However, ML and MAP estimators are nonlinear functions of the data that are computed using iterative algorithms. Closed-form solutions are not generally available. The situation is further complicated in the case of ML methods, since the iterations are often terminated before convergence [2]. Estimator performance can always be evaluated using Monte Carlo techniques, but the computational costs will often make this approach impractical. A great deal of progress has recently been made in developing approximate closed-form expressions for the bias and variance of these estimators, as we briefly review below. In this paper we build on this work by developing new approximations for the con-

trast recovery and voxelwise variances of MAP reconstruction algorithms.

Since the EM algorithm for ML reconstruction [3] is rarely iterated to convergence, Barrett *et al.* [2] derived approximate formulae for the mean and covariance of the reconstructed image as a function of the iteration number. Monte Carlo validations [4] showed that these theoretical predictions matched Monte Carlo estimates for the earlier iterations at which the algorithm is usually terminated. In high count situations, these results were also accurate for larger numbers of iterations. Wang and Gindi [5] extended this approach to two special cases of MAP EM algorithms: a MAP EM algorithm incorporating an independent gamma prior [6], and the one-step-late MAP-EM algorithm using a multivariate Gaussian prior [7].

This iteration-based approach is attractive for methods that are terminated before convergence, as is common practice for the EM algorithm and its ordered-subsets variants [8]. However, for the MAP algorithms in [5], the complexity of these expressions gives little intuitive insight into the effect of the smoothing parameter on bias and variance and evaluation of the expressions for large numbers of iterations is time consuming. These methods also require explicit update equations, so that they are inapplicable to numerical optimization methods such as gradient or coordinate-wise ascent that involve line searches.

An alternative approach was proposed by Fessler and Rogers [9], [10] who analyzed the mean, variance, and spatial resolution at a fixed point of the objective function. The resolution and noise properties are computed at the fixed point using partial derivatives and truncated Taylor series approximations. These results are independent of the particular optimizing algorithm used and require only that the algorithm be iterated to effective convergence. The derivation does not account for positivity constraints, so the results are valid only where the solution is strictly positive. These results show very good agreement with Monte Carlo results, except in areas where the activity is very low. By using the image estimated from a single data set in place of the true image, a plug-in form of the variance approximation can be used to compute a surprisingly good estimate.

While closed-form equations are derived in [10], they involve computing the inverse of a Hessian matrix or solving a related set of linear equations. These expressions provide little direct insight into the relationship between estimator bias and variance and the parameters of the imaging system and reconstruction algorithm. In addition, computational cost remains high except for the case of MAP with a modified quadratic prior, where a set of linear equations can be presolved [11].

Manuscript received June 29, 1998; revised February 4, 1999. This work was supported in part by the National Cancer Institute under Grants R01 CA59794 and R01 CA56655. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was J. Fessler. *Asterisk indicates corresponding author.*

\*J. Qi and R. M. Leahy are with the Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: leahy@sipi.usc.edu).

Publisher Item Identifier S 0278-0062(99)04503-6.

Here, we analyze the resolution and variance properties of MAP reconstruction methods. Because several fast convergent algorithms have been developed for MAP reconstruction [12], [13], [14] and penalized weighted least squares [15]–[17], we have adopted the approach in [9] and [10] and investigate the properties of the MAP estimator at a fixed point of the objective function, rather than as a function of iteration. The resolution is characterized by the local impulse response contrast recovery coefficient (CRC). We derive simplified approximate equations for local impulse response CRC's and variance for each voxel. The approximations used are in the spirit of previous studies of resolution and variance [9]–[11] and in the design of preconditioners for conjugate gradient methods [17]. However, the approximation used here results in expressions that are rational polynomials of the eigenvalues of matrices computed from the system geometric response and the Gibbs energy of the prior. The eigenvalues are found using a two-dimensional (2-D) Fourier transform by assuming that the geometric response of PET is spatially invariant. The rational polynomials are easily computed in comparison to the iterative formulae, or those based on inverse Hessian computations. We use these results to examine the relationship of the CRC and variance to the hyperparameter and voxelwise Fisher information.

Using the plug in approach, we can also use these CRC and variance expressions to select the smoothing parameter  $\beta$ . This parameter controls the tradeoff between smoothing (bias) and noise propagation (variance). Within the Bayesian framework, the parameters of the prior, if unknown, should be treated as random quantities for which a hyperprior density is specified. The parameters can then either be marginalized out or estimated directly from the data. In practice, these parameters are more commonly chosen either using approximate maximum likelihood methods [18]–[20] or cross validation techniques [21], [22].

An alternative approach to parameter selection was proposed by Fessler and Rogers [10] who noted that using a constant  $\beta$  parameter for a quadratic prior with Poisson data results in a spatially variant resolution. This observation was then used as the basis for defining a spatially variant  $\beta$  which compensates for this effect to produce uniform resolution throughout the image. Here we describe an alternative means of hyperparameter selection that may be more appropriate for use in lesion detection tasks. The expressions developed for CRC and noise variance can be used to compute a voxelwise contrast to noise ratio (CNR) as a function of the data and the hyperparameter. In computing these factors we use the observed data in place of its expected value in the approximate CRC and variance expressions. We show that the CNR has a unique maximum so that  $\beta$  can be specified to maximize CNR at each voxel. The cost of computing these optimized parameters is small compared to that of reconstructing the image. Since the prior in this case is an explicit function of the data, the MAP method described here is not a true Bayesian estimator. For consistency with our previous publications, however, we will continue to refer to the general procedure of maximizing over the posterior probability as a MAP method, regardless of the method by which the hyperparameter is selected.

The paper is organized as follows. In Section II we develop approximate expressions for the CRC and variance of each

voxel. Although we focus here on PET image reconstruction, these results are extendible to other linear image reconstruction problems using photon limited data. In Section III we study these approximations for the special case of Gaussian priors. The method for choosing the smoothing parameters to maximize the CNR is described in Section IV. Simulation studies which evaluate the accuracy of the approximations developed in the paper are included in Section V.

## II. THEORY

### A. MAP Reconstruction

We use the standard Poisson model for the PET data with log-likelihood

$$L(\mathbf{y} | \mathbf{x}) = \sum_i y_i \log \bar{y}_i - \bar{y}_i, \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^{N \times 1}$  is the image,  $\mathbf{y} \in \mathbb{R}^{M \times 1}$  is the measured sinogram, and  $\bar{\mathbf{y}} \in \mathbb{R}^{M \times 1}$  is the mean of the sinogram. The mean sinogram  $\bar{\mathbf{y}}$  is related to the image,  $\mathbf{x}$ , through an affine transform

$$\bar{\mathbf{y}} = \mathbf{P}\mathbf{x} + \mathbf{r} \quad (2)$$

where  $\mathbf{P} \in \mathbb{R}^{M \times N}$  is the detection probability matrix and  $\mathbf{r} \in \mathbb{R}^{M \times 1}$  accounts for the presence of scatter and randoms in the data. Here we assume that the data are not precorrected for randoms and that an estimate of  $\mathbf{r}$  is available. Randoms corrected data are not Poisson so the following expressions are not applicable. However, the shifted-Poisson model [23], which accounts for the increased variance that results from random subtraction, can be used as the basis for developing analogous expressions to those below.

In our previous work [24], [25] we have used a factored detection probability matrix  $\mathbf{P}$  to accurately model factors such as depth-dependent geometric sensitivities and spatially variant detector response. In the following we use the simpler model  $\mathbf{P} = D[n_i]\mathbf{G}$  where  $D[n_i], i = 1, \dots, M$  is a diagonal matrix containing the product of the detector normalization, dead-time, and attenuation correction factors.<sup>1</sup>  $\mathbf{G}$  is the geometric projection matrix representing the probability that an emission from each voxel in the image produces, in the absence of attenuation effects, a photon pair at each of the detector pairs in the system. We will later assume that  $\mathbf{G}$  has the property that the operator  $\mathbf{G}'\mathbf{G}$  produces a shift-invariant blurring. This model cannot, therefore, directly include shift-variant detector blurring, nor is it directly extendible to fully three-dimensional (3-D) data acquisition. We address these issues in a separate publication [26].

The results below apply to any prior with a Gibbs energy  $U(\mathbf{x})$  that can be written in the form [17]

$$U(\mathbf{x}) = \sum_{k=1}^P \phi_k([\mathbf{C}\mathbf{x} - \mathbf{m}]_k) \quad (3)$$

where  $\mathbf{m} \in \mathbb{R}^{P \times 1}$  is an arbitrary vector and  $\mathbf{C} \in \mathbb{R}^{P \times N}$  is a sparse neighborhood matrix whose  $k$ th row has nonzero

<sup>1</sup> $D[n_i]$  denotes a diagonal matrix with the  $(i, i)$ th diagonal element equal to  $n_i$ .

elements corresponding to the voxels which form the  $k$ th clique. The specific values of these elements will in most cases, including those considered here, correspond to the weights necessary to compute approximate first- or second-order spatial derivatives of the image. For example, in a simple pairwise membrane model,  $\mathbf{m} = \mathbf{0}$  and  $[\mathbf{C}\mathbf{x}]_k$  is equal to the difference between the  $k$ th pair of neighboring voxels. The function  $\phi_k(\cdot)$  is a convex function of its argument. The number of terms  $p$  depends on the size of the neighborhood and the number of potential functions defined on cliques within this neighborhood system [27]. As with the geometric projection matrix, we later assume that  $\mathbf{C}'\mathbf{C}$  is a shift-invariant operator. This property will hold in the interior of the image, provided that the neighborhood system and cliques are homogeneous throughout the image. Behavior of  $\mathbf{C}'\mathbf{C}$  near the image boundary will depend on the boundary conditions used in defining the cliques and clique potentials.

The MAP reconstruction is found as the maximizer of the log posterior probability

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg \max_{\mathbf{x}} L(\mathbf{y} | \mathbf{x}) - \beta U(\mathbf{x}) \quad (4)$$

where  $\beta$  is the hyperparameter that determines the relative influence of the prior and likelihood terms.

### B. Resolution and Local Impulse Response CRC's

In this and the following subsection, we develop approximate expressions for the local impulse response CRC and for the variance of each voxel of the MAP reconstruction. The initial development follows the methods used in [9], [10], and [17], however, diagonalization of an approximate matrix inverse allows us to obtain simpler expressions in terms of the eigenvalues of the matrices  $\mathbf{G}'\mathbf{G}$  and  $\mathbf{C}'\mathbf{C}$ .

The local impulse response [10] for the  $j$ th voxel is defined as

$$l^j(\hat{\mathbf{x}}) = \lim_{\delta \rightarrow 0} \frac{\mathcal{E}\hat{\mathbf{x}}(\mathbf{y}(\mathbf{x} + \delta\mathbf{e}_j)) - \mathcal{E}\hat{\mathbf{x}}(\mathbf{y}(\mathbf{x}))}{\delta} \quad (5)$$

where  $\mathcal{E}$  denotes the expectation operator,  $\hat{\mathbf{x}}(\mathbf{y})$  is the reconstruction from data  $\mathbf{y}$ ,  $\mathbf{y}(\mathbf{x})$  is the projection data from tracer distribution  $\mathbf{x}$ , and  $\mathbf{e}_j$  is the  $j$ th unit vector.  $l^j(\hat{\mathbf{x}})$  is dependent on the object, the system, and the estimator.

Using a first-order Taylor series approximation and the chain rule, one can approximate the local impulse response  $l^j(\hat{\mathbf{x}})$  of the MAP reconstruction as [10]

$$l^j(\hat{\mathbf{x}}) \approx \frac{\partial}{\partial x_j} \hat{\mathbf{x}}(\mathbf{y}(\mathbf{x})) \approx [\mathbf{F} + \beta \mathbf{C}'\mathbf{D}_{\ddot{\phi}}(\hat{\mathbf{x}})\mathbf{C}]^{-1} \mathbf{F}\mathbf{e}_j \quad (6)$$

where  $\mathbf{F} \stackrel{\text{def}}{=} \mathbf{P}'D[1/\bar{y}_i]\mathbf{P}$  is the Fisher information matrix,  $\mathbf{D}_{\ddot{\phi}}(\hat{\mathbf{x}}) = D[\ddot{\phi}_k([\mathbf{C}\hat{\mathbf{x}} - \mathbf{c}]_k)]$  and  $\ddot{\phi}(t) = d^2/dt^2\phi(t)$ .

In the following, we will consider the local impulse response CRC [28]  $crc_j \stackrel{\text{def}}{=} l^j_j(\hat{\mathbf{x}})$  as a measure of resolution where  $l^j_j(\hat{\mathbf{x}})$  is the  $j$ th element of  $l^j(\hat{\mathbf{x}})$ . While the measure does not directly reflect the conventional measures such as full width at half maximum (FWHM) and tenth maximum (FWTM), these measures will tend to be highly correlated since a larger CRC will produce a narrower FWHM by a simple conservation of activity argument. For the task of lesion detection, the CRC

is also a measure that has been shown to be correlated with lesion detectability [29]. The obvious advantage of working with the CRC is that it is a simple function of the local impulse response, while the FWHM is not.

We now adopt the approximations proposed in [10] and [17] for the Fisher information matrix and the second-order derivative of the prior

$$\mathbf{F} \approx \mathbf{D}_{\kappa} \mathbf{G}'\mathbf{G} \mathbf{D}_{\kappa} \quad (7)$$

$$\beta \mathbf{C}'\mathbf{D}_{\ddot{\phi}}(\hat{\mathbf{x}})\mathbf{C} \approx \mathbf{D}_{\kappa} \mathbf{D}_{\eta}(\hat{\mathbf{x}})\mathbf{C}'\mathbf{C} \mathbf{D}_{\eta}(\hat{\mathbf{x}})\mathbf{D}_{\kappa} \quad (8)$$

where  $\mathbf{D}_{\kappa} = D[\kappa_j]$  with

$$\kappa_j \stackrel{\text{def}}{=} \sqrt{\frac{\sum_i g_{ij}^2 n_i^2 / \bar{y}_i}{\sum_i g_{ij}^2}} \quad (9)$$

$\mathbf{D}_{\eta}(\hat{\mathbf{x}}) = D[\eta_j(\hat{\mathbf{x}})]$  with

$$\eta_j(\hat{\mathbf{x}}) \stackrel{\text{def}}{=} \sqrt{\frac{\beta \sum_k c_{kj}^2 \ddot{\phi}_k([\mathbf{C}\hat{\mathbf{x}} - \mathbf{m}]_k)}{\sum_k c_{kj}^2}}. \quad (10)$$

Using (7) and (8), (6) can be simplified to

$$l^j(\hat{\mathbf{x}}) \approx \mathbf{D}_{\kappa}^{-1} \mathbf{B}(\hat{\mathbf{x}})^{-1} \mathbf{G}'\mathbf{G} \mathbf{D}_{\kappa} \mathbf{e}_j \quad (11)$$

where

$$\mathbf{B}(\hat{\mathbf{x}}) \stackrel{\text{def}}{=} \mathbf{G}'\mathbf{G} + \mathbf{D}_{\eta}(\hat{\mathbf{x}})\mathbf{C}'\mathbf{C} \mathbf{D}_{\eta}(\hat{\mathbf{x}}). \quad (12)$$

Using the local property of the local impulse response  $l^j$ , we can approximate the inverse of  $\mathbf{B}(\hat{\mathbf{x}})$  using

$$\mathbf{B}(\hat{\mathbf{x}})^{-1} \mathbf{e}_j \approx \mathbf{K}(\eta_j^2(\hat{\mathbf{x}}))^{-1} \mathbf{e}_j \quad (13)$$

where  $\mathbf{K}(\eta) = \mathbf{G}'\mathbf{G} + \eta \mathbf{C}'\mathbf{C}$ . This approximation is essentially the same as that used in (34) in [10]. Although the approximation has no apparent optimal properties, note that in the event that  $\mathbf{D}_{\eta}(\hat{\mathbf{x}})$  is a constant diagonal matrix, the approximation becomes exact. Provided the diagonal elements are smooth (see Section V) in the vicinity of the  $j$ th voxel (as determined by the spatial extent of  $\mathbf{B}(\hat{\mathbf{x}})^{-1} \mathbf{e}_j$ ) then the approximation should be reasonably accurate. Note also, that rather than being a single approximation, this is in fact a sequence of approximations, one per voxel, so that the CRC can be computed using a different approximation for each voxel.

Substituting (13) into (11) we can compute the  $j$ th element of  $l^j(\hat{\mathbf{x}})$  as

$$\begin{aligned} l^j_j(\hat{\mathbf{x}}) &\approx \mathbf{e}'_j \mathbf{D}_{\kappa}^{-1} \mathbf{B}(\hat{\mathbf{x}})^{-1} \mathbf{G}'\mathbf{G} \mathbf{D}_{\kappa} \mathbf{e}_j \\ &= \mathbf{e}'_j \kappa_j^{-1} \mathbf{B}(\hat{\mathbf{x}})^{-1} \mathbf{G}'\mathbf{G} \kappa_j \mathbf{e}_j \\ &= \mathbf{e}'_j \mathbf{K}(\eta_j^2(\hat{\mathbf{x}}))^{-1} \mathbf{G}'\mathbf{G} \mathbf{e}_j. \end{aligned} \quad (14)$$

Up to this point, the results are essentially a combination of the development in [10] and [17]. Now we proceed by using Fourier transform theory to analyze (14), which results in a simplified expression for the local impulse response. We also apply the same method to the variance analysis in Section II-C.

Since  $\mathbf{G}'\mathbf{G}$  and  $\mathbf{C}'\mathbf{C}$  correspond to shift-invariant blurring operators, they have a block Toeplitz structure and can be

approximately diagonalized using the 2-D discrete Fourier transform (DFT) [17], [30], i.e.,

$$\mathbf{G}'\mathbf{G} \approx \mathbf{Q}'D[\lambda_i]\mathbf{Q} \quad (15)$$

$$\mathbf{C}'\mathbf{C} \approx \mathbf{Q}'D[\mu_i]\mathbf{Q} \quad (16)$$

where  $\mathbf{Q}$  and  $\mathbf{Q}'$  represent the Kronecker form of the 2-D DFT matrix and its inverse, respectively [30]. The  $\lambda_i$ 's are the eigenvalues of the system geometric response matrix inner product  $\mathbf{G}'\mathbf{G}$ , and the  $\mu_i$ 's are the eigenvalues of the neighborhood matrix inner product  $\mathbf{C}'\mathbf{C}$ .

Combining these diagonalizations with (14) and using the unitary property of the DFT

$$\begin{aligned} l_j^j(\hat{\mathbf{x}}) &\approx \mathbf{e}_j'(\mathbf{Q}'D[\lambda_i]\mathbf{Q} + \eta_j^2(\hat{\mathbf{x}})\mathbf{Q}'D[\mu_i]\mathbf{Q})^{-1}\mathbf{Q}'D[\lambda_i]\mathbf{Q}\mathbf{e}_j \\ &= \mathbf{e}_j'\mathbf{Q}'D\left[\frac{\lambda_i}{\lambda_i + \eta_j^2(\hat{\mathbf{x}})\mu_i}\right]\mathbf{Q}\mathbf{e}_j \\ &= \frac{1}{N}\sum_{i=0}^{N-1}\frac{\lambda_i}{\lambda_i + \eta_j^2(\hat{\mathbf{x}})\mu_i}. \end{aligned} \quad (17)$$

The local impulse CRC can therefore be expressed in the compact form

$$crc_j = l_j^j(\hat{\mathbf{x}}) = \frac{1}{N}\sum_{i=0}^{N-1}\frac{\lambda_i}{\lambda_i + \eta_j^2(\hat{\mathbf{x}})\mu_i}. \quad (18)$$

### C. Approximation of the Variance

One can similarly show that the covariance of the MAP reconstruction can be approximated as [9], [11]

$$\text{Cov}(\hat{\mathbf{x}}) \approx [\mathbf{F} + \beta\mathbf{C}'\mathbf{D}_{\phi}(\hat{\mathbf{x}})\mathbf{C}]^{-1}\mathbf{F}[\mathbf{F} + \beta\mathbf{C}'\mathbf{D}_{\phi}(\hat{\mathbf{x}})\mathbf{C}]^{-1}. \quad (19)$$

Using the fact that  $l^j(\hat{\mathbf{x}}) \approx \partial\hat{\mathbf{x}}/\partial x_j$ , it can be easily verified that the right-hand side of (19) is approximately equal to the Cramer–Rao bound (CRB) for a biased estimator [31]

$$\sigma^2(\hat{\mathbf{x}}) \geq [\nabla\hat{\mathbf{x}}]'\mathbf{F}^+[\nabla\hat{\mathbf{x}}] \quad (20)$$

where  $\nabla\hat{\mathbf{x}} = [l^1(\hat{\mathbf{x}}) \ l^2(\hat{\mathbf{x}}) \ \dots \ l^N(\hat{\mathbf{x}})]'$  is the gradient matrix and  $\mathbf{F}^+$  denotes the Moore–Penrose pseudoinverse of the Fisher information matrix.

Substituting (7) and (8) into (19) we have

$$\text{Cov}(\hat{\mathbf{x}}) \approx \mathbf{D}_{\kappa}^{-1}\mathbf{B}(\hat{\mathbf{x}})^{-1}\mathbf{G}'\mathbf{G}\mathbf{B}(\hat{\mathbf{x}})^{-1}\mathbf{D}_{\kappa}^{-1}. \quad (21)$$

Here we are interested only in the variance at each voxel, i.e., the diagonal elements of  $\text{Cov}(\hat{\mathbf{x}})$ . Using the approximation (13) in Section II-B, the variance at the  $j$ th voxel can be expressed as

$$\begin{aligned} var_j(\hat{\mathbf{x}}) &\approx \mathbf{e}_j'\mathbf{D}_{\kappa}^{-1}\mathbf{B}(\hat{\mathbf{x}})^{-1}\mathbf{G}'\mathbf{G}\mathbf{B}(\hat{\mathbf{x}})^{-1}\mathbf{D}_{\kappa}^{-1}\mathbf{e}_j \\ &= \mathbf{e}_j'\kappa_j^{-1}\mathbf{B}(\hat{\mathbf{x}})^{-1}\mathbf{G}'\mathbf{G}\mathbf{B}(\hat{\mathbf{x}})^{-1}\kappa_j^{-1}\mathbf{e}_j \\ &= \kappa_j^{-2}\mathbf{e}_j'\mathbf{K}(\eta_j^2(\hat{\mathbf{x}}))^{-1}\mathbf{G}'\mathbf{G}\mathbf{K}(\eta_j^2(\hat{\mathbf{x}}))^{-1}\mathbf{e}_j \\ &= \kappa_j^{-2}\mathbf{e}_j'\mathbf{Q}'D\left[\frac{\lambda_i}{(\lambda_i + \eta_j^2(\hat{\mathbf{x}})\mu_i)^2}\right]\mathbf{Q}\mathbf{e}_j \\ &= \frac{\kappa_j^{-2}}{N}\sum_i\frac{\lambda_i}{(\lambda_i + \eta_j^2(\hat{\mathbf{x}})\mu_i)^2}. \end{aligned} \quad (22)$$

Equations (18) and (22) are the key results of the paper. These are simple and readily computed expressions for the local contrast recovery and variances of individual voxels. Since our starting points, (6) and (19), are first-order Taylor series approximations involving derivatives of the objective function up to the second-order only, we would expect the contrast recovery and variance expressions to be most accurate for near quadratic cost functions. The Poisson distribution is well approximated by a Gaussian distribution for all but the smallest mean count rates ( $\leq 7$ ), so that we may expect the approximations to be fairly accurate for MAP estimates with Gaussian priors. For non-Gaussian priors, accuracy will depend on the degree of local nonquadratic behavior. In the following two sections we consider the case where the prior is Gaussian, which leads to further simplification of the results (18) and (22). We return to non-Gaussian priors in Section II-C.

### III. THE CRC AND VARIANCE FOR GAUSSIAN PRIORS

Here we examine the case of Gaussian Markov random field priors with the potential function  $\phi(t) = t^2/2$ . This class includes a number of Gibbs energy functions popular in the PET imaging literature, including the pairwise quadratic or membrane energy function defined on first- (four-neighbor) or second- (eight-neighbor) order neighborhoods [32] and the thin-plate spline energy defined on a third-order (12-neighbor) neighborhood [32], [33].

For these energy functions  $\mathbf{D}_{\phi}(\hat{\mathbf{x}}) = \mathbf{I}$ , so the factors  $\eta_j(\hat{\mathbf{x}})$  in (10) simplify to

$$\eta_j(\hat{\mathbf{x}}) = \frac{\sqrt{\beta}}{\kappa_j}. \quad (23)$$

The CRC and variances are then

$$crc_j \approx \frac{1}{N}\sum_{i=0}^{N-1}\frac{\lambda_i}{\lambda_i + \beta\kappa_j^{-2}\mu_i} \quad (24)$$

$$var_j \approx \frac{\kappa_j^{-2}}{N}\sum_{i=0}^{N-1}\frac{\lambda_i}{(\lambda_i + \beta\kappa_j^{-2}\mu_i)^2}. \quad (25)$$

When  $\beta = 0$ , (24) and (25) simplify to

$$crc_j \approx 1 \quad (26)$$

$$var_j \approx \frac{\kappa_j^{-2}}{N}\sum_{i=0}^{N-1}\frac{1}{\lambda_i}. \quad (27)$$

In this case, the MAP estimate reduces to ML. The right-hand side of (27) is the  $(j, j)$ th diagonal element of the inverse of the Fisher information matrix  $\mathbf{F}^{-1}$  under approximation (7), i.e. the CRB for an unbiased estimator.

Most priors, including those using membrane and thin-plate energies, are based on balanced voxel differences so that the matrix  $\mathbf{C}'\mathbf{C}$  will have a zero eigenvalue,  $\mu_0 = 0$ , corresponding to the DC component of the 2-D DFT. For these priors, when  $\beta \gg \lambda_i\kappa_j^2/\mu_i, i = 1, \dots, N-1$ , (24) and (25) reduce to

$$crc_j \approx \frac{1}{N} \quad (28)$$

$$var_j \approx \frac{\kappa_j^{-2}}{N}\frac{1}{\lambda_0}. \quad (29)$$

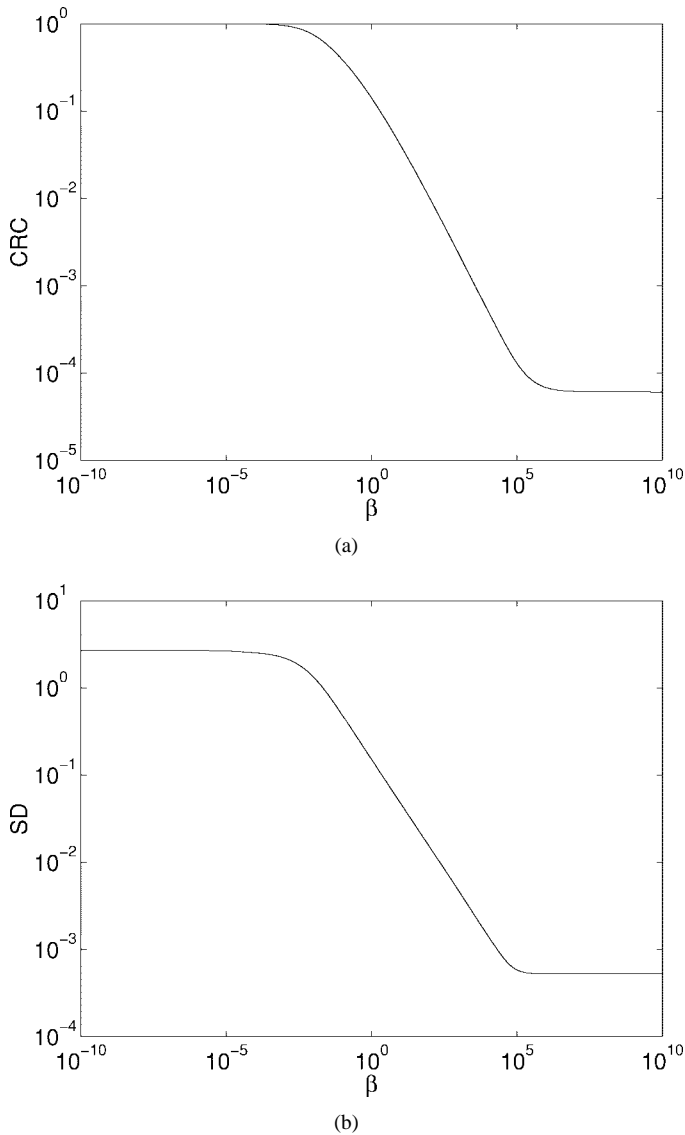


Fig. 1. (a) CRC and (b) standard deviation versus  $\beta$  curves computed using (24) and (25) with  $\kappa = 1$  for a quadratic prior with a second-order neighborhood.

This case is of little practical interest since our experience is that such a large value of  $\beta$  is of no practical value.

When  $\beta\kappa_j^{-2}$  is between these two extremes ( $0 \leq \beta \ll \max_{1 \leq i \leq N-1} \{\lambda_i \kappa_j^2 / \mu_i\}$ ), which corresponds to the practical operating range for MAP estimators, both CRC and variance reduce monotonically with  $\beta$  where the lower inflection point is approximately  $\beta \approx \kappa^2 \min_{1 \leq i \leq N-1} \{\lambda_i / \mu_i\}$  and the higher  $\beta \approx \kappa^2 \max_{1 \leq i \leq N-1} \{\lambda_i / \mu_i\}$ .

To illustrate these results, we first plot the CRC and standard deviation versus  $\beta$  curves keeping  $\kappa = 1$ , in Fig. 1. The PET system that was simulated to compute these curves is the CTI EXACT HR+ scanner with a sinogram size equal to  $288 \times 288$ . The reconstruction voxel size was  $2.3 \text{ mm} \times 2.3 \text{ mm}$ . The prior used was the membrane potential with a second-order neighborhood. The  $\min_i \{\lambda_i / \mu_i\}$  and  $\max_i \{\lambda_i / \mu_i\}$  in this case are  $2.13 \times 10^{-3}$  and  $1.65 \times 10^4$ , respectively. Note that these curves are dependent only on the system geometry, the prior, and the voxel size, and are independent of the

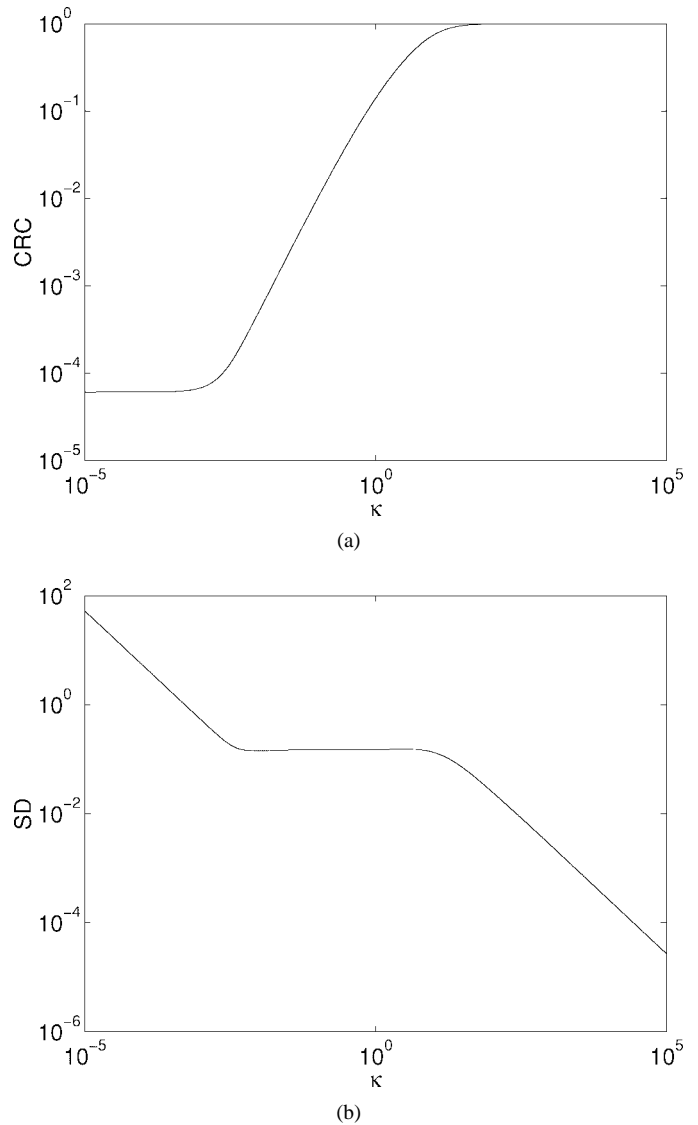


Fig. 2. (a) CRC and (b) standard deviation versus  $\kappa$  curves computed using (24) and (25) with  $\beta = 1$  for a quadratic prior with a second-order neighborhood.

particular image. In Fig. 2 we show the CRC and standard deviation versus  $\kappa$  curves, this time keeping  $\beta = 1$ . Equation (24) shows that  $\beta$  and  $\kappa^{-2}$  have the same effect on the CRC. Fig. 2(b), however, shows that the standard deviation remains nearly constant for values of  $\kappa$  in a middle range, and increases linearly with  $\kappa$  on either side of this range.

Our experience has shown that the useful range of  $\beta\kappa^{-2}$ , i.e., that which produces generally reasonable reconstructions, is approximately 0.1–10 (assuming  $\sum_i g_{ij}^2 \approx 1$ ). The results in Fig. 2(b) show that if a constant  $\beta$  value is used across the entire image, then in the range  $0.1 \leq \beta\kappa^{-2} \leq 10$ , the reconstructions will have near constant variance and CRC's approximately proportional to  $\kappa_j$ . This result is consistent with the examples of nearly constant variance in MAP reconstructions reported in [5] and [11]. For different PET system configurations, or different priors, the curves in Figs. 1 and 2 will change. It is interesting to note, for example, that for the case of the thin-plate prior there is no constant standard deviation range for  $\kappa$ .

#### IV. MAXIMUM DETECTABILITY HYPERPARAMETER SELECTION

One of the major problems in using MAP methods in practical applications is selecting an appropriate hyperparameter  $\beta$ . Fessler and Rogers [10] proposes using a spatially variant  $\beta$  that compensates for the spatially variant resolution properties that result from using a quadratic penalty or prior. Here we propose an alternative hyperparameter-selection scheme based on the results derived above.

It is well known that detection of lesions, one of the primary applications of PET, is affected by both the level of background noise and by contrast of the lesion. Selecting the hyperparameter effectively involves performing a tradeoff between low background noise and good contrast recovery. Although lesion detectability is affected by the correlation of the noise in the reconstruction [34], empirical studies have shown that human observer performance is reasonably well correlated with measures of the ratio of contrast recovery to background noise standard deviation [29]. Consequently, we propose choosing  $\beta$  to maximize this ratio. The CNR can be computed in a straightforward manner using the results derived above as

$$\text{CNR} \stackrel{\text{def}}{=} \frac{\text{crc}_j}{\sqrt{\text{var}_j}}. \quad (30)$$

Assuming that the reconstructions are locally ergodic (i.e., the ensemble variances approximated using the expression above are close to the spatial variances of a local region computed from an individual reconstruction), then the CNR is equivalent to the SNR used in [29] and [35], which has been shown to be correlated with both human and machine observers for cold-spot detectability.

From (18) and (22) we know that the CNR's are different for each voxel location so the hyperparameter  $\beta$  needs to be spatially variant in order to maximize the ratio. The maximum detectability hyperparameter for quadratic priors at the  $j$ th voxel can be computed as

$$\beta_j^* = \arg \max_{\beta} \frac{\text{crc}_j}{\sqrt{\text{var}_j}} \quad (31)$$

$$= \arg \max_{\beta} \psi(\beta \kappa_j^{-2}) \quad (32)$$

where

$$\psi(\eta^2) \stackrel{\text{def}}{=} \sum_i \frac{\lambda_i}{\lambda_i + \eta^2 \mu_i} / \sqrt{\sum_i \frac{\lambda_i}{(\lambda_i + \eta^2 \mu_i)^2}} \quad (33)$$

and  $\psi(\eta^2)$  is a function of  $\eta$  only. This means that the optimal  $\eta$  is independent of the data. Consequently, we can first find the optimal  $\eta$  without knowledge of the data and then compute the optimal  $\beta$  for each voxel by computing  $\kappa_j$  from (9) and using (23). When applied in practice, we need to replace the mean data in (9) with its observed realization, but the smoothing effect of the modified backprojection in (9) results in estimated  $\kappa_j$  values which are very similar to those that would be obtained using the true mean data.

Fig. 3 shows the CNR as a function of  $\beta$  with  $\kappa = 1$  for the EXACT HR+ PET system with first- and second-order membrane potentials and a third-order thin-plate potential.

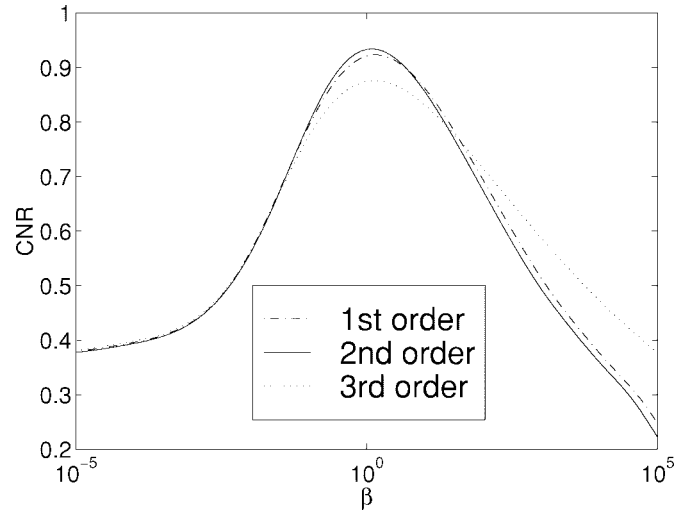


Fig. 3. CNR versus  $\beta$  curves for three different priors (first- and second-order membrane and third-order thin-plate spline energy functions). Note the unique maximum for each prior and that the second-order membrane prior produces larger maximum CNR.

These curves show unique global maxima, all at approximately  $\eta^{*2} (= \beta \kappa^{-2}) \approx 1.26$ . The fact that the second-order neighborhood shows the highest CNR among the three, indicates that it may be most suitable for small lesion detection. The thin-plate model does not work well here because the local impulse response has relatively large second-order derivatives and hence is highly penalized by the thin-plate model.

For quadratic priors, the hyperparameter  $\beta_j$  can be computed from  $\eta^*$  using (23)

$$\beta_j = \eta^{*2} \kappa_j^2. \quad (34)$$

The analysis above assumes, when deriving the CRC and variance for a particular voxel, that  $\beta$  is constant across the image. Since the  $\beta$  that maximizes the CNR is a function of the  $\kappa_j$ 's, which are in turn a function of a modified backprojection of the data, we would expect the optimal  $\beta$ 's to vary smoothly across the image. The  $\beta$  value used for each potential term is computed as the geometric average of the optimal  $\beta$ 's computed for each of the voxels contributing to the clique on which the potential is defined, i.e., we use  $\sqrt{\beta_i \beta_j}$  for terms involving the two voxels  $i$  and  $j$  and  $(\beta_i \beta_j \beta_k)^{1/3}$  for terms involving three voxels  $i$ ,  $j$ , and  $k$ .

From (24) and (25) we know that when using the optimal  $\beta$ 's,  $\beta \kappa^{-2}$  is constant and hence the local impulse response CRC's will be approximately constant across the field of view and the voxel variances will be proportional to  $\kappa_j^{-2}$ . This uniform contrast recovery property may be helpful for lesion detection by human operators.

To summarize the hyperparameter selection scheme for maximizing CNR when using quadratic priors.

- 1) Compute  $\kappa_j$ ,  $j = 1, 2, \dots, N$  from (9) using  $\hat{y}_i = \max\{1, y_i\}$  in place of  $\bar{y}_i$ .
- 2) Compute  $\beta_j$ ,  $j = 1, 2, \dots, N$  from (34).
- 3) Find the MAP estimate using the spatially variant hyperparameters  $\beta_j$ ,  $j = 1, 2, \dots, N$ .

## V. SIMULATION RESULTS

### A. Validation of Approximations of CRC and Variance: Quadratic Priors

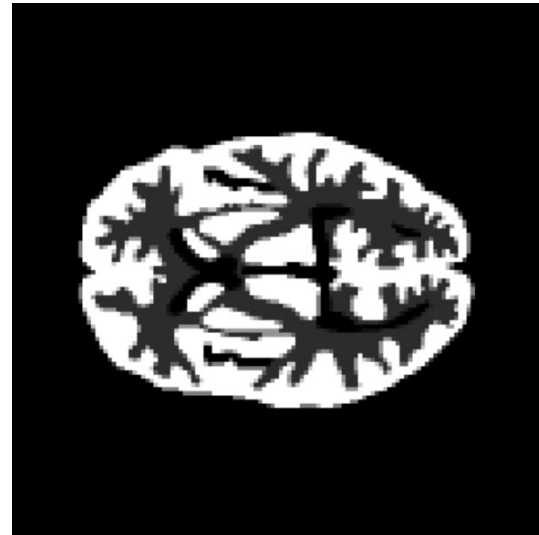
To test the accuracy of the approximate CRC and variance expressions, we simulated data for the CTI EXACT HR+ scanner using a 2-D slice of the Hoffman phantom [36] (Fig. 4) with activity ratios in grey matter, white matter, and CSF of 5:1:0. The attenuation factors were computed by assigning a constant attenuation coefficient of 0.095/cm to the support of the phantom. A preconditioned conjugate gradient method [12] was used in computing MAP reconstructions. For each reconstruction we ran 100 iterations to ensure effective convergence. Shown in Fig. 4(c) is an image of the value of  $\kappa^{-2}$  at each voxel. Note the smoothness of this function within the object, which is essential for the approximation (13) to be accurate.

We first evaluated the CRC approximation. Three points of interest were selected, as shown in Fig. 4(b). Since several investigators have observed that the ensemble mean of ML and MAP estimators is approximately equal to the reconstruction that is obtained by applying the estimator to the expectation of the data [2], [4], [5], [9], [37], we measured the CRC from reconstructions of two noiseless data sets: 1) the original phantom sinogram and 2) the sinogram of the phantom after perturbation of a single voxel. Fig. 5 shows a comparison of the measured and theoretically predicted CRC's for each point of interest. The approximations are in good agreement with the measured CRC's, with a maximum error of 6%. When the true projection is unavailable, one can use the projection of the reconstructed image, or even the noisy data, to estimate  $\bar{y}$ . Fig. 5 also shows the CRC's computed using a single realization of the Poisson data. The theoretical predictions do not show a significant increase in error.

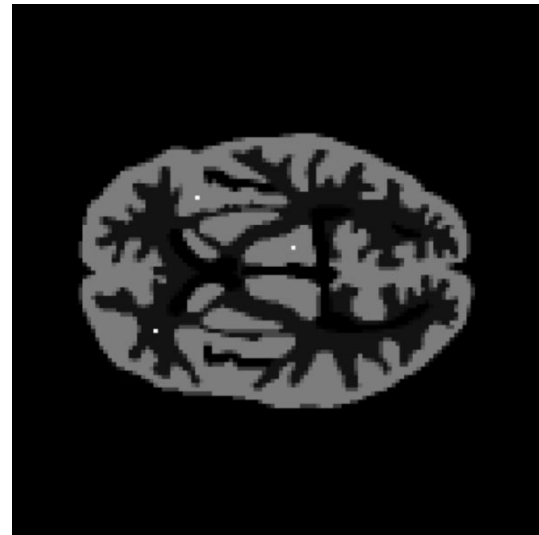
We then computed the voxelwise variances from 8000 reconstructions ( $\beta = 0.0004$ ) from independent data sets, with a total of 200 K mean counts per image. The reconstructed activity in grey matter is about 100 count per voxel. Fig. 6 shows the comparison between the Monte Carlo variances and the theoretically predicted variances. Again, the agreement is good except in regions with zero activity. The reason for the discrepancy in these regions is that the nonnegativity constraint used in the reconstruction algorithm was not included in the theoretical analysis.

### B. Ergodicity: Ensemble Variance Versus Spatial Variance

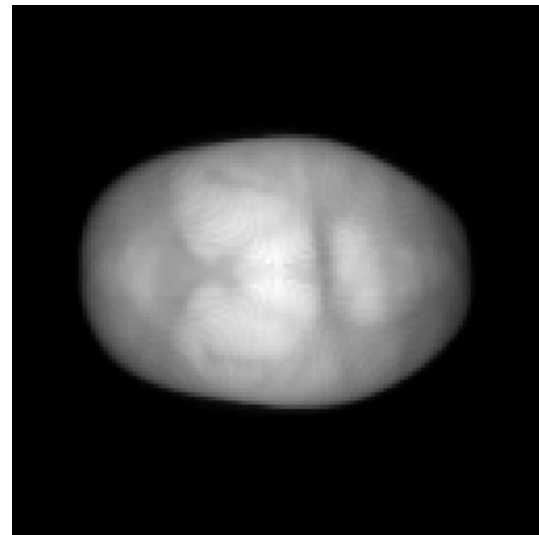
In using the CNR as a figure of merit for choosing the smoothing parameters, we are implicitly assuming ergodicity in the noise in the reconstructions, i.e., we assume that the ensemble variances computed from our theoretical analysis reflect the spatial variance in images reconstructed from a single data set. It is only if this assumption is correct that optimizing CNR could be expected to lead to improved lesion detectability in individual PET scans. We investigated this issue using a uniform elliptical phantom. The total number of counts was again 200 K. Several sets of images were reconstructed from 50 independent noisy data sets, each set corresponding to a different value of  $\beta$  which, in each case, was kept constant throughout the image.



(a)

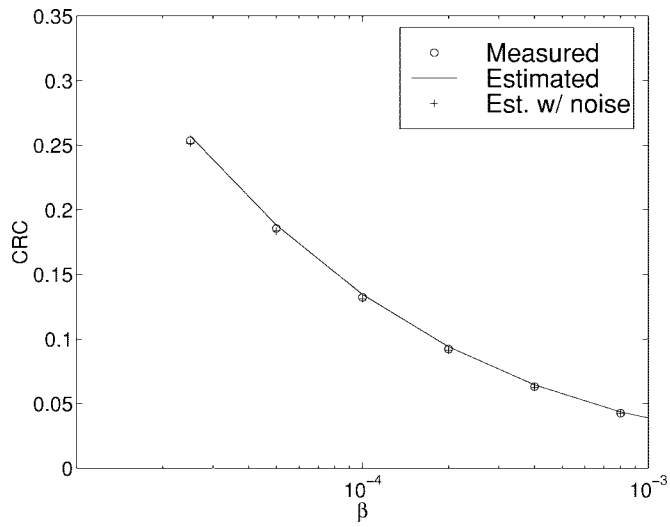


(b)

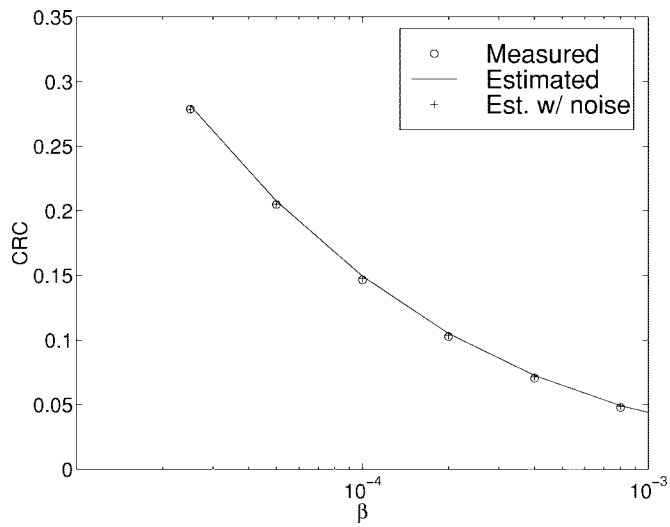


(c)

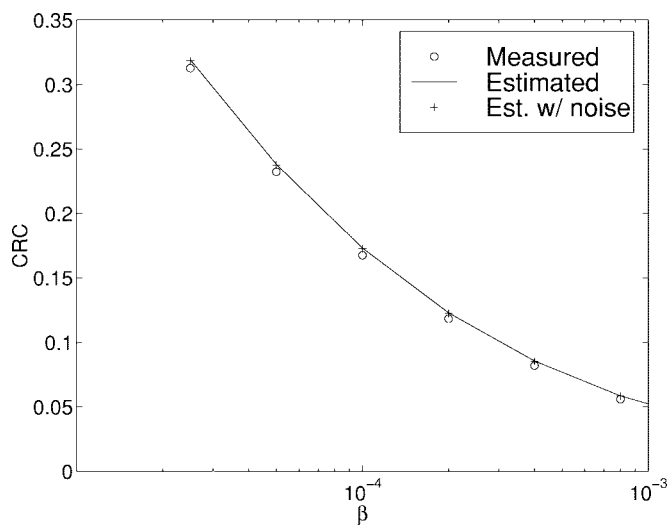
Fig. 4. (a) Digital Hoffman phantom used in numerical evaluations. (b) The three bright spots indicate the locations of the points used in evaluating the accuracy of the CRC formula(24). (c) Image of  $\kappa^{-2}$  for this phantom.



(a)

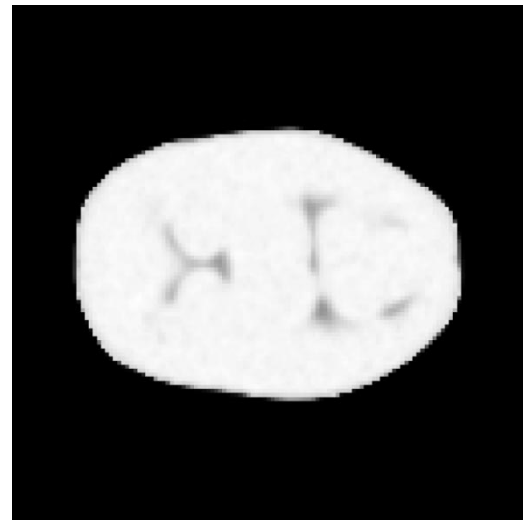


(b)

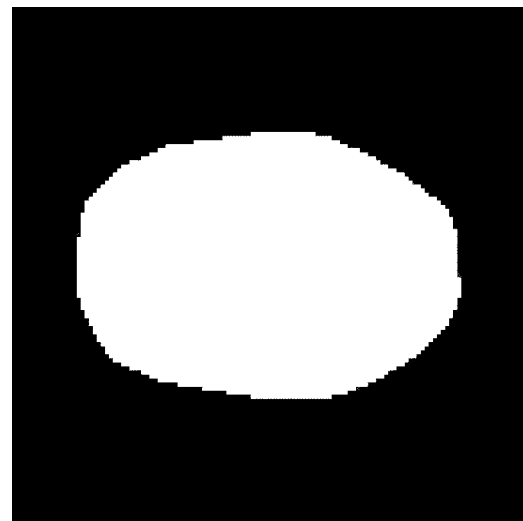


(c)

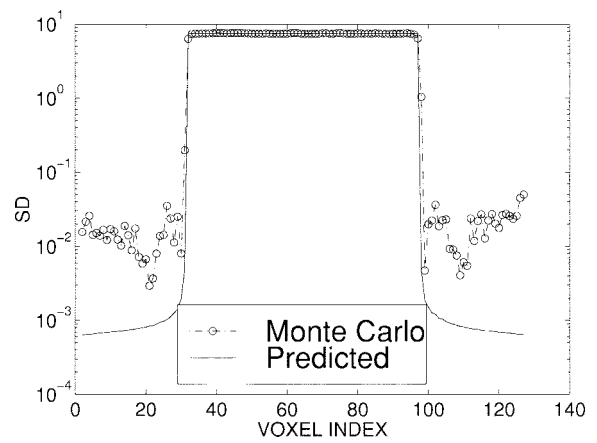
Fig. 5. CRC's computed using (24) with the mean of the data and a single realization of noisy data compared with CRC's measured using direct reconstruction of noiseless data, with and without perturbation: (a) center point in grey matter, (b) off-center point in grey matter, and (c) off-center point in white matter.



(a)



(b)



(c)

Fig. 6. Comparison of voxelwise variances computed using (a) Monte Carlo method and (b) the theoretical approximation in (25). (c) The center profiles through (a) and (b).



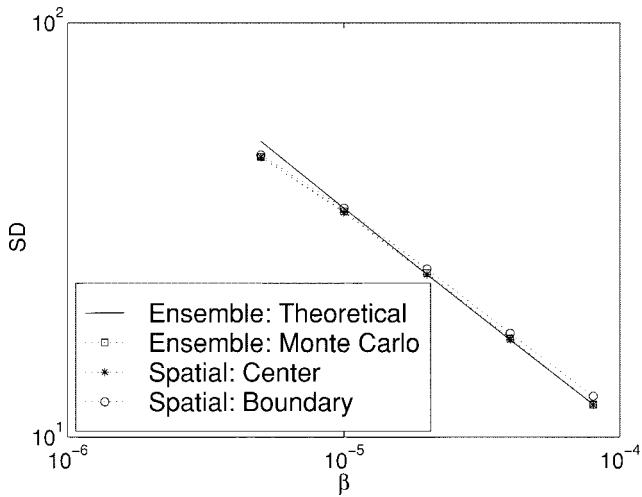


Fig. 7. Comparison of the spatial and ensemble variance computed from reconstructions of a uniform ellipse versus  $\beta$ . Spatial variances were computed in ROI's at the center and edge of the ellipse, ensemble variances were computed at the center of the ellipse across 50 independent reconstructions.

Since a constant  $\beta$  was used, the variance in the reconstructions was approximately uniform inside the field of view (FOV). We could therefore use large regions of interest (ROI's) to compute the spatial variance. Two ROI's were selected: one was a circular region at the center (1976 voxels), the other a nonintersecting boundary region (1980 voxels). The variance across the voxels in each ROI was computed for each individual reconstruction. The mean spatial variance was then computed by averaging the spatial variances across the 50 independent reconstructions. Ensemble variances were computed for each voxel as the variance of that voxel across the 50 reconstructions. The voxelwise variances were then averaged over the voxels in the center ROI. Fig. 7 shows a comparison of the ensemble mean of the spatial variances and the spatial mean of the ensemble variances of the reconstructed ellipse using several different values of  $\beta$ . The variances of the center and boundary ROI's are almost equal, which supports the earlier observation that a constant  $\beta$  will produce near uniform variance. More importantly, the spatial variances agree well with the ensemble variances, both Monte Carlo and theoretical, indicating that the implied ergodicity assumption in our development is reasonable.

### C. Approximations of CRC and Variance: Huber Priors

We used the Huber prior to study the accuracy of our approximations for nonquadratic priors. The Huber potential function is defined as

$$\phi(t) = \begin{cases} \frac{1}{2\delta_h} t^2, & \text{if } |t| \leq \delta_h \\ |t| - \frac{\delta_h}{2}, & \text{otherwise} \end{cases} \quad (35)$$

where  $\delta_h$  is a small constant.

We again used the Hoffman phantom in a simulated CTI EXACT HR+ scanner. The parameter  $\delta_h$  was chosen to be 1% of the grey-matter intensity, which makes the second-order derivative of the prior highly spatially variant. Since the Huber prior is nonquadratic, the measured local impulse response CRC is dependent on the amplitude of the pertur-

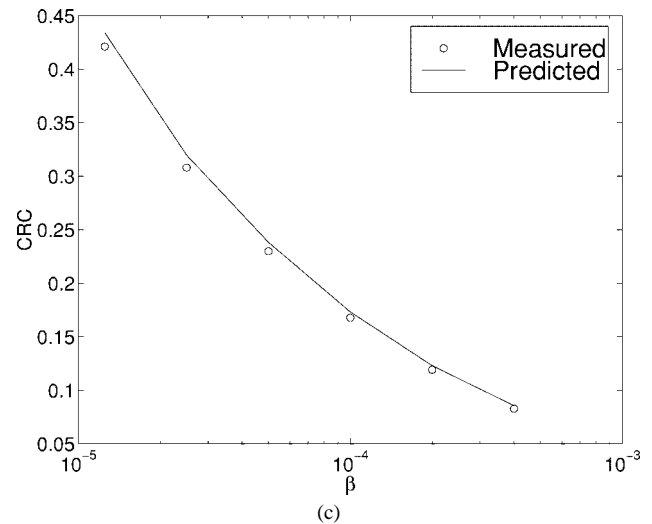
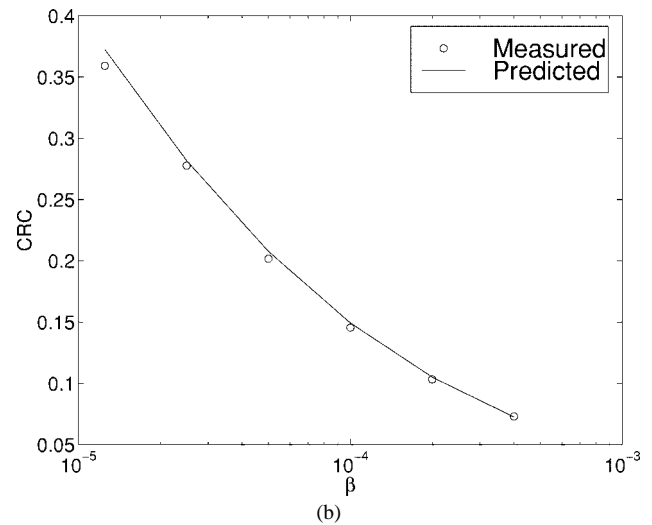
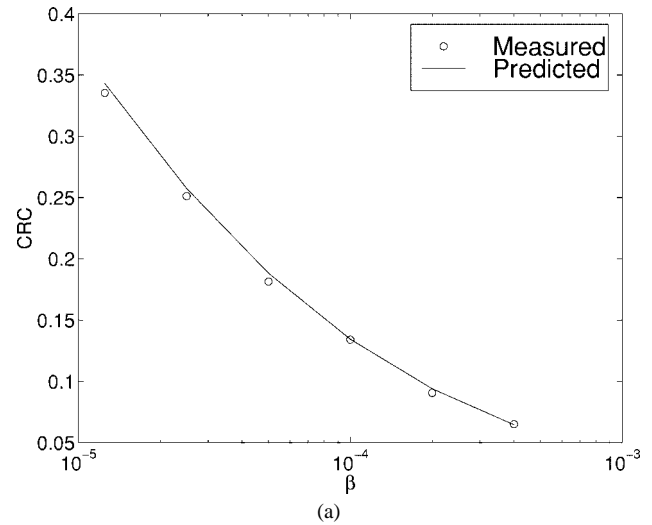


Fig. 8. Comparison of measured and predicted CRC's for small perturbations at three points of interest using the Huber prior (a) center point in grey matter, (b) off-center point in grey matter, and (c) off-center point in white matter.

bation. We studied the CRC's with small ( $=\delta_h$ ) and large ( $=100\delta_h$ ) perturbations. The results are shown in Figs. 8 and 9, respectively. The approximate CRC works quite well with the small perturbation, but breaks down when the perturbation

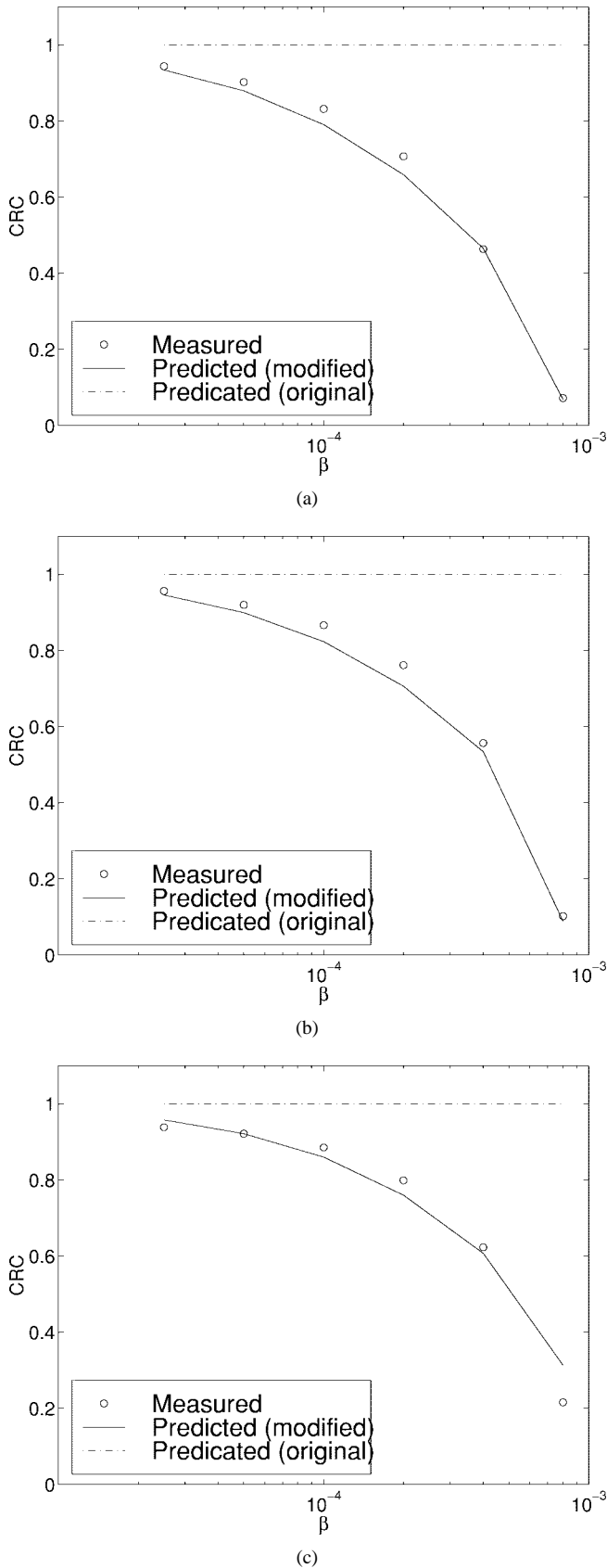


Fig. 9. Comparison of measured and predicted CRC's for large perturbations of three points of interest for Huber prior: (a) center point in grey matter, (b) off-center point in grey matter, and (c) off-center point in white matter. In this case, a finite difference computation of the second derivative of the prior produced reasonably accurate predictions of CRC where the true second-derivative method fails.

is large. Since the local perturbation exceeds the threshold  $\delta_h$  in this situation, the corresponding second derivative becomes zero and hence the CRC's are predicted to be one. However, we found that if we approximate the second derivatives of the nonquadratic prior as the finite difference

$$\ddot{\phi}(t) \approx \frac{\dot{\phi}(2t) - \dot{\phi}(0)}{2t} \quad (36)$$

the computed CRC's show good agreements with the measured results as indicated in Fig. 9. Note the (36) is an *ad hoc* choice that works for the Huber prior studied here. It may not work for other nonquadratic priors. In general, the second derivative of a nonquadratic prior may not be sufficient for analyzing large perturbations. Higher order derivatives or other information may be needed in these situations.

From the above CRC comparison, we can expect that the accuracy of the variance estimate (22) for the Huber prior will be dependent on  $\delta_h$  and the variance in the reconstruction. If the ratio of  $\delta_h$  over standard deviation is large ( $>3$ ), (22) will be fairly accurate in uniform activity regions. When the ratio is small, the estimate will become worse. Figs. 10 and 11 show comparisons between theoretical variance estimates and the results from reconstructions of 500 independent Monte Carlo data sets (total count = 200 k) with  $\delta_h = 30$ ,  $\beta = 0.012$  and  $\delta_h = 10$ ,  $\beta = 0.004$ , respectively. Reconstructions using these two sets of prior parameters have the same CRC for small perturbations ( $<\delta_h$ ). However, they have different variances. In Fig. 10, the theoretical approximations match the Monte Carlo results very well because  $\delta_h$  is far greater than the variance of noise, while there are noticeable errors in Fig. 11 since  $\delta_h$  is comparable to the standard deviation.

#### D. Example Using Maximum Detectability Hyperparameter Selection

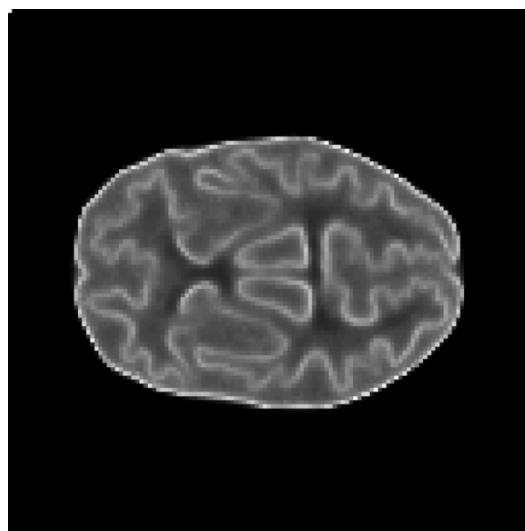
Following the procedure described in Section IV, we reconstructed the Hoffman phantom with the hyperparameter selected to maximize CNR. One hundred iterations of a preconditioned conjugate gradient MAP algorithm were used with a second-order membrane prior. We added two 2-mm diameter lesions in the grey matter with an activity ratio relative to grey matter of 10 : 1. The total number of counts was 200 K. Fig. 12 shows an example of an image reconstructed using this hyperparameter selection scheme in comparison with images reconstructed using 0.5 and 2.0 times these values. It is difficult to judge these single realizations, so a subsequent Monte Carlo evaluation was performed.

We reconstructed 50 independent data sets, each with several  $\beta$  values which were kept constant over the entire image. The background standard deviation was computed from a local grey-matter region surrounding the lesion. The mean CNR (averaged over the 50 realizations) versus the  $\beta$  curve for the center lesion is plotted in Fig. 13. The CNR was normalized to the optimal  $\beta$  computed for the center of the lesion using the CNR optimization technique described in Section IV. Fig. 13 shows that the optimal CNR was indeed achieved at the value computed using (34).

These results are very promising although we know that the simple measure, CNR, is not fully correlated with human ob-



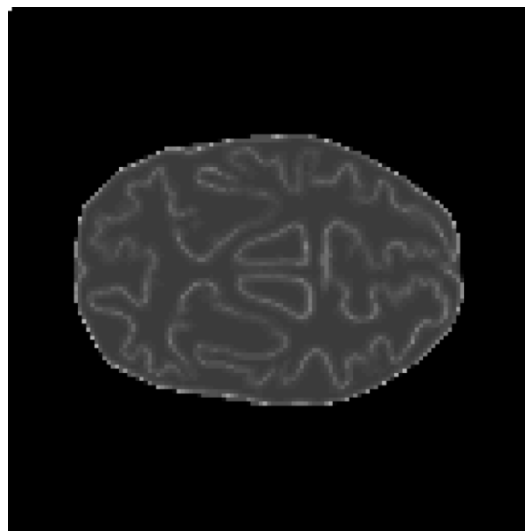
(a)



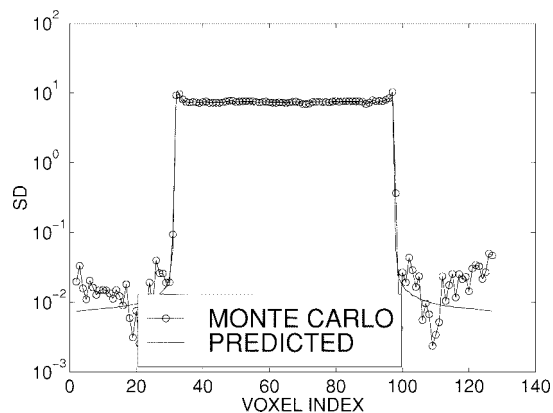
(a)



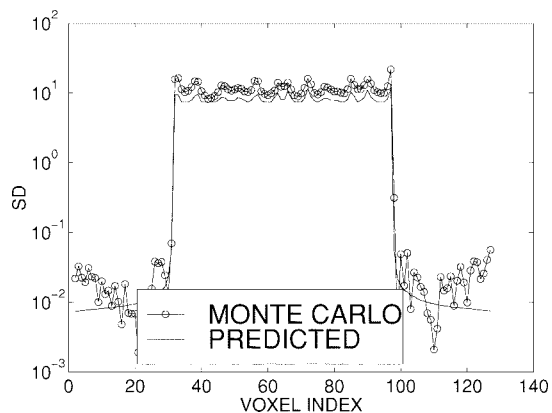
(b)



(b)



(c)



(c)

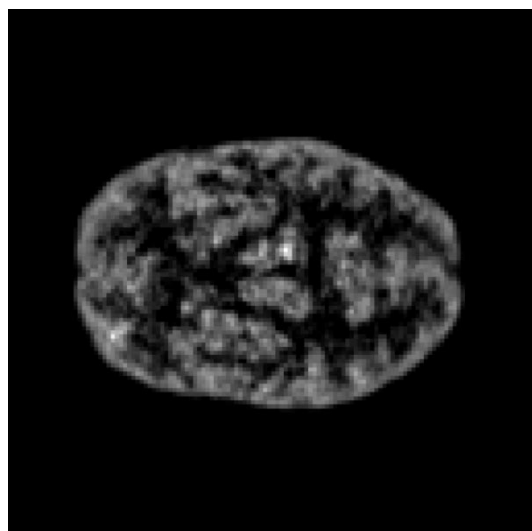
Fig. 10. Comparison of (a) Monte Carlo and (b) theoretical voxelwise computations of variance for the Huber prior with  $\delta_h = 30$ ,  $\beta = 0.012$  (total counts = 200 k). (c) Center profiles through (a) and (b).

Fig. 11. Comparison of (a) Monte Carlo and (b) theoretical voxelwise computations of variance for the Huber prior with  $\delta_h = 10$ ,  $\beta = 0.004$  (total counts = 200 k). (c) Center profiles through (a) and (b).

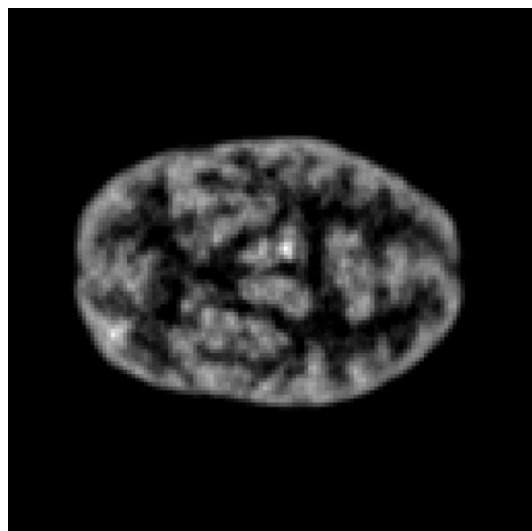
server performance. Receiver operating characteristic (ROC) studies comparing this method with other hyperparameter selection schemes are needed to further evaluate this method.

## VI. CONCLUSION

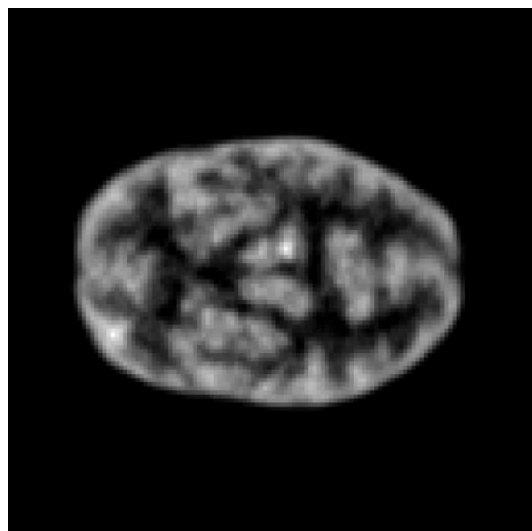
We have derived simplified approximate equations for the local impulse response CRC and variance of each voxel in im-



(a)



(b)



(c)

Fig. 12. MAP images reconstructed from 200 K counts with  $\beta$  selected to maximize CNR, then scaled as follows: (a) half of the optimal hyperparameter, (b) optimal hyperparameter, and (c) double the optimal hyperparameter.

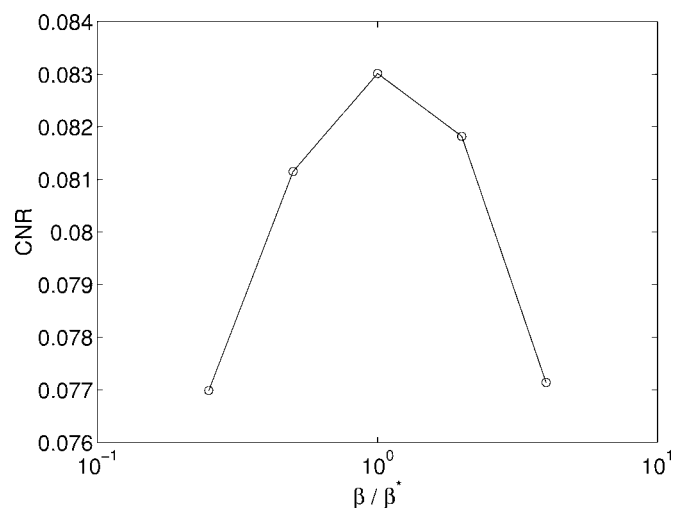


Fig. 13. The CNR versus the  $\beta$  curve averaged over 50 independent reconstructions from 200 K counts, normalized to optimal  $\beta$  computed using (34).

ages reconstructed using MAP estimators. These expressions reveal how resolution and variance changes as a function of the system geometry and the source image. They also show the effect of the hyperparameter on these factors. Computer simulations showed that the approximations agree reasonably well with Monte Carlo results. Using the CRC and variance results, we also proposed a hyperparameter selection scheme that optimizes lesion detectability by maximizing the CRC-to-noise ratio in the reconstructed image. Again, Monte Carlo evaluation showed this technique can indeed maximize lesion CNR ratios.

## REFERENCES

- [1] R. Huesman, "A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography," *Phys. Med. Biol.*, vol. 29, pp. 543–552, 1984.
- [2] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, "Noise properties of the EM algorithm: I. Theory," *Phys. Med. Biol.*, vol. 39, pp. 833–846, 1994.
- [3] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, Oct. 1982.
- [4] D. W. Wilson, B. M. W. Tsui, and H. H. Barrett, "Noise properties of the EM algorithm: II. Monte Carlo simulations," *Phys. Med. Biol.*, vol. 39, pp. 847–872, 1994.
- [5] W. Wang and G. Gindi, "Noise analysis of MAP-EM algorithms for emission tomography," *Phys. Med. Biol.*, vol. 42, pp. 2215–2232, 1997.
- [6] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imag.*, vol. 6, pp. 106–114, June 1987.
- [7] P. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, pp. 84–93, Mar. 1990.
- [8] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, pp. 601–609, 1994.
- [9] J. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Processing*, vol. 5, pp. 493–506, Mar. 1996.
- [10] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Spatial-invariant tomographs," *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 1346–1358, Sept. 1996.
- [11] J. A. Fessler, "Approximate variance images for penalized-likelihood image reconstruction," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, Albuquerque, NM, 1997.

- [12] E. Mumcuoglu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. Med. Imag.*, vol. 13, pp. 687–701, Dec. 1994.
- [13] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Processing*, vol. 5, pp. 480–492, Mar. 1996.
- [14] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image Processing*, vol. 4, pp. 1417–1429, 1995.
- [15] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Processing*, vol. 41, pp. 534–548, Feb. 1993.
- [16] J. A. Fessler, "Penalized weighted least squares image reconstruction for PET," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, June 1994.
- [17] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant image reconstruction," *IEEE Trans. Image Processing*, to be published.
- [18] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," in *Proc. 46th Session ISI*, vol. 52, 1987.
- [19] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Trans. Med. Imag.*, vol. 16, pp. 516–526, 1997.
- [20] Z. Zhou, R. Leahy, and J. Qi, "Approximate maximum-likelihood hyperparameter estimation for Gibbs-priors," *IEEE Trans. Image Processing*, vol. 6, pp. 844–861, 1997.
- [21] P. Craven and G. Wahba, "Smoothing noisy data with spline function," *Numerische Mathematik*, vol. 31, pp. 377–403, 1979.
- [22] V. Johnson, "A model for segmentation and analysis of noisy images," Institute Stat. Decision Sci., Duke Univ., Discussion Paper 91-A151991.
- [23] M. Yavuz and J. A. Fessler, "Objective functions for tomographic reconstruction from randoms-precorrected PET scans," in *Proc. IEEE Nuclear Science Symp. and Medical Imaging Conf.*, Anaheim, CA, 1996, vol. 2, pp. 1067–1071.
- [24] E. Mumcuoglu, R. Leahy, S. Cherry, and E. Hoffman, "Accurate geometric and physical response modeling for statistical image reconstruction in high resolution PET," in *Proc. IEEE Nuclear Science Symp. and Medical Imaging Conf.*, Anaheim, CA, 1996, pp. 1569–1573.
- [25] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar, "High resolution 3D Bayesian image reconstruction using the microPET small animal scanner," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 1001–1013, 1998.
- [26] J. Qi and R. M. Leahy, "Resolution and noise properties of map reconstruction for fully 3-D pet," in *Proc. Int. Meeting Fully 3D Image Reconstruction Radiology Nuclear Medicine*, 1999, to be published.
- [27] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.
- [28] R. M. Kessler, J. R. Ellis Jr., and M. Eden, "Analysis of emission tomographic scan data: Limitations imposed by resolution and background," *J. Comput. Assist. Tomogr.*, vol. 8, pp. 514–522, 1984.
- [29] L.-N. D. Loo, K. Doi, and C. E. Metz, "A comparison of physical image comparison indices and observer performance in the radiographic detection of nylon beads," *Phys. Med. Biol.*, vol. 29, no. 7, pp. 837–857, 1984.
- [30] A. Rosenfeld and A. Kak, *Digital Picture Processing*, 2nd ed. New York: Academic, 1982, vol. 1.
- [31] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Processing*, vol. 44, pp. 2026–2041, Aug. 1996.
- [32] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT, 1987.
- [33] S.-J. Lee, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Trans. Med. Imag.*, vol. 14, pp. 669–680, Dec. 1995.
- [34] K. J. Myers, H. H. Barrett, v M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in Med. Imag.," *J. Opt. Soc. Amer. A*, vol. 2, pp. 1752–1759, 1985.
- [35] J.-S. Liow and S. C. Strother, "Practical tradeoffs between noise, quantitation, and number of iterations for maximum likelihood-based reconstructions," *IEEE Trans. Med. Imag.*, vol. 10, pp. 563–571, Dec. 1991.
- [36] E. Hoffman, P. Cutler, W. Digby, and J. Mazziotta, "3D phantom to simulate cerebral blood flow and metabolic images for PET," *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 616–620, Apr. 1990.
- [37] R. Carson, Y. Yan, B. Chodkowski, T. Yap, and M. Daube-Witherspoon, "Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm," *IEEE Trans. Med. Imag.*, vol. 13, pp. 526–537, Sept. 1994.